Beyond the Individual in the Evolution of Language

David J. C. Hawkey
M.Phys, M.Sc.

A thesis submitted in fulfilment of requirements for the degree of
Doctor of Philosophy

to
Linguistics and English Language
School of Philosophy, Psychology and Language Sciences
University of Edinburgh

September 2008

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

David J. C. Hawkey

Abstract

This thesis concerns the evolution of language. A proliferation of theoretical models have been presented in recent years purporting to offer evolutionary accounts for various aspects of modern languages. These models rely heavily on abstract mechanistic models of the production and reception of language by modern humans, drawn from various approaches in linguistics which aim at such models. A very basic and ubiquitous assumption is that expressions have meaning in virtue of being associated with internal representations, and that therefore the evolution of language can be modelled on the basis of individuals trying to produce external manifestations of these internal "meanings". I examine the role of this assumption in language evolution theorising, and review evidence from neuroscience and first language acquisition relevant to the validity of this assumption. The chaotic nature of the relationship between "meaning" and the brain undermines the supposition that the evolution of language was driven by spontaneous association between internal structures and external forms.

I then turn to the philosophical basis of language evolution theorising, adopting a Wittgensteinian perspective on the cognitive interpretation of linguistic theories. I argue that the theoretical apparatus of such approaches is embedded in language games whose complicated rules relate to linguistic behaviour (and idealisations of that behaviour) but not to neural organisation. The reinterpretation of such descriptions of language as descriptions of the internal structures of language users is rejected as a grammatical confusion: if the rules for constructing linguistic theory descriptions do not mention neural structures, then theoretical descriptions of the linguistic abilities of an individual say nothing non-trivial about their internal brain structure. I do not deny that it would, in principle, be possible to reduce linguistic theories (reinterpreted as mechanistic descriptions) to neural structures, but claim that this possibility is guaranteed only by leaving the practice of re-describing physical brain descriptions entirely unconstrained.

Thus the idea that we can reasonably infer the behaviour of humans and prehumans in more primitive communicative environments by manipulation of the models of linguistic theories is unfounded: we have no idea how such a manipulation would translate into statements about neural organisation, and so no idea how plausible such statements about earlier neural organisation (and the resultant behaviours) are. As such, cognitive interpretations of linguistic theories provide no better ground for statements about behaviour during earlier stages in the evolution of language than guessing.

Rejecting internal-mechanism based accounts as unfounded leaves the evolution of language unexplained. In the latter parts of this thesis, I offer a more neutral approach which is sensitive to the limited possibilities available for making predictions about human (and pre-human) behaviour at earlier stages in the evolution of language. Rather than focusing on the individual and imputed internal language machinery, the account considers the communicative affordances available to individuals. The shifts in what individuals can learn to do in interaction with others, that result in turn from the learning of interactive practices by others, form the basis of this account. General trends in the development of communicative affordances are used to account for generalisations over attested semantic change, and to suggest how certain aspects of modern language use developed without simply assuming that it is "natural" for humans to (spontaneously) behave in these ways. The model is used in an account of the evolution and common structure of colour terms across different languages.

Acknowledgements

To Eilidh for her love, support and encouragement, for putting up with the imposition of this thesis in our lives, and for agreeing to marry me.

To Mum and Dad for being so supportive throughout my time as a student and for sparking my interest at such an early age in the importance of interaction for language learning and use.

Thanks to my supervisors Simon Kirby and Jim Hurford, both for their direct help in the production of this thesis and for maintaining such an open intellectual environment in the LEC. I am grateful also to Ronnie Cann for several enjoyable and stimulating meetings during my second year. Thanks to the other members of the LEC, especially Stefan Höfler and Clare MacCumhaill for innumerable thought provoking and entertaining discussions.

To the Economic and Social Research Council for financial support.

To Sadivino for the sandwiches.

Contents

List of Tables

List of Figures

Abbreviations

BB	*The Blue and Brown Books* (Wittgenstein 1969)
OC	*On Certainty* (Wittgenstein 1975)
OED	Oxford English Dictionary
PG	*Philosophical Grammar* (Wittgenstein 1974)
PI	*Philosophical Investigations* (Wittgenstein 1967)
RFM	*Remarks on the Foundations of Mathematics* (Wittgenstein 1978)
RoC	*Remarks on Colour* (Wittgenstein 1977)
WCS	World Colour Survey
Z	*Zettel* (Wittgenstein 1981)

CHAPTER 1

Introduction

“Only humans have language; other animals do not.” This statement, may be seen as one source of interest in, and motivation to study, language. Unique to and ubiquitous among humans, language appears to be the key to understanding “what exactly it is that makes us human,” (Christiansen and Kirby 2003b, p. 1). However, the statement opening this paragraph can be taken in several different ways. For example, “only humans have language” may be taken to mean that humans are unique from a biological perspective in that they alone possess the capacities to learn and exploit the communicational conventions of the communities they are born into. In contrast, in spite of attempts to teach them, animals appear incapable of learning these interactive techniques. Another (related, though not identical) way of taking the statement is that it is only among humans that such practices, which individuals come to master, exist. Here “having language” is taken not as having a capacity or biological endowment, but as something like having a certain kind of system of communication or interaction (and different groups of humans may have different systems). What humans “have” in this sense is open to view rather than hidden in their minds or in their brains.[1]

These ways of taking the statement are clearly related: if humans did not have the capacity to master certain techniques, then these techniques could not exist

[1] Languages notoriously resist any principled delimitation, the distinction between two “different languages” often being as much a political matter as a reflection on the degree to which individuals can successfully communicate. Nothing in this thesis depends on a rigid definition of such theoretical entities as “the English language”, and the term “language(s)” is used in a rough manner to denote ways in which more or less well defined groups interact communicatively with each other.

among humans to be mastered. But having the capacity to engage in some technique leaves open questions about how it came to pass that such techniques exist (just as prior to 1968 the biological capacity to perform a Fosbury Flop was no less present in the human species than it is today, yet the now–widespread high jump technique did not exist). It is this kind of question, about the emergence and development of languages rather than the biological evolution of the capacities to learn and use languages, which this thesis primarily focuses on. If we do not assume that languages spontaneously appeared overnight, how can we account for their gradual emergence, perhaps from systems of inter-individual interaction we would not recognise as languages, and can such an account shed light on why modern languages are as they are?

Languages are culturally inherited systems (or groups of techniques), passed down by individuals learning to use them (through interaction with and observation of others), thereby becoming (amongst other things) individuals from whom the language may be learned. This is a dynamic process on which numerous factors can impact and thereby affect the languages that are inherited and maintained in this way. How should we think about the processes of cultural evolution when trying to account for the emergence and development of languages?

One possibility is to return to the first way of interpreting the statement that "humans have language" and to consider the nature of the human capacities to learn and use language. Many approaches in the modern study of language present models of brain-internal processes that take place when language is learned, produced and understood (e.g. Chomsky 1995; Sperber and Wilson 1995; Croft and Cruse 2004; Hurford 2003; Gallese and Lakoff 2005; Tomasello et al. 2005; Evans and Green 2006 to name but a few). Such models offer the hope of constructing a *mechanistic* model of the cultural evolution of language by acting as a bridge between individuals' experience with others and their own linguistic behaviour. If the likely language-relevant behaviour of individuals can be computed (on the basis of internal mechanisms) given a context of others whose communicative behaviour may differ from modern humans', then it will be possible to create plausible models of the evolution of language by computing the behaviours of individuals in populations, iteratively learning and producing a language (Kirby and Hurford 2002). This approach can have quite a powerful impact on our view of the development of languages. If, for example, the process of language learning involves learners with "some initial delimitation of a class of possible hypotheses about language structure" (Chomsky 1965, p. 30) from which they select a structure on the basis of their experience, then it seems possible to model numerous individuals selecting structures on the basis of their experience and producing language according to that structure from which other individuals may then learn (i.e. select a structure). As individuals need not necessarily select the same structure as those from whom they learn (perhaps their experience is limited or noisy) pathways by which languages develop, or likely stable states for languages to be in may be determined (e.g. Kirby et al. 2007).

Such "iterated learning models" can suggest interesting theoretical possibilities for the evolution of multi-agent interactions. However, as (perhaps abstract) accounts of the actual emergence and development of human languages, they are only as convincing as the cognitive models on which they rely. Similarly, other accounts of language evolution which may not as explicitly embrace the idea of iterated learning but nonetheless rely on behavioural assumptions drawn from models of internal processes, are beholden to the validity of the models of cognition they employ. A subsidiary, though important question in this thesis is to what degree can such approaches, which rely on contemporary models of internal processes underpinning our linguistic capacities, contribute to our understanding of the cultural evolution of our languages?

1.1 Thesis structure

An interest in language can take various forms. One may, for example, consider the sounds used in speech and the relationships between them, or one may be interested in the structure of spoken sentences or sentences that are judged to be grammatical. Attempting to cover every aspect of languages from the perspective of cultural evolution would be an undertaking beyond the scope of this thesis. Instead, the focus here is primarily on the meanings we are able to express with our languages. How did the transition from animal-like interactions to modern possibilities for "meaning something" occur?

Focusing on meaning is a useful way of getting a handle on the suitability for an account of language evolution of contemporary models of internal processes, as it is a near universal assumption of such models that the internal processes associated with language involve representations of meaning. Chapter 2 looks at the role of internal representations of meaning in theories of language evolution, arguing that assumptions made by theories about meaning at earlier stages of language evolution can (and often do) rely on the inference that our ancestors would have had internal representations akin to our own, and that these would have driven them to use language in meaningful ways that parallel our own uses of language (e.g. by using some expressions as names for kinds of object). These assumptions are substantial, involving individuals spontaneously using expressions in complex and diverse ways.

Chapter 3 goes on to examine whether this kind of inference is warranted. Neuroscientific evidence concerning the internal processes involved in language is reviewed to see if any support for the assumptions detailed in chapter 2 can be found. On the basis of neuropathology and imaging studies, it is argued that the complexity of structures that may be thought of as being representations of meaning suggests it is unlikely that these would play the roles attributed to them in language evolution theories. Evidence for substantial plasticity in language-related neural structures is reviewed and used to argue that the kinds of neural structure we see in relation to the meanings of words develop through experience of language, and are therefore poor candidates as drivers of the development of such meanings. The role of learning is emphasised in a brief review of first language acquisition literature which casts doubt on the idea that children show spontaneous usage generalisations, independent of experience, that would be expected were they mapping external forms onto existing internal representations.

If internal representations of meaning cannot be relied on as the driving force behind the development of meaning in language, how should we think about meaning in the evolution of language? Chapter 4 turns to conceptual questions about language and meaning, and adopts Wittgenstein's perspective on language developed in his later philosophy. Rather than thinking of language as the external manifestation of hidden structures, language is seen on an analogy with games as a series of rule-governed practices. The goals of chapter 4 are two-fold: to introduce a way of thinking about language that doesn't rely on internal representations of meaning, and to examine the conceptual foundations of cognitive models of language that may be employed in theories of language evolution.

Languages are populated with expressions whose uses in language games are governed by a great diversity of rules: some, but not all, relate to physical objects; some, but not all, relate to other expressions, etc. The diversity of rules is illustrated in chapter 4 by considering ways in which psychological expressions and expressions from the metalanguages used in descriptions of language function. In detailing these rules, we get a clearer picture of what it amounts to for a description employing these terms to be true (for example, what statements like "utterances are constructed from words" or "speech consists of a series of phonemes" amount to). I will argue that since these language games all proceed without reference to neural structures, the fact that such a statement may be true does not say anything significant about the brain. So, for example, we may grant that when a person means something by uttering *x* then they "intended the utterance of *x* to produce some effect in the audience by means of the recognition of this intention," (Grice 1957, p. 385) without concluding that this says anything significant about mechanistic processes of language production and comprehension. This is important for the role of cognitive models in mechanistic accounts of language evolution, as generally these models are couched in terms of such expressions, and so their interpretation as descriptions of the *mechanisms* by which humans learn, produce and understand language is mistaken. I also consider the possibility that expressions in these models may be interpreted as re-using existing expressions in novel ways (namely as labels for entities or processes inside the brain). While it remains a theoretical possibility that rules could be devised to relate neural processes to cognitive models, I argue that, since such relationships are unknown (and the review of empirical studies in chapter 3 suggest they would be highly complex), the cognitive models of language available to us offer no basis for predicting the behaviour of humans in the circumstances of earlier stages in the evolution of language.

Chapter 5 addresses the challenge of accounting for language evolution without relying on cognitive models. If language is conceived of as a series of normatively governed activities, an account of the cultural evolution of language should describe how such activities (and the boundaries between correct and incorrect use of expressions) can develop. This chapter presents a way such accounts can be constructed, by considering what is available to be learned in an individual's environment, and what impact his learning can have on the communicational environments of others. This way of thinking about the development of language is applied to attested semantic change in order to show that an account that does not depend on a cognitive model can be constructed and can offer insight into the development of language. Lessons drawn from this way of accounting for semantic change are used to sketch accounts of the evolution of some very basic aspects of our languages (e.g. that we can use expressions to "refer to" objects) from what were presumably very different communicative systems. Rather than seeing the ways we use language as coming naturally to humans as individuals, they are seen as the result of cultural evolution.

Finally, chapter 6 offers an account of the development of colour terms based on the approach described in chapter 5. The possibility for an expression to develop into a colour term is seen as the result of various correlated colour distinctions in the human-relevant environment. Similarities found in the colour term systems of the world are thus accounted for by arguing for similarities in humans' colour environments.

1.2 Thesis overview

The "take home message" of this thesis is that when thinking about the cultural evolution of language (and when trying to address questions about why languages are as they are), we should go beyond the individual in two senses. First, our current ignorance of the workings of the brain and the high likelihood that contemporary cognitive models make misleading behavioural predictions in contexts with which we are unfamiliar, oblige us to construct accounts of the evolution of language which do not rely on models of individuals' cognition. Second, the steps made in this thesis toward such an account indicate that it is a mistake to view properties of languages and trends in language change as reflections of the properties of individuals or individually based tendencies to alter language in certain ways. In going beyond the individual, we will be able to shed light on how humans came to have the languages we have.

CHAPTER 2

Externalisation as a window on language evolution

> Spoken words are the symbols of mental experience and written words are the symbols of spoken words. (Aristotle, *De Interpretatione*, chapter 1)

> Language encodes and externalises our thoughts by using symbols. (Evans and Green 2006, p. 21)

2.1 The Windows Approach to language evolution

A commonly cited problem in tackling questions of language evolution is evidential paucity. Direct evidence concerning the factors relevant to the evolution of language (both in terms of the biological evolution of language users and the evolutionary development of languages) is not available for obvious reasons. However, in his review of a number of contemporary perspectives on language evolution, Botha (2003) concluded that the "main obstacle to gaining a better understanding of central aspects of the evolution of language is a poverty of a restrictive theory" (p. 7). Botha argued that in the absence of direct evidence, theorists were constructing theories on the basis of inferences from phenomena about which direct evidence is available to phenomena of earlier stages in language evolution about which direct evidence is unavailable, and that often these inferences were based on poorly developed assumptions. In subsequent work, he has sought to tackle this problem by identifying these inferences (which he

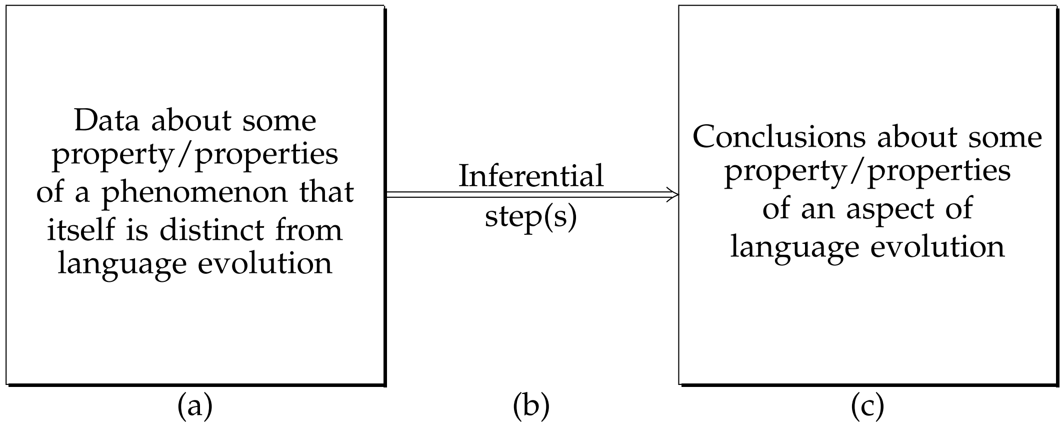

Data about some property/properties of a phenomenon that itself is distinct from language evolution
Inferential step(s)
Conclusions about some property/properties of an aspect of language evolution
(a)
(b)
(c)

Figure 2.1: The structure of a window on language evolution, adapted from Botha (2006a)

calls "windows on language evolution") and suggesting a number of ways in which they may be appraised (Botha 2006a, 2006b, 2007).

This chapter adopts Botha's scheme and presents a particular inference about our ancestors during earlier stages in the evolution of language that can be used to justify assumptions about meaning in earlier communication systems. The inference is based on the notion of externalisation: utterances of contemporary languages are meaningful because they are the external manifestation (or externalisation) of brain-internal representations (c.f. Evans and Green 2006; Fitch 2007). The inference is that at earlier stages in the evolution of language, individuals would have had systems of internal representation similar to our own and that communication systems would have been based on the externalisation of these representations. Similarities in internal representations may be used (by language evolution theories making this inference) to justify assertions that the meaning of communicational units (e.g. utterances or parts of utterances) at earlier stages in the evolution of language would have been similar to contemporary meanings. I will refer to this inference as the externalisation window on language evolution.

To clarify the concept of a window on language evolution and the ways such windows may be judged, Botha (2006a) introduces a schematic structure for a window inference (figure 2.1). Botha identifies three basic features which determine how good a window is as an inferential device: groundedness, warrantedness and pertinence. These features refer to parts (a)–(c) of figure 2.1 respectively.

Groundedness The groundedness of a window refers to how well the phenomenon (or phenomena), schematically represented in part (a) of figure 2.1, is (or are) understood. If a window on language evolution is not well grounded then either the inference is based on a poorly understood or poorly defined phenomenon, or it is based on a misunderstanding of a phenomenon. In the former case it is difficult to have confidence in the inference as there is a danger that the description of the phenomenon is chosen to fit the inference being made. In the latter case, while the logic of the inference may be sound, the conclusion (part c of figure 2.1) does not apply as the premise(s) do not hold. The groundedness of the externalisation window will be discussed in chapter 4.

Warrantedness Warrantedness refers to the validity of the inferential steps represented by the arrow (b) in figure 2.1. If a window phenomenon is grounded, then one has a clear characterisation of it. If a window is warranted, then there are clear reasons for accepting the phenomenon as actually showing something about earlier stages in language evolution. In order for a window inference to be warranted, it requires a "bridge theory" linking the currently observable phenomenon to the stage of language evolution being inferred. Chapter 3 tackles the warrantedness of the externalisation window by looking at empirical studies into internal representations and at the processes of language development in infants and young children.

Pertinence The pertinence of a window inference is a property of what is inferred (represented by (c) in figure 2.1). A window is pertinent to the degree that conclusions drawn on the basis of the window are actually conclusions about language evolution. The rest of this chapter discusses the pertinence of the externalisation window within theories of language evolution.

2.2 Pertinence of the externalisation window

As the notion of externalisation isn't always explicitly embraced in theories of language evolution, this section reviews ways in which certain theories of language evolution either rely on externalisation or may be interpreted as so relying.

2.2.1 *Theories of "protolanguage"*

Derek Bickerton (1990) is generally credited with having introduced the term "protolanguage" into the field of language evolution. The term refers to a hypothesised stage in language evolution that "helps to bridge the otherwise threatening evolutionary gap between a wholly alingual state and the full possession of language as we know it" (Bickerton 1995, p. 51). ("Protolanguage" in this sense is distinct from the kinds of proto-language reconstructed by historical linguists, such as Proto-Indo-European.) While Bickerton uses the term freely to refer to both this hypothesised stage and certain contemporary forms of communication (including pidgins, the language of young children, and the use of symbols by trained apes; Bickerton 1995), not all theories accept this identification (e.g. Arbib 2003). Here I will use the term "protolanguage" to refer only to a hypothesised stage in the evolution of language.

A broad distinction can be drawn between suggestions of the form of protolanguage. "Synthetic" models (e.g. Bickerton 1990; Jackendoff 2002) imagine protolanguage consisted of units whose meanings were analogous to modern words which could be combined to create messages though syntactic phenomena were either absent (Bickerton 1995) or limited (Jackendoff 2002). The term "synthetic" is used to highlight the notion that utterances were put together from a stock of symbols. "Holistic" models (e.g. Wray 1998; Arbib 2003) imagine protolanguage consisted of units whose meanings were less like modern words and more akin to modern utterances. Alison Wray (2000) emphasises the relationship between the functions of modern "formulaic sequences" and chimpanzee holistic noise/gesture signals while Arbib (2003, p. 183) suggests protolanguage utterances may have been "names" for "complex but frequently important situations". One route from a holistic protolanguage to modern language involves breaking holistic utterances down into smaller components. This process was suggested by Wray (1998) and features in a number of computational simulations of language evolution (e.g. Kirby 2002; Vogt 2003). Because this suggested route from protolanguage to modern language involves taking utterances apart it is sometimes referred to as the "analytic route" though this term may cause confusion as the "analytic system" is sometimes used to refer to the system which (purportedly) builds modern grammatical sequences from word-like symbols (Wray 2002a).

2.2.2 Synthetic protolanguage

2.2.2.1 Differences between synthetic protolanguage and animal communication

As a bridge between animal communication and human languages, synthetic protolanguage is conceived by its proponents as sharing with language some properties that are absent in animal communication (e.g. the use of symbols). These properties resulted in an open ended system which allowed protolanguage users to convey "an infinite amount of information" (Bickerton 1995, p. 14). Bickerton (1995) lists a number of properties he attributes to language (and to protolanguage) but not to animal communication:

- Animal communication uses *iconic* signals for which the relationship between the message expressed and the form expressing it is "straightforward and transparent." Examples given by Bickerton include lowering the head and/or gaze or presentation of the rump to indicate submission.[1] Linguistic symbols, in contrast, are arbitrary in the sense that "they lack any apparent connection with the objects or actions they represent" (Bickerton 1995, p. 19). Vervet alarm calls may constitute an exception to the rule that animal signals are iconic, and Bickerton (1995, p. 18) sees arbitrary alarm calls as plausible candidates for the first use of arbitrary symbolism.
- Many animal signals are *gradient*, while the symbols of language are not Bickerton (1995, pp. 18–20). For example, a bird determined to defend its territory to the death will sing louder and longer than one whose intent is weaker. In contrast, a *l-o-o-o-o-ng* journey is not necessarily longer than a *long* journey, and *warm* is not an intermediate form (in length, tone or vowel quality) between *hot* and *cold*. Additionally, linguistic symbols are "produced by the recombination of a small number of fairly abstract units (in vocal language, speech sounds; in sign language, hand shapes)" (p. 26).
- Animal signal systems lack systematic relationships present in linguistic units. While animals may have signals expressing anger, Bickerton (1995) states that no known species has different signals for "expressing anger at close kin, anger at a non-kin conspecific, anger at a member of another species, and so on" (p. 19). Languages, in contrast, have subset-superset

[1] Bickerton (1995) doesn't explain why this particular action straightforwardly and transparently expresses submission, but appeals to the ubiquity of this signal and the absence of a reverse relationship (i.e. gaze lowering or rump presentation to signal dominance) in animal communication. There may be arguments concerning e.g. the relationship between body size and likely success in a violent interaction that justify seeing gaze lowering and rump presentation as iconic, but Bickerton doesn't make them.

> relations: in English, for example, a *bluebottle* is a specific kind of *fly*, and a *fly* is a specific kind of *insect*.

Bickerton (1995, p. 26) offers an explanation why linguistic symbols should have these properties. Language plays the role of a representational system: "an ordered picture of the world, arranged so that the items of information in it can be swiftly and easily located" (Bickerton 1995, p. 20). The profusion of things in the world to represent means that the symbols of a language must be numerous. A large number of symbols is "more economically produced" by combining small abstract units (ibid. p. 26). In order to reduce the complexity of the sensory world to "a manageable level of simplicity," Bickerton states the representational system must categorise, which "rules out gradient, non-discrete units," (ibid. p. 26). Finally, the subset-superset relationships among linguistic symbols allow them to be "filed hierarchically, not thrown together like junk in a closet" (ibid. p. 27) which aids rapid recovery of units.

The development of these features in protolanguage can therefore be seen as the result of protolanguage being a system for representing the world. According to Bickerton, this representational function of protolanguage is tied to the fact that, somehow, (proto-) humans developed the ability "to attach symbols to categories and to use those symbols correctly [. . .] without training," (ibid. p. 52). As will be seen below (section 2.2.2.4) Bickerton cashes out these ideas about representation and categories in terms of internal representations. Therefore one important role of externalisation in Bickerton's protolanguage is as a driver for the development of aspects of protolanguage absent from animal communication systems.

2.2.2.2 *The development of syntax*

Synthetic protolanguage is conceived by its proponents as "differing from fully developed modern language in its vocabulary size, its lack of syntax and its lack of a modern phonology, but in no other significant respects" (Bickerton 2007, p. 515). Since the units of protolanguage are presented as having the meanings of some of the units of modern languages (predominantly nouns and verbs), the absence of syntax is taken to imply that protolanguage was inherently ambiguous (Bickerton 1995; Calvin and Bickerton 2000; Jackendoff 2002). For example, the synthetic protolanguage equivalent of *I saw Og taking Ug's meat* (which would lack the morphemes *-ing* and *'s* and thus be rendered as *I saw Og take Ug* *meat*) would pose a number of difficulties to a protolanguage hearer, as numerous other meanings could be expressed by the same "bag of words". Calvin and Bickerton (2000) describe the process of interpreting the final three words thus:

> No surprise to find two nouns together (this happens all the time in protolanguage) but their relationship is up for grabs, even with the verb "take" around to help out. Could it mean that Ug was taken to the meat, or from the meat, or the meat was taken to him or from him? (Calvin and Bickerton 2000, pp. 141–142)

Jackendoff (2002) imagines a similar scenario in which synthetic protolanguage utterances had ambiguous meanings:

> The first essential innovation would be the ability to concatenate two or more symbols into a single utterance, with the connection among them dictated purely by context. For example, *Fred apple* (imagine this uttered by an eighteen-month-old or a signing chimp) might express any number of connections between Fred and apples, expressible in modern languages in sentences such as *That's Fred's apple*, *Fred's eating an apple*, *Fred likes apples*, *Take the apple from Fred*, *Give the apple to Fred*, or even *An apple fell on Fred*. (Jackendoff 2002, pp. 245–246)

The ambiguity of synthetic protolanguage (which derives from its units having modern-word-like meanings) is important to both Bickerton's and Jackendoff's accounts of syntax. Both theorists use the term *syntax* with systematic ambiguity (c.f. Chomsky 1965, p. 25) between (broadly) a description of utterances and a (biologically based) cognitive faculty, specific to language. Thus, in their schemes, "the evolution of syntax" refers to a process of biological evolution which impacts on the forms and uses of utterances. The inherent ambiguity of synthetic protolanguage provides the grounds for identifying a fitness differential between individuals without and individuals with syntax, as syntax reduces ambiguity/increases expressive power in communication (Jackendoff 2002, p. 237). It is therefore important to the role of protolanguage in this story about the evolution of syntax that protolanguage's (syntactic-less) units had the same meanings as the units of modern (syntactical) language: this is the basis on which it is assumed that protolanguage would have been inherently ambiguous; and the ready stock of modern like meaningful units is what syntax emerged to operate on and thereby produce modern-like languages.

2.2.2.3 *Meaning in synthetic protolanguage*

What do the assertions that synthetic protolanguage units and utterances would have had the meanings and ambiguities attributed to them amount to? Jackendoff's suggestion that we "imagine [*Fred apple*] uttered by an eighteen-month-old or a signing chimp," (see above) is potentially misleading as both these cases involve an already established language. The child or chimp is learning a system in which utterances with Jackendoff's suggested meanings may be degraded to give the proto-utterance *Fred apple*; that is, given that the child/chimp's ambient language allows the expression of Jackendoff's connections between Fred and apples with combinations of symbols which include *Fred* and *apple*, the proto-utterance *Fred apple* may result as an approximation to the expression appropriate to the target language. However, in the case of protolanguage as an evolutionarily earlier stage in language evolution, no such ambient language exists against which to judge an utterance's possible meaning in this manner.

In order to clarify what is being claimed when meanings are attributed to protolanguage, I propose we use a simple thought experiment along the lines of Quine's famous *gavagai* thought experiment (1960). The method is to imagine a field linguist charged with the task of translating the unknown language (or protolanguage) into English. The goal here is to shed light on what is being claimed when meanings are attributed to protolanguage by asking what sort of observation would be cited in (and generally accepted as) defence of such an attribution. While Quine's interest was to show that no amount of evidence — not even learning the language — could decide between radically different translation schemes, the purpose of the thought experiment here is more mundane, and will rely on the kinds of observation that would reasonably lead a field linguist to particular meaning attributions and not to others. We may ignore radical translation possibilities if they depend on kinds of glosses it would be radically unlikely for a field linguist to think up, such as *all–and–sundry–undetached–rabbit–parts* as a possible alternative to *rabbit*.

When field linguists translate languages into English, they rely on observation of the *range* of ways an expression can be used. That a particular utterance of the target language may appear to perform the same function in a particular context as a particular utterance in English is regarded as insufficient grounds to claim that they have the same meaning: other uses may diverge to a degree that makes the translation unacceptable (Everett 2005, p. 623). However, we need not stipulate that the evidence relevant to the meaning of a unit be restricted to only observations of the use of that unit. We will allow our field linguist to appeal to the uses of other protolanguage units to justify his translations. So, for example, if he claims that a protolanguage utterance is ambiguous between two meanings, he must give be able to give reasons for supposing that that utterance *could* be variously used in accordance with those meanings. The evidence for this assertion need not be actual observation of those various uses. He may, for example, back up his assertion by pointing to the ways in which other similar utterances were used and claiming a systematic relationship. Thus our thought experiment need not insist that every use the field linguist attributes to units is observed (or even actually happens), and therefore need not be overly restrictive in the degree to which protolanguage uses of an element match uses of an English element with the same meaning. But it will insist that patterns of use be observable from which it would be reasonable to deduce the protolanguage units *could* be used (by the protolanguage users) according to their attributed meanings.

Single unit utterances Jackendoff's (2002) multiword protolanguage is thought of as developing from a one-word stage, and this may be used as a starting point for thinking about the kinds of evidence a field linguist would require. A property that these single word utterances must have to be translated as single words is a degree of flexibility of use. As Jackendoff notes,

> The word *kitty* may be uttered by a baby to draw attention to a cat, to inquire about the whereabouts of the cat, to summon the cat, to remark that something resembles a cat, and so forth. Other primates' calls do not have this property. (Jackendoff 2002, p. 239)

This flexibility would be crucial for a field linguist to confidently translate symbols from a one-word stage protolanguage. Without this flexibility, it would be difficult for a field linguist to exclude a number of plausible possibilities. If a word which the field linguist thought might mean the same as *kitty* were used for only one purpose (for example, to inquire about the whereabouts of the cat) it would be difficult to justify ruling out an alternative holistic meaning (for example, the meaning of *where's the cat?*). In order for a unit to translate as *kitty* it must be used with at least some variety, and those various uses must correspond to some degree to the various modern uses of the word.

Of course, if a protolanguage is in a one-word stage, its units cannot correspond perfectly with modern words in terms of use: on the whole, modern words are not used individually. Whether this difficulty means that protolanguage units in a one-word stage simply *cannot* have meanings analogous to modern words is a decision our field linguist will have to make: the decision depends on how the notion of meaning in a one-word protolanguage is operationalised. If, for example, the protolanguage unit *elppa* is used to request apples and to report the presence of apples, one could define the meaning as the disjunction of the meanings of *can I have an apple* and *there are apples*; alternatively, the meaning could be defined as the common element (i.e. something like the set of all apples); or one could insist that for a unit of a (proto)language to have the same meaning as a word in a modern language it must play substantially the same role in both systems, and that the difference between units used alone and units used predominantly in combinations is an insurmountable barrier to sameness of meaning. Rather than ruling the one-word protolanguage with word-like meanings impossible by definition (as, perhaps, Deacon 1997 would), we may allow our field linguist to operationalise meaning in favour of word like meanings: the kinds of evidence our field linguist would require for a one-word translation would be observation that the various uses of the protolanguage unit correspond to a variety of uses of modern (multiword) utterances in which the translation word appears. While this kind of criterion may rely on the field linguist's judgement, it is clear that the evidence necessary for a one-word translation is an appropriate *variety* of use.

There are at least two ways in which a variety of use can be appropriate for a one-word translation. First, if the protolanguage unit *elppa* is to mean the same as the English word *apple*, then the objects commonly involved in situations in which *elppa* is used should be the kinds of objects which English speakers would call *apples*. If, for example, *elppa* can be used in relation to various activities involving apples and bananas, the field linguist might think *fruit* a more appropriate translation. Secondly (and more importantly) the variety of uses of the term should be appropriate to rule out a holistic interpretation (e.g. that *elppa* means only what the English *give me an apple* means).

Multi-unit meanings In Bickerton's protolanguage, and secondary stages of Jackendoff's, units are combined to form multi-unit utterances in which the order of units makes no difference to the imputed meanings (thus, "*eat apple Fred* and *eat Fred apple* might be used to convey the same message" Jackendoff 2002,

p. 247). A field linguist would have to observe a number of features in order to draw such conclusions about the identities and meanings of protolanguage units. For the field linguist to conclude that *Fred apple* (Jackendoff 2002, p. 246) was the concatenation of two symbols rather than simply one, perhaps haltingly articulated, complex symbol (whose parts are not individually meaningful), he would have to observe a range of utterances systematically related to each other. Given the non-syntactic nature of this presumed protolanguage, he might, for example, observe the parts being used in a different order (i.e. *apple Fred*) without observing a systematic difference in the ways these two utterances are used. Alternatively he might observe paradigmatic relations between utterances (e.g. *Peter apple*, *Fred banana*, *Peter banana*, etc.) whose various uses are appropriately related; for example, the differences between the range of uses of *Fred apple* and *Peter apple* shouldn't be too much more than a systematic exchange of one individual for another. If this kind of relationship is not observed, it would be difficult for the field linguist to argue that what he was observing were multi-unit utterances rather than single unit (perhaps holistic) utterances (*Fredapple*).

In terms of attributing the meanings of a compositional approach to protolanguage, the field linguist would need to observe appropriate variety in symbol use analogous to the one-word stage case. If the units of *Fred apple* could be convincingly identified by comparison with other utterances, the field linguist would still require an appropriate variety of uses of the component units. If, for example, every utterance in which the protolanguage unit *elppa* occurred was used to instruct the giving of an apple (to an individual identified by the other part of the utterance) it wouldn't be obvious that the meaning of *elppa* was more like the meaning of modern English *apple* than *give an apple to....*

The absence of syntax is taken to be a hindrance for a protolanguage in the compositional approach, forcing its users to rely on context and pragmatics for interpretation. For a field linguist to come to this conclusion he would need some reason for supposing that an utterance *could* be used in different ways corresponding to the different interpretations for which it is ambiguous. Thus, to conclude that *take Og meat* (Calvin and Bickerton 2000, p. 141) would be ambiguous as to whether it was Og that was taken to/from some meat or some meat that was taken to/from Og, our field linguist would have to justify why he thought that either of these were possibilities. For example, he would have to be able to cite evidence that *take* could be used in various utterances that related both to animates and to inanimates being *take*n. Without observations supporting these as possibilities for *take*, there would be no reason for preferring the attribution of the meaning of English *take* over some other meaning like English *take* but (e.g.) restricted to inanimates (thereby reducing the ambiguity of *take Og meat*).

It would be a difficult task to delineate further the kinds of evidence a field linguist would require in order to assign the meanings assumed by compositional approaches to protolanguage. Different meanings would require different kinds of evidence, and there may well be alternative sets of observations sufficient to assign a particular meaning. What is important is that protolanguage units be observed being used in a variety of ways appropriate to the imputed meaning. This raises the question: why would the units of a protolanguage have these modern-meaning-appropriate varieties of uses?

Spontaneous symbol use Bickerton (2002, p. 221) notes that, given the current state of our knowledge of the situations in which protolanguage(s) began, attempts to reconstruct the details will involve the telling of just-so stories. However, he does speculate that the origins or protolanguage lay in various interactions with the environment: "exchange of information gleaned in foraging, interpretation of natural signs, warnings of the young against dangers," (Bickerton 2002, p. 221), and offers three just-so stories to illustrate how symbols may have come about. While these scenarios are offered as comments on the evolutionary adaptiveness of protolanguage, they reveal that Bickerton conceives of the re-use of an extant communicative expression in novel ways (appropriate to the attribution of modern-like meaning) to be a naturally occurring spontaneous event:

> Assume that some ancestral species had warning calls that related to major predators, as Vervet alarm calls do today. Such calls (perhaps with a different inflection), if coupled with pointing at a python-track, pawprint, bloodstain, or other indication of a possible nearby predator, *could very likely have been understood* as a warning that did not require immediate reaction, but rather a heightened awareness and preparedness for action. (Bickerton 2002, pp. 220–221, emphasis added)

The idea that individuals would spontaneously produce and understand protolanguage expressions in ways appropriate to the meanings attributed by Bickerton and Jackendoff is a substantial assumption. In these theorists' writings on language evolution, it seems this assumption derives from the externalisation window inference: that contemporary neural representation of meanings can be projected back onto our ancestors and appealed to as the source of usage generalisations appropriate to synthetic protolanguage meanings.

2.2.2.4 *Representational grounding*

Both Bickerton and Jackendoff have idiosyncratic views on the ways in which meanings are represented in the heads of modern language users, and it seems to be the externalisation window that justifies for them the assumptions concerning meanings in their approaches to protolanguage.

Jackendoff's approach to reference Jackendoff characterises his approach to reference as "Pushing 'the world' into the mind" (Jackendoff 2002, title of his §10.4). His goal appears to be an account of how individuals can find referring expressions meaningful without appealing to a "magic" (p. 268) or "mystical" (p. 303) connection from the mind (which Jackendoff identifies with the brain) to the world. Precisely what Jackendoff means by a "mystical connection" is never spelled out, but it appears to be motivated by a comparison between reference and a physical connection between objects. Figure 2.2 (his figure 10.4) shows Jackendoff's schematic characterisation of the view he attributes to Fodor: "language is a mental faculty that accesses concepts [which] have a semantics; they are connected to the world in virtue of being 'intentional'," (Jackendoff 2002, p. 300). The locus of Jackendoff's difficulty with a theory that involves a relationship between the mind and the world (the wiggly arrow in figure 2.2) is that "there is no physically realisable causal connection between concepts and objects," (Jackendoff 2002, p. 300). While he never makes clear what such a causal connection would be like, it seems Jackendoff thinks of the connection as being something physical occurring *at the moment* a referring expression is understood, analogous to the light from an object of sight impinging on the retina at the moment the object is seen (Jackendoff 2002, p. 299). In the absence of such a connection, Jackendoff is drawn to the conclusion that it is not entities in the world which a referring expression connects to, but entities in the world as conceptualised by the individual. Jackendoff can be understood as insisting that there *must* be something causally connected to a referring expression in every act of reference (that is, the connection must occur as the expression is understood), and in the absence of such a connection to entities outside the mind/brain the connection must be a neurally instantiated connection within the mind/brain.

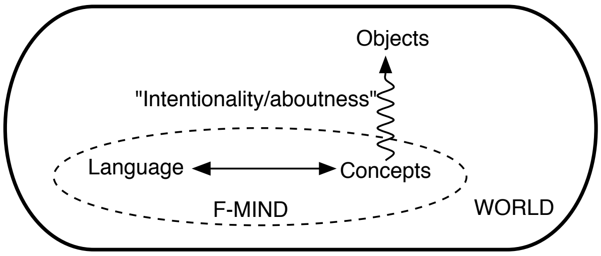


Figure 2.2: Adapted from Jackendoff's (2002) figure 10.4 showing a "mystical" connection between concepts and objects

The entities which Jackendoff sees as connected to referring expressions are not restricted to a linguistic role, but are thought of as being "the neural correlates of consciousness" (p. 310). The difference between conscious perception of something real and an imagined perception is explained by an associated variable (or set of variables) taking particular values (e.g. *external* and *non-self-produced* for perception of something in the world versus *internal* and *self-produced* for an imagined image). Similar variables associated with representations would presumably distinguish between perception of something and being told about something.

Thus Jackendoff's assumption that protolanguage units had word like meanings may be justified by assuming that protolanguage was (for some reason) also a system which linked expressions to the neural entities which underlay the consciousness of protolanguage users. Protolanguage consisted of units with the meanings of words because it utilised the same, or similar mechanisms to those by which modern words are associated with entities existing in the mind.

Bickerton's notion of meaning Like Jackendoff, Bickerton sees the meaningfulness of language as being something to do with an internal representation of the world:

> For the moment, let's just say that words represent something, somehow. They serve to focus your mind on some aspect of reality — or rather, I should say, of the picture of reality you carry about with you in you brain. (Calvin and Bickerton 2000, p. 15)

Also, like Jackendoff, this internal picture is thought of as relating to conscious experience or perception:

> When you hear the word “orange,” this may suggest to you just some vague picture (“kind of fruit”) or it may evoke the taste of an orange, or its colour (ripe or unripe), or its smell, or the texture of its skin, or — if you happen to be a fruit-grower in Italy — probably also the soft thud that an overripe orange makes when it falls and hits the ground, as well as probably lots of other things that might seem obvious to Italian fruit-growers but lie wholly outside the knowledge of you and me. (Calvin and Bickerton 2000, p. 16)

However, unlike Jackendoff, Bickerton draws a distinction between the internal representations of animals and those that underpin meaningful language. The former are involved in “on-line” thinking and exist in the “Primary Representational System” while the latter are involved in “off-line” thinking and exist in a neurally distinct secondary system.

> The level of brain structure at which the creature processes and analyses *all* its sensory input — its Primary Representational System or PRS (Bickerton 1990) — hooks up directly with output systems. Or, as the saying has it, “Monkey see, monkey do.” But if humans see, they may stop, think things over, and maybe do something later, maybe not. (Bickerton 1995, p. 56)

Because the protolanguage Bickerton imagines would have been used for the exchange of information (rather than the immediate manipulation of another’s behaviour), it would also have relied on the secondary representational system: “such representations must be kept free of direct contact with motor centers, if inappropriate reactions to words (reactions of the type appropriate to calls) are to be avoided,” (Bickerton 1995, p. 59). Bickerton speculates that this secondary system might have predated the emergence of protolanguage, perhaps as a consequence of phylogenetic increases in brain size triggered by some other cause:

> Once this development [the expansion of the brain] had begun, the new areas might have been “colonized” by ensembles of cells converting adjacent cells into functional copies of themselves along lines suggested, very plausibly, by Calvin (1993). This could conceivably have created multiple representations of whatever was already in the PRS. Moreover, since these new areas would not yet have established behavior-triggering links with motor regions, the representations

> they contained could have provided both the materials and the kind of environment required for off-line thinking. Protolanguage would then have developed as soon as links were formed between the new areas and those that controlled the vocal organs. (Bickerton 1995, pp. 59–60)

While this is not the only possible evolutionary scenario envisaged by Bickerton for the origins of a secondary representational system (nor even the one he favours[2]), the above quote displays his commitment to protolanguage being the externalisation of a representational system whose representational units' identities determine the meanings of protolanguage units. The mechanism which links vocal symbols to representations in the secondary system in modern humans is thought of as being responsible for the meanings of protolanguage units: the reason why protolanguage units would have been used in such a way as to warrant the glosses Bickerton offers is their attachment to off-line representations. Thus, for example, the reason why the use of an alarm call would be spontaneously extended as Bickerton (2002) describes (see section 2.2.2.3 above) is because protolanguage users would have associated the call with the off-line representation.

2.2.3 Holistic protolanguage

The holophrasis approach to protolanguage envisages a different kind of meaning for protolanguage units than the synthetic approach. Rather than utterances of protolanguage consisting of combinations of units with word-like meanings, the holophrasis approach hypothesises that protolanguage utterances consisted of single units with semantics corresponding to whole modern sentences/utterances.

The most detailed defence of the holophrasis approach is given by Alison Wray (1998, 2000, 2002a; Wray and Grace 2007). In part, the motivation for a holistic protolanguage derives from an alternative view of the end-state of language evolution (i.e. modern language). While Bickerton and Jackendoff are concerned to offer accounts of an evolutionary process whose end-state is a system which syntactically combines words, Wray (2002a) sees such an "analytic strategy" as

[2] Bickerton does not favour this *evolutionary* scenario which dissociates protolanguage from the selection pressure for a larger brain. But what he finds wrong with it is the notion that the secondary representational system was selected for reasons other than the use of protolanguage (e.g. the secondary system was a consequence of brain expansion which developed because the brain's venous system serves as a radiator), not that it assumes protolanguage units' meanings were determined by externalisation.

only one component of the processing of modern language. In addition, modern language users exploit a holistic strategy which "takes from the lexicon prefabricated strings of morphemes, words, phrases, clauses, or whole sentences — even whole texts" (Wray 2002a, pp. 113–114). Wray (2000, 2002a) sees the communicative functions of these modern formulaic sequences as being analogous to the functions of chimpanzee cries as described by Reiss (1989). This continuity in function leads Wray to the conclusion that the modern holistic strategy and chimpanzee cries are homologous, sharing a common evolutionary root. Thus, the holistic strategy would be available during all periods of language evolution.

Like the compositionality approach to protolanguage, Wray's protolanguage differs from animal signals in terms of iconicity and gradience (in Bickerton's senses). The point in the evolution of language at which Wray's protolanguage comes is after the development of arbitrary phonetic form. Like Bickerton, Wray (2000, p. 293) identifies the greater number of discrete (and distinguishable) units arbitrary phonetic form affords as the selective advantage driving the (biological) evolution of these forms. The externalisation window is an inference which may be appealed to to explain the expansion of this communicative system.

2.2.3.1 *Meanings in holistic protolanguage*

The similarity between modern formulaic sequences and chimpanzee cries Wray identifies are that "in both species they are used for social interaction, where their purpose is the manipulation of the hearer, either to act in the interests of, or to recognise the identity and status of the speakers" (Wray 2000, p. 289). The functions of chimp cries are seen as a subset of the functions of formulaic sequences. This focus on social function carries through into Wray's description of protolanguage. Wray (2002a, p. 117) contrasts this kind of manipulative message with what she sees is the function of Bickerton's protolanguage, namely the exchange of referential information. In spite of this difference, the units of Wray's protolanguage are described as having meanings which, when glossed in English, contain similar kinds of referential terms to the glosses of Bickerton's protolanguage.

This focus on the social function of protolanguage units doesn't seem to constrain the meanings Wray is prepared to associate with units. She (2002a, p. 120) states that, "in principle, you can have a holistic message with any meaning you like, from 'come here' to 'five minutes ago I saw a buck rabbit behind the stone at the top of the hill'." While Wray does offer reasons for thinking the former is more likely than the latter (the infrequency with which the latter is used would increase the likelihood that the community would forget its meaning) this quotation suggests that Wray does not see the holistic nature of the form of protolanguage utterances being an impediment to their range of possible meanings. If, in principle a holistic message can have "any meaning you like" and protolanguage utterances consist of a single holistic unit, it seems that Wray equates the meaning of holistic utterances with the meanings of modern utterances. Thus protolanguage units may convey semantically complex messages where "[a] semantically complex message is one that, despite having no internal structural composition itself, would require several words and some grammar to translate into a language like English,"[3] (Wray and Grace 2007, p. 569).

This semantic complexity strains the link between holistic protolanguage and chimpanzee cries. The essay by Reiss (1989) which Wray (2000, p. 289; 2002, p. 113) refers to in support of her claim that chimpanzee calls fulfil similar functions as formulaic utterances is concerned with the analysis of chimpanzee calls within a revised version of Searle's (1979, 1983) taxonomy of speech acts. Reiss's concern is to show that the intentions of chimpanzee communicative acts can be analysed in the same terms as some uses of human language (whether formulaic or not). The analysis is in terms of the different kinds of effect a chimpanzee communicative act can have on its audience (e.g. whether the effect of the act is to make the addressee expect that the addressor will do something in the future or whether it is to make the addressee do something themselves). This kind of feature of speech acts appears to be orthogonal to the kinds of meanings Wray attributes to holistic utterances:

> "What a nice day" may be an expressive thanking (expression of gratitude) for what is presupposed to have been a nice day (in some way offered by the hearer to the speaker), or it may have been an assertive speech act whose sincerity condition is the speaker's belief that it is a certain kind of day. . . (Reiss 1989, p. 291)

[3] This is a strange criterion given that Wray sees continuity between holistic protolanguage and formulaic sequences. There seems to be no reason why a modern language could not have a holistic unit whose meaning is identical to the meaning Wray attributes to a given protolanguage unit. Thus if protolanguage *tebima* means what English *give that to her* means, there seems to be no barrier to imagining a modern language identical to English with the exception that it includes the formulaic utterance *tebima*. Translation of protolanguage *tebima* into this imagined relative of English would produce a single word and hence lead to the conclusion that protolanguage *tebima* did not convey a semantically complex message.

The meanings of chimpanzee cries (in terms of conventional contents) are impoverished in comparison to the meanings Wray attributes to her protolanguage units. None of the examples listed by Reiss, for example, involve a third party.

It is worth considering what Wray's attribution of complex semantics to the utterances of a holistic protolanguage amount to. If we consider again a time travelling field linguist confronted with Wray's protolanguage, we will again find that if evidence for Wray's glosses is to be convincing it will have to cover a variety of uses. Consider *pubatu* which she glosses as 'help her' (2000, p. 294). Suppose the protolanguage users of this this unit were fond of climbing trees, but sometimes females got tangled up in the branches. If these kinds of situations were the ones in which the field linguist observed *pubatu* being used with the semi-reliable consequence that the addressee helped the stuck individual, the field linguist might gloss it as 'help her'. However, other equally valid glosses would be available with no obvious way of choosing between them: *pubatu* could be glossed as 'get her down' or 'she's in trouble' or 'shake that branch' (if this was the most common way in which help was given).

Observation of various uses (gestures semi-reliably leading to the addressee helping a female third party in a variety of situations) would rule out some glosses. However, some aspects of the meaning attributed to the holistic unit would still raise difficulties. What kinds of evidence would lead the field linguist to conclude that *pubatu* specified the sex of the to–be–helped third party? Suppose the field linguist made observations that led him to conclude that no individual would ever want another individual to help a third if the third individual were male (perhaps males never need help). Should *pubatu* be glossed as 'help her' or 'help that one' (with meaning-external factors accounting for the fact that it is only ever used when the addressor knows that the one–to–be–helped is female)? One sort of evidence that could make up the field linguist's mind on this issue would be a contrasting holistic unit whose function seemed close enough to *pubatu* with the difference that the one–to–be–helped was male. That is, one source of evidence that could be used to justify the specificity of Wray's glosses would be systematic relationships between holistic units.

The specificity and complexity of the meanings Wray attributes to protolanguage implies complex patterns of usage (and, perhaps, expressions whose usage patterns are systematically related to each other). One way the development of these meanings could be accounted for is by appeal to the externalisation window.

It may be objected that usage patterns appropriate for a field linguist to light on a specific gloss to the exclusion of other possibilities are not necessary to the idea of a holistic protolanguage, and that to gloss a holistic protolanguage utterance as e.g. 'help her' does not have to imply that the use of that utterance is such that other glosses would not be equally appropriate. However, as we will see, Wray's account of the development from holistic protolanguage to a system which uses recombinable meaningful parts relies crucially on utterances having the meanings she attributes to them.

2.2.3.2 Fractionation

In various publications, Wray suggests that the "analytic strategy" developed from a holistic system by the decomposition of holistic units on the basis of chance correspondences between sub-parts of form and sub-parts of meaning. Wray (1998, pp. 55–57) gives the following as examples of holistic protolanguage utterances:

/mɛbita/	*give her the food*
/ikatubɛ/	*give me the food*
/kamɛti/	*give her the stone*

The phonetic forms of these holistic utterances are arbitrary and any relations between them are supposed to have come about by accident. By chance this holistic protolanguage has a number of regularities: /mɛbita/ and /kamɛti/ both contain the syllable /mɛ/ in their forms and a singular female recipient in their meanings. Wray (1998, pp. 55–56) suggests that this would lead individuals to create a "morpheme boundary" around /mɛ/, and to give it the meaning of *her*. Likewise, /ikatubɛ/ and /kamɛti/ both contain the syllable /ka/ in their form and their meanings share "the idea represented in English by the word *give*," leading individuals to associate this syllable and meaning Wray (1998, p. 56). The consequences of these associations is that individuals then go on to use form/meaning pairings in *other* utterances. For example, the association between /ka/ and 'give' could lead an individual to "hypercorrect" /mɛbita/ (whose meaning also contains 'give') to /mɛbika/ (ibid., p. 56).

The driving force behind this process is internal representations of meaning. While the external forms of holistic utterances are not built up of units each bearing parts of the meaning, Wray considers holistic utterances as "mapping" onto semantic representations that have parts. To illustrate, Wray (2002a) compares

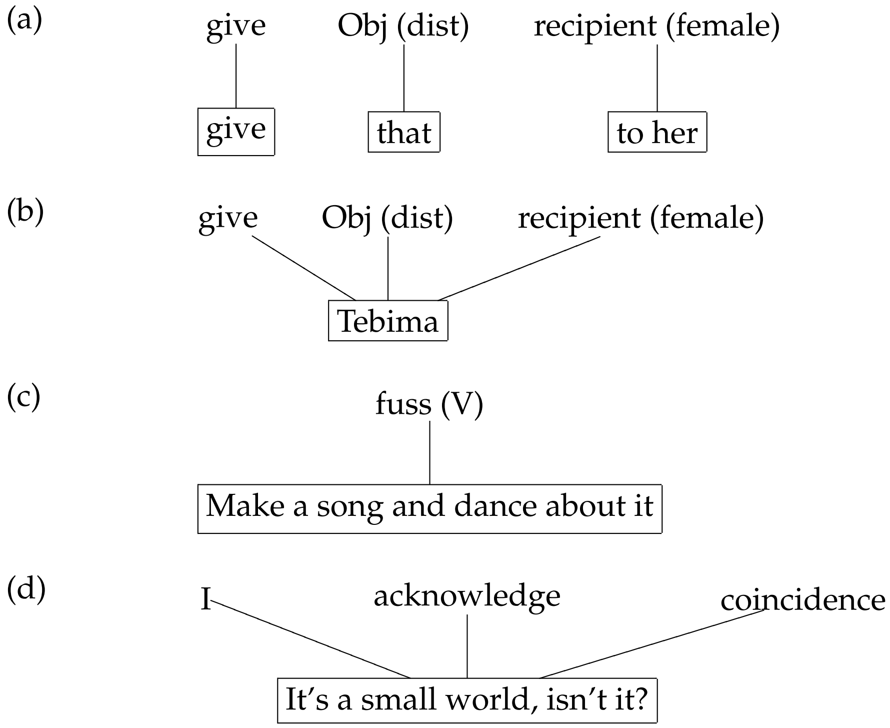


Figure 2.3: The relationship between semantic representation and utterances illustrated by Wray (2002a, p. 119, her Fig. 6.1)

the "semantic representation" associated with English *give that to her*, with the "semantic representation" associated with an imagined protolanguage utterance *tebima*, (Figs. 2.3a and 2.3b respectively; 2.3c and 2.3d are offered by Wray to illustrate the idea that the forms of modern formulaic utterances do not correspond to the forms of their associated internal representations). The justification for the assumptions that holistic utterances could have "any meaning you like", and that individuals would identify and re-use parts of expressions in ways appropriate to the meaningful parts of modern utterances derive from her view that protolanguage utterances would have been the externalisation of internal structures which can be inferred via the externalisation window.

The fractionation process thus makes similar assumptions to those made when the units in the compositionality approach to protolanguage are assumed to have the meanings of words: protolanguage users generalise the uses of symbols in ways appropriate to modern symbol uses in virtue of the constitution of their internal representations.

To further emphasise this point, consider the identification of the syllable /ka/ with "the idea represented in English by the word *give*." The English word *give* can be used with various different recipients. The holistic utterances from which /ka/ is derived have different recipients (/ikatubɛ/: first person recipient; /kamɛti/: second person female recipient). Why would a protolanguage user assume that their protolanguage ought to have a word that can be used when the speaker is to be the recipient and when a female third party is to be the recipient? To those of us who have learned a language in which this kind of generalisation is common, such a feature seems natural. The fractionation process assumes that it would be just as natural for protolanguage users without such structure already present in their protolanguage. This assumption rests on the inference of the externalisation window.

Protolanguage summary Holistic and synthetic protolanguage theories can be seen as alternative conceptions of what language–without–syntax would be like. In considering protolanguage as an asyntactical version of modern language, defenders of holistic and synthetic protolanguage attribute modern–language–like meanings to protolanguage. These meaning attributions are crucial to the processes by which protolanguage develops into language, and in both cases the justification for meaning attributions can be found in the projection of modern internal representations back into the brains of our ancestors.

2.2.4 Computational simulations

A number of computational models of language evolution rely explicitly on the externalisation of internal structures. For example, models of the evolution of vocabulary (Smith 2004) and grammatical structures (Batali 1998; Kirby 2002) can be seen as models of the cultural evolution of shared systems of form/meaning mappings. These simulations consist of agents who associate external signals with internal representations of meaning whose identities are stipulated to correspond to the meanings of expressions in modern languages (most commonly, English words). Through communication, agents learn to associate the same external forms with the same internal meaning representations. The tacit assumption of these simulations is that throughout the evolution of language, expressions would have modern-like meanings because of their relationships with internal representations of meaning whose existence may be inferred via the externalisation window.

Some other computational models (e.g. Vogt 2003; Steels and Kaplan 2002) share a commitment to externalisation, but with a more flexible internal component. In these models, the internal representations which are externalised relate to a number of perceptual variables: representations are ranges of values these perceptual variables take when an object is perceived. These representations may be refined through experience of the world and experience of other agents' use of language. While these latter kinds of model show a greater degree of flexibility, the assumption of externalisation is still a powerful factor in their success. While certain details of internal representations may develop through experience with language, these models nonetheless rely on the externalisation of specific internal structures to imbue expressions with particular kinds of meaning. That an expression in such a model is, for example, used as the name of a colour is determined by its being associated with internal representations of colour. The structure of the externalisation mechanisms means that these agents are forced to, and can only have communication systems which consist of a restricted set of concepts.

These computational simulations assume a communicational scenario (the language game) driven by externalisation (of entities given in some cases and of learned regions on perceptual scales others). In fact, the force of these agents' impulses to externalise their internal representations is so strong that they attempt to do it even in the absence of any communicational success.

2.3 Summary

This chapter has sought flesh out in behavioural terms the attributions of meaning to earlier stages in the evolution of language by considering what observations could lead a field linguist to make those attributions. In doing this, substantial (sometimes tacit) assumptions made by language evolution theorists about the uses of expressions throughout language evolution were revealed. The externalisation window can (and often is) used to justify these assumptions about meaning in language evolution theories by, for example, backing up assertions about the spontaneous creative use of expressions by individuals. The externalisation window has the property of pertinence in the context of these theories of language evolution: in the case of computer simulations, externalisation of internal representations is the only way in which meaning features in the accounts offered; in theories of protolanguage, externalisation of the representations of the meanings of modern expressions is crucial not only to the characterisation of meaning in protolanguage, but also to the processes by which protolanguage is conceived of as developing into a more modern-language like system. If the externalisation window lacks the properties of warrantedness and groundedness, these theories of language evolution rest on unfounded assumptions. The next chapter asks whether there is any empirical evidence to support the warrantedness of the externalisation window, and the window's groundedness will be questioned in chapter 4.

CHAPTER 3

Warrantedness of the externalisation window

Through the externalisation window, the meanings of words of modern languages are projected onto earlier stages of the evolution of language on the basis that a word has meaning in virtue of being the external expression of an internal mental representation. This chapter considers whether the window has the property of warrantedness; that is, whether the assumptions about early meaning embodied in language evolution theories can be justified by assuming that the relationships between internal neural structures and external expressions present in modern language-proficient humans can safely be assumed to have also been present earlier in the evolution of languages. The warrantedness of the externalisation window is approached from three directions. Section 3.1 considers the state of our knowledge about the neural structures related to meaning, and whether it is plausible to assume that usage patterns assumed by language evolution theories would plausibly spring forth spontaneously from analogous structures in our distant ancestors brains. Section 3.2 considers the possibility that the human brain may be innately endowed with special externalising machinery that could be assumed present at earlier stages in the evolution of language; this possibility is approached obliquely by considering the attribution of innate functions to the classical language areas. Finally, section 3.3 considers whether the patterns of development shown by children learning their first language support the view that humans spontaneously produce the kinds of usage generalisations discussed in chapter 2.

3.1 The structure of internal representations

While many theories of language evolution (and many theories of language, Chomsky 1995; Croft and Cruse 2004, etc.) appeal to the notion of internal representation to explain meaningful language, the neural instantiation of such representation is rarely touched on. There is a danger that notations for internal representations disguise the complexity of these internal structures and lull theories into a false sense that attaching labels to internal structures is a straightforward matter. Additionally, as Saunders and van Brakel (1997, p. 173) note, "the cautions and hesitancies of the neurophysiologist are frequently lost when adjacent disciplines adopt their findings." This section, therefore, touches on a number of issues concerning the structure of the neural underpinnings of meaningful language and tries to give a fair reflection of the degree to which there is agreement concerning these structures.

In the neuroscientific literature, the neural structures thought to underpin meaningful language use are referred to as "semantic memory". This was defined by Warrington (1975) as "that system which processes, stores and retrieves information about the meaning of words, concepts and facts." Humphreys and Forde (2001, p. 455) note that "despite the general use of the term, or perhaps because of it, there has been little attempt to provide a more rigorous definition." It is notable that the role of semantic memory in language is definitional rather than empirical: "semantic memory" is not the label for a neural system which has been *found* to process, store and retrieve certain kinds of information; it is, rather, an aspect of the way neuroscientific investigations into meaning and language are structured. As such, the definition of semantic memory as a "system" cannot be used to infer that the same system would have been present at earlier stages in the evolution of language and performing a particular role (specifically, producing expressions whose usage patterns fit the meaning ascriptions discussed in chapter 2). "Semantic memory" reflects the same externalisation assumption about modern language as language evolution theories, but have the investigations structured around "semantic memory" produced evidence that would warrant the externalisation window?

3.1.1 Category specific disorders

A particularly fertile ground for evidence and theories of semantic memory organisation has been the phenomenon of category specific disorders (Caramazza and Mahon 2006). Category specific (CS) impairments are impairments to an individual's ability to perform tasks when items from certain categories but not others are involved. Although many cases of such CS deficits have been reported (Capitani et al. 2003), the first published description of such disorders appeared only about 60 years ago (Forde and Humphreys 1999). Nielsen (1946) reported the cases of two patients: the first, Flora D., presented with visual agnosia for animate objects but not for inanimate objects while the second patient, CHC, showed the opposite pattern, displaying visual agnosia for inanimate objects but not for animate objects. Nielsen used the fact that these two patients had different patterns of neurological damage to project the recognition of animate objects and inanimate objects onto the sites of Flora D.'s and CHC's damage respectively.

As more cases of CS impairments were reported, the picture became increasingly complicated. The first systematic investigation of a patient with selective semantic impairment Warrington and Shallice (1984) showed what appeared to be a deficit affecting living but not non-living things. However, when the range of categories tested was extended, the patient (JBR) was found to also perform poorly with types of cloth, musical instruments, precious stones and food items. The phenomenon of 'damage to inanimate objects' also appeared to be more complex than Nielsen's (1946) cases suggested: Warrington and McCarthy (1987) reported a patient (YOT) who appeared to have selective impairment *within* the domain of inanimate objects. YOT was more impaired (on a spoken–word–to–picture matching task) for small manipulable things than for larger outdoor objects like houses, ships, bridges, etc. (Note, however, that whether this is the appropriate interpretation of YOT's performance is controversial, Capitani et al. 2003).

Category specific impairments are a principle source of evidence for the structure of semantic memory (Caramazza and Mahon 2006). For example, if different category impairments are regularly associated with different kinds of neural damage, then categories can be correlated with neural structures (at least, with neural structures essential for semantic memory of those categories). Indeed, an account of category specific deficits is generally seen as a desirable feature of a good description of the structure of semantic memory (Humphreys and Forde 2001), and failure to account for patterns of deficit is a commonly used criticism of particular theories of semantic memory (e.g. Caramazza and Shelton 1998; Cree and McRae 2003). Therefore, CS semantic disorders are one possible source of evidence for the nature of the internal representations on which the warrantedness of the assumptions about meaning discussed in chapter 2 could be based.

3.1.1.1 *What are the categories in category specific impairments?*

If CS impairments show consistent patterns over categories, this would suggest the semantic memory representations within these categories share certain properties at a neural level, and that this is consistent across individuals. This could be a first step in a chain of argumentation that based the warrant for the externalisation window on the fact that semantic representations are coherent neural entities. However, the categories of CS impairments are a matter of some controversy.

Cree and McRae (2003) analysed ten studies which, in their opinion, covered the range of deficits typically reported, and which reported the items used in testing. They found the following eight trends in the patterns of CS semantic deficit:

1. Creature categories (including mammals, birds and reptiles) pattern together and can be separately impaired.
2. Nonliving-things pattern together and can be separately impaired, though these exclude musical instruments and food.
3. Fruits/vegetables pattern together and can be separately impaired.
4. Fruits/vegetables can pattern with either creatures or nonliving things.
5. Nonliving foods can be impaired along with living things.
6. Musical instruments can be impaired along with living things.
7. Living things deficits are more frequent than nonliving things deficits.
8. Body parts can be impaired with nonliving things.[1]

The categories of musical instruments and body parts, and their potential to associate with nonliving and living things respectively appear to contradict a simple interpretation of the phenomena along the lines of natural kinds. However, there is disagreement as to whether this interpretation of the data on musical instruments and body parts is correct. Capitani, Laiacona, Mahon, and Caramazza (2003) reviewed the clinical evidence (published by 2001) on category specific semantic impairments. With respect to paradoxical associations with *musical instruments* and with *body parts*, they found the evidence unconvincing. Patients have been reported whose impairments support these associations, but there are also a number of patients whose impairments contradicted the associations (e.g. impaired performance with living things but spared performance with musical

[1] In Cree and McRae's (2003) study, this eighth trend was not addressed as they didn't collect data relevant to it.

instruments). Capitani et al. (2003) suggest that factors other than the organisation of semantic memory may account for the fact that some patients show poor performance on *musical instruments* while having a category specific impairment for *living things*. This conclusion is supported by Barbarotto, Capitani, and Laiacona (2001) who tested a number of patients suffering a variety of brain insults commonly associated with lexical-semantic impairment, and found that *musical instruments* generally produced poorest performance while *body parts* produced the best performance. However, items within each category varied in terms of certain features which may be taken as indications of how experienced people in general are with particular items (e.g. age of acquisition and lexical frequency of items). When these variables were controlled for, neither *musical instruments* nor *body parts* emerged as particularly impaired or spared categories. Therefore, it is *possible* that the observed poor performance with *musical instruments*, and relatively spared performance with *body parts*, seen by patients who show *living things* deficits, result from unmatched test items. The possibility that *some* reported CS phenomena could result from poorly controlled stimuli will be returned to in section 3.1.1.4.

A limitation of the work on category specificity is that it commonly focuses on the same categories, and this might be part of the reason why certain categories appear as impaired in many reports. Support for this notion comes from investigations which test previously ignored domains. On the basis of their theoretical account of category specific impairments, Borgo and Shallice (2003) predicted that the semantic domain of *mass kinds* (categories whose exemplars lack consistent form such as such as materials, edible substances and drinks) should be impaired along with *living things*. Consistent with their theoretical predictions, the patient they reported was impaired for this novel domain as well as for living things, but had relatively spared performance on the domain of artifacts. (However, Carroll and Garrard (2005) found that not all patients with impairments for living things showed an impairment on this newly identified domain.)

While there is general agreement that living things deficits are more common than non-living things deficits (though see section 3.1.1.2 below), and that animate living things can be impaired separately from inanimate living things, precisely what the categories of CS impairments are, and particularly whether they can be confined to strict animate/living/nonliving boundaries are controversial matters (Rogers and Plaut 2002). That such disagreement exists should sound a note of caution: specific categories do not clearly and uncontroversially emerge from studies of CS impairments. It is possible that the categories animate/living/nonliving are artefacts of the way patients are tested and CS impairments are reported (as it is possible that only those CS impairments that seem to conform to these categories are detected and reported).

3.1.1.2 *How is category specificity defined?*

Commonly, category specificity has been determined by a within patient analysis: comparing a patient's test scores for a given category with their scores on a different category (i.e. lower scores with living things being interpreted as a living things deficit). The problem with this method is that performance is defined relative to the test rather than relative to a measure of normal performance (that is, the cause of poorer performance with certain categories could be something to do with the test items rather than the patient). Laws (2005) suggests that a better criterion for category specific deficits is poor performance *relative to matched control performance*, combined with a difference in performance across categories greater than would be expected on the basis of category differences seen for normal controls. The importance of choosing an appropriate criterion for category specificity is highlighted by a study by Laws, Gale, Leeson, and Crawford (2005) which compared different methods of assessing data for the presence of CS deficits and found that different methods produced striking differences in results. Most notably, the methods Laws (2005) argues against produced higher estimates of the prevalence of living things impairments than nonliving things, while the method favoured by Laws identified only nonliving things impairments.

The difference between within patient analyses and Laws' method can be understood by inspecting figure 3.1. Naming accuracy (x-axis) is plotted against a hypothetical cumulative distribution (that is, the proportion of the population which is at or below the corresponding naming accuracy, y-axis). The performance of this hypothetical normal population (on a hypothetical set of stimuli) is on average better for living things than artefacts, though the range of performance in the normal population is greater for living things than nonliving things. Consider a patient who can name approximately 40% of items from both categories: comparing their scores for the two categories does not indicate a CS impairment, comparing their performance with population averages shows a greater impairment for living things, and comparing their performance

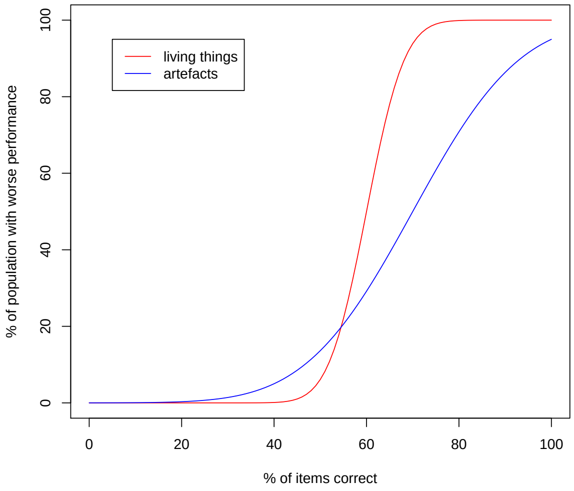


Figure 3.1: Hypothetical distribution of naming ability in normal population

with the distribution of normal performance suggests an impairment for artefacts (about one person in twenty would have the same or worse performance on living things while only about one person in one thousand would how such performance on nonliving things). The latter comparison is that suggested by Laws as the most appropriate.

Thus there is a danger that patterns of results across studies (particularly the higher rate of living things deficits than nonliving) may be artefacts of inappropriate methods of analysis. However, there is some dispute as to whether Laws' is the appropriate analysis. Capitani and Laiacona (2005) disagree that control data should be used as Laws suggests (i.e. to estimate what proportion of the population would show worse performance than a subject) and instead favour analyses which use control data to measure the difficulty of test items. These difficulty measures are then entered as a model variable in a logistic regression analysis. These two analyses seem to have different strengths and weaknesses: Laws' analysis takes into account normal variability amongst controls but ignores effects that may arise differences in normal performance associated with particular items while Capitani and Laiacona's method accounts for stimulus differences but not normal population variability.

Again, the confusion and controversy here should serve as a warning: damage to the brain does not produce striking unequivocal CS effects (if such effects were present then we might expect different plausible methods of detecting them to yield the same results). These complications leave open the possibility that "semantic memory" is far more complex than the picture of individuable representations, grouped by category, would suggest.

3.1.1.3 What abilities are impaired by damage to semantic memory?

The definition of "semantic memory" implies that individuals with "semantic deficit" should show impaired performance on a *variety* of tasks which rely on the ability to "process, store and retrieve information about the meaning of words, concepts and facts". Tasks which patients classified as suffering semantic deficits fail on include naming pictures, naming to descriptions, recognizing animals from their characteristic sounds, distinguishing between pictures of real or unreal objects (object decision) and judging the relative sizes of objects (Caramazza and Shelton 1998). In contrast, patients who, for example, only have difficulty on tasks involving visual recognition are classified as having visual agnosia (which can also be category specific, Dixon et al. 2002). However, a patient need not show impairments on *all* tasks to be classified as having a semantic impairment. For example, Caramazza and Shelton (1998) grant that patients with CS impairments may perform within normal limits on an object decision task.

Laws and Sartori (2005) note that there do not exist agreed explicit criteria for establishing CS semantic deficits. They state that the one source of evidence consistently taken as indicating CS semantic deficit is poor picture naming, and in some cases this is the *only* source of evidence on which a claim of CS impairment is made. Damasio et al. (2004, p. 217) also complain that semantic deficits are often inferred on the basis of poor performance on just one kind of task.

Laws and Sartori (2005) report patterns of deficit that highlight the problems associated with such inferences. They reported patients with paradoxical deficits: two patients showed a living-things deficit when tested on a feature verification task but a nonliving-things deficit when tested with a picture naming task; and one patient showed a living-things deficit for feature verification but a nonliving-things deficit when naming to description. While Laws and Sartori (2005) note that they didn't retest these patients (so the results may not be reliable), these data serve to highlight a difficulty in drawing conclusions from reports of CS semantic deficits in the absence of an agreed set of tests to define a CS semantic deficit. There is a danger that more complex deficits (affecting performance on different tasks in different ways) have been missed in the CS literature, and the appearance that brain damage has resulted in deficient semantic memory is, in some cases, an artefact of investigators' assumptions.

This possibility has important implications for the reliance on internal representations to produce appropriate usage generalisations: if brain damage produces CS impairments for some tasks and not others, the assumption that there exist neural structures which can be relied upon to produce meaning-appropriate usage generalisations is questionable.

Imaging results support a dissociation between neural structures which, when damaged, produce CS effects specific to different tasks. Two studies (Zahn et al. 2006; Brambati et al. 2006) both correlated CS performance with brain damage location and found quite different results. Brambati et al. (2006) correlated performance on a picture naming task with grey matter volume, and found an association between naming living things and the medial portion of the right anterior temporal lobe. Difficulties naming nonliving things were associated with damage to the left posterior middle temporal gyrus. Brambati et al. (2006) interpreted their results in terms of semantic processing. In contrast, Zahn et al. (2006) correlated damage with performance on a semantic verification task (in which subjects answer true/false questions such as "a zebra has stripes") and found impaired visual property verification for living things was associated with damage to the left posterior fusiform gyrus while functional property verification for living things showed no association with any particular brain region. Visual and functional property verification for nonliving things correlated with overlapping areas in left anterior temporal and bilateral premotor areas. Thus although both studies interpreted their findings in terms of semantic representations, the differences between tasks produced strikingly different results (e.g. different *hemispheres* implicated in the representation of living things). Similarly, Damasio et al. (2004) found different lesion sites to be associated with CS naming and recognition deficits.

The effects of neural damage may be interpreted in various ways (Hughlings-Jackson 1878; Pulvermüller 2002). The findings of Brambati et al. (2006), Zahn et al. (2006) and Damasio et al. (2004) do not necessarily have to be interpreted in terms of damage to semantic representations (some damage may, for example, be thought of as producing effects "downstream" of semantic representations). However, the neuroimaging results do not *support* the conclusion that unitary semantic representations exist.

3.1.1.4 *Effects of premorbid experience*

Lambon Ralph et al. (2003) suggest that at least some cases of CS impairment may simply reflect levels of premorbid knowledge of the 'impaired' and 'spared' categories. Such variation does exist in control subjects' knowledge of animate, plant and artifact domains (Funnell and De Mornay-Davies 1996). Dixon et al. (2002) report the case of ELM who suffered prosopagnosia and CS visual agnosia (poor performance on biological items). Prior to his stroke, ELM had been a bugle player in a military band, and when tested on picture–word matching and attribute listing tasks with musical instruments, showed a deficit pattern mirroring his experience of instruments (stringed instruments impaired, brass instruments spared).

Premorbid semantic memory abilities are difficult to assess simply because individuals only come to investigators' attention after suffering damage to the brain. In some cases, attempts are made to control for levels of experience by using population average measures (e.g. word frequency, mean age of acquisition, mean familiarity rating) as proxies for measures of the individuals' experience of the categories being tested (e.g. Albanese, Capitani, Barbarotto, and Laiacona 2000). These are somewhat unsatisfactory as individuals suffering from apparent CS semantic deficits are by definition different to the ordinary population: these differences *may* be rooted in selective damage to semantic memory, or they *may* reflect unusual premorbid experience (in which case population average measures would be inappropriate to control for the subject's experience). Caramazza and Shelton (1998) report the case of EW who showed a CS deficit for animals across a wide variety of tasks, but the difficulty in assessing whether this was simply a reflection of premorbid relative ignorance about animals is apparent in Caramazza and Shelton's description of EW's experience:

> EW reported that, "I'm not good with animals. They always give me trouble." Later, we asked whether she had problems with animals because she never had much experience of them. She responded, "Well, I did know some things about animals. I used to take my kids to the zoo and such. It's worse since I had my stroke. Now I have no idea about animals at all. They [her speech therapists] used to give me pictures and I told them I couldn't do animals. I can't even think too much about what animals look like" (Caramazza and Shelton 1998, p. 24)

It is difficult to reach any conclusion about EW's premorbid experience on this basis (especially since it is possible to have relatively limited knowledge without realising).

However, premorbid experience cannot simply account for all cases of CS impairment. For example, Michelangelo, a patient who suffered CS impairments for animals on a visual recognition task and on perceptual questions, was a member of the World Wildlife Fund and could identify large numbers of animals before his brain damage (Humphreys and Forde 2001).[2] In cases where premorbid experience is difficult to assess, spared performance with a particular category on some tasks and CS impairments on others may be used to argue that the category specificity is not a reflection of premorbid experience, but as noted above, these cases do not suggest that internal representations have the coherence required to drive the usage generalisations of chapter 2.

3.1.2 *Distributed, overlapping representations*

A number of theoretical models have been developed to account for CS deficits (see Forde and Humphreys 1999 and Caramazza and Mahon 2006 for reviews). A common feature of these models is that semantic representations are distributed over a number of neural structures, and different semantic representations overlap in terms of the neural structures involved (this is true even for the Domain Specific Knowledge hypothesis according to which the brain has category organisation hardwired, Mahon and Caramazza 2003).

Investigations into CS activity in normal subjects have identified different brain regions associated with different categories (Martin et al. 1996; Chao et al. 1999; Ishai et al. 1999). However, Devlin et al. (2002) criticised some of these studies on a number of methodological grounds, and noted the large degree of inter-subject variability in imaging results. Devlin et al. performed a meta-analysis of 9 studies using Positron Emission Tomography (PET) to look for an anatomical basis for differences between natural kinds and artifacts, and noted that there was little consistency across studies. For example, sixteen of the regions associated with domain differences were found only in single studies, and one region was associated with natural kinds in one study and artifacts in another. The only finding Devlin et al. (2002) found to be consistent across studies (appearing in seven out of nine) was a region of the left posterior middle temporal gyrus which responds

[2] N.B., Capitani et al. (2003) reanalysed Michelangelo's data and concluded that he did not display an animals deficit on associative questions.

preferentially to tools. Devlin et al. (2002) suggest that the inconsistent reports of category specific activation may reflect type I errors (false positives) resulting from failures to correct statistical tests for multiple comparisons. Devlin et al. combined the results of their meta-analysis with their own imaging results, and concluded that these experiments identified a distributed neural system common to both natural kinds and artefacts.

Striking evidence for the overlapping nature of neural activity interpreted as semantic representations comes from a study by Ishai, Ungerleider, Martin, Schouten, and Haxby (1999) which compared activity in posterior ventral temporal cortex in response to pictures of houses, faces and chairs. The locations of peak activation for houses and faces were consistent with other studies, while peak activity in response to chairs did not fall in the region more responsive to houses than faces (also, more responsive to tools than animals) as some theories would predict on the basis that houses and chairs are both artefacts. Instead the peak activity response to chairs was situated in such a way that the peak response to faces was *between* chairs and houses. Ishai et al. also found that each category provoked significant levels of activity in regions that responded maximally to stimuli from the other two categories. Ishai et al. compared activity during passive viewing and a delayed matching task to investigate the effects of increased attentional load. They found that attentional modulations for a given category were not confined to the region that responded maximally to that category. For example, the effect of increased attention when dealing with houses was greatest in the region that responded maximally to chairs. Ishai et al. conclude that representations of different categories are distributed over overlapping neural assemblies, and that increasing attentional demands do not produce uniform increases in activity over these assemblies.

The general agreement (Devlin et al. 2002; Chao et al. 1999; Ishai et al. 1999; Martin and Chao 2001) that distributed overlapping neural structures are involved in tasks interpreted as relying on semantic representations adds extra difficulty to the attempt to demonstrate the warrantedness of the externalisation window. Where semantic representations appear in language evolution theories, they are generally presented as distinct unitary entities; as such, their attachment to an external gesture may appear unproblematic.

Neuropsychological theories of language which employ the concept of semantic memory are generally not concerned with accounting for the evolution of language. When these theories consider the relationship between internal and external representations, an ambient language with the appropriate meanings is assumed. For example, Pulvermüller and colleagues appeal to Hebbian learning to account for the result that words whose meanings are related to different parts of the body evoke activity in the relevant parts of the somatotopically organised motor strip (Hauk et al. 2004; Pulvermüller et al. 2005). Hebbian learning (Hebb 1949) relies on the principle that "neuronal correlation is mapped onto connection strength" (Hauk et al. 2004, p. 301). Thus "words frequently encountered in the context of body movements may produce meaning-related activation in the frontocentral motor areas" (ibid). Neural network models designed to implement different theoretical accounts of CS impairments also rely on correlated activity between representations of words and concept-related representations such as visual/perceptual and functional/associative information (Farah and McClelland 1991; Devlin et al. 1998).

When it comes to explaining usage generalisations in the evolution of language, Hebbian learning raises some difficulties for the viewpoint that such generalisations naturally arise because of a relationship between external gestures and internal representations. If a particular gesture is used in a restricted way (too restricted for our field linguist to be confident about a meaning ascription), then neural activity associated with producing the gesture will co-occur with activity related to all the features of the restricted usages. The Diana monkeys' "eagle" alarm call (Zuberbühler et al. 1997) may be interpreted as being associated with neural activity representing eagles, but it may also be associated with neural activity related to fear, aggression, vigilance, etc. If representations of objects are distributed and overlapping, an account of how a connection between a representation and a gesture could develop would seem to have to allow the connection to depend on patterns of activity over a number of neural structures. But if this were the case it would seem that such a connection would not necessarily be limited to a representation corresponding to just one aspect of the restricted use of the gesture. If a restricted usage allows a field linguist a multiplicity of different interpretations of the "meaning" of the signal, then any neural structure that correlates with any of these candidate meanings could potentially be co-active with the gesture.

Getting "the right" neural structures linked to an external gesture is an important part of an externalisation based attempt to explain the origins of the varieties of usages words with different kinds of meanings have. While it is conceivable that there might be properties of individuals that favour particular kinds of representation to be individually linked to gestures, these properties of individuals need not be invoked to account for modern gesture/representation connections. The extant variety of usages modern words allow may be part of the explanation for the links between internal representations and external gestures in those who have learned a language. This possibility makes it difficult to judge what the constraints are on how internal representations and external gestures may become associated by investigating individuals who have learned a language. Knowing how external gestures become linked to internal representations would help delimit the kinds of usage generalisations that could be assumed in the evolution of language, but without a clear understanding of this process (and what it would produce in different circumstances) it seems that we can do little more than just guess what kinds of usage generalisations would have been made during the evolution of language.

This problem is most obvious for theories which rely on individuals to produce the appropriate range of uses. The holistic route from protolanguage, in contrast, could appeal to random chance as the locus of the origin of various uses. Modern-like correlations between internal representation activity and external gestures could be thought of as arising accidentally, thus producing the right kind of association. However, in addition to the capacity for this process to produce the right kind of association, it also has the capacity to produce the wrong kinds of association too (that is, associations between external gestures and patterns of internal representational activity that do not correspond the associations presumed to underpin modern patterns of usage). If the holistic route is to appeal to internal representations to account for the units of meaning involved in analysis, it will rely on those representations being special in terms of their capacity to associate with external gestures. Without this assumption, it seems that the possibilities for association of (parts of) external gestures and internal representational activity are just as unconstrained as Synthetic route accounts. Both rely on certain patterns of internal representation having special status with respect to their association with external gestures.

3.1.3 Semantic memory: summary

The phenomenon of CS impairments is puzzling (Lambon Ralph et al. 2003). The literature reporting the phenomena and their interpretations is rather complex (Rogers and Plaut 2002). Differences in methodology (Laws 2005) complicate the issue and lead to disagreements about what the phenomena of CS impairments actually are (e.g. Capitani et al. 2003 versus Cree and McRae 2003). Scepticism has been expressed as to the generality of CS impairments reported as *semantic* deficits, and whether category specific double dissociations have in fact been demonstrated (Laws 2005). The extent to which relatively pure cases of CS semantic deficit are due to damage to particular internal representations rather than unusual premorbid experience is unclear (Lambon Ralph et al. 2003). Imaging studies of both brain damaged and normal individuals suggest that a description of the brain in terms of semantic memory will rely on distributed overlapping structures for different semantic representations.

That semantic memory is "a system" is part of its definition, not a clear finding of neuropsychology. The picture of language as a translation from internal representations to external gestures, on which the externalisation window depends, is adopted by many of the neuroscientists whose work has been reviewed in this section, guiding their research (to the extent that the possibility that more mundane factors such as individual experience may account for some cases of CS deficit is a source of frustration, such factors being referred to as "nuisance variables", e.g. Capitani and Laiacona 2005). The theoretical possibility that the neural workings of the brain could in principle (though not necessarily in practice) be re-described in terms of internal representations of meaning will be discussed in chapter 4. However, the findings discussed here throw doubt on the idea that these internal representations are suitable to spontaneously produce usage generalisations in the absence of an ambient language as assumed by theories of language evolution discussed in chapter 2, and therefore suggest that the externalisation window is not warranted.

3.2 Experience–dependence of language related neural structures

The difficulties associated with internal representations of meaning raised in the previous section cast doubt on the externalisation window. However, as the mechanics of the brain involved in language use are so poorly understood, these doubts rely on a degree of intuition as to what form an internal representation which could drive usage generalisations should take. A defender of an externalisation based language evolution theory could respond that while the structure of internal representations are poorly understood and may be rather complex, there may (or even must) also be (complex and poorly understood) neural mechanisms which connect internal representations with external gestures, and that these mechanisms may be innately specified; as such, the externalisation window may be warranted by the inference (from biological continuity) that our ancestors also possessed such mechanisms. The idea that the brain has innately specified language-machinery is not unusual in the study of language (e.g. Chomsky 1986; Bickerton 1995; Petitto et al. 2000; Hauser et al. 2002), and the "necessity" of such externalising mechanisms could be inferred from "poverty of the stimulus" arguments (noting that experience is generally insufficient to determine how a word in the language one is learning is to be used). Such arguments have not been put forward specifically for the relationship between internal representations and external gestures, if only because such a relationship has not been considered problematic in the study of language.

The classical language areas (Broca's area and Wernicke's area) have often been cited in defence of the view that linguistic functions are "hard wired" in the brain (see Bates, Vicari, and Trauner 1999 for a review). Broca (1861) dissected the brains of several patients who had suffered from expressive aphasia and found that they all had lesions in the left inferior frontal lobe (now known as Broca's area). Wernicke (1874) found an association between receptive aphasia and damage also in the perisylvian region, but more posterior than Broca's. Lichtheim (1885) proposed a neurological model which identified Broca's area as the location of the articulatory patterns of words, and Wernicke's area as the location of words' "acoustic images". Were such functions thought to be innately specified in the classical language areas, these could be interpreted as showing (or at least supporting the contention that) the externalisation window has the property of warrantedness: usage generalisations may be spontaneously produced in virtue of Broca's area (somehow) connecting a semantic representation to a motor program.

Since Lichtheim's proposal, the role of the classical language areas has been discovered to be more subtle (for example, individuals suffering lesions to Broca's area commonly show comprehension deficits specific to certain sentence types such as the passive construction, Caramazza and Zurif 1976). Additionally, not every investigation into the relationship between language and the brain implicates the classical language areas (for example, Dronkers et al. 2004 found that lesions to the classical language areas were not associated with performance on a sentence/picture matching task). While an argument that the brain has innately specified externalising mechanisms is unlikely to identify those mechanisms with the classical language areas, this section indirectly considers the possibility that such functions may be "hardwired" by considering whether the classical language areas perform some innately specified function.

3.2.1 *Effects of language structure on Broca's aphasia*

Certain difficulties Broca's aphasics have in producing language have led to the claim that Broca's area is crucial for grammatical processing. Ullman et al. (1997) reported results from a variety of patients' use of past-tense inflection tasks in English that appear to justify such a claim. When English speaking aphasic patients with damage to frontal areas (including Broca's area) were tested, their ability to produce irregular past tense forms was relatively spared, but they had difficulties converting regular forms and nonce-words into the past tense. Aphasic patients with damage to posterior areas (including Wernicke's area) showed the opposite pattern. Ullman et al. (1997) interpreted this result as showing that damage to Broca's area disrupted morphological processes but left stored irregular forms intact, while damage to Wernicke's area affected stored items but spared the ability to perform morphological operations on words.

On the basis of these and other results, Pinker and Ullman (2002) propose a model of the relationship between language and the brain in which the function of Broca's area is seen as manipulating words and morphemes that are stored in other parts of the brain. However, this view of the role of Broca's area as the area that performs grammatical operations is not unproblematic. Cross linguistic studies suggest that in languages with more elaborate inflectional systems than English, Broca's aphasics' rates of omission of grammatical morphemes are markedly lower (e.g. Bates, Wulfeck, and MacWhinney 1991). Penke and Westermann (2006) point out that the English past tense is not ideally suited to test the hypothesis that Broca's aphasics have suffered damage to a device devoted to manipulation of parts. The English past tense confounds regularity with the kind of difference between stem and past tense forms of a verb. The regular past tense form in English is related to the verb stem by the affixation of the suffix *-ed* (e.g. *jump — jumped*) while irregular forms often only differ from the infinitive form in terms of the stem vowel (e.g. *dig — dug*).

In German and Dutch, both regular and irregular past participles consist of the verb stem with a prefix and a suffix. Irregular past participles generally take a different suffix to the regular participle and have a modified stem vowel. According to the kind of mechanism proposed by Pinker and Ullman (2002), German and Dutch speaking Broca's aphasics should show impairments in producing regular past participles while irregular participles (assumed to be "stored") would be relatively unaffected. However, Penke and Westermann (2006) report results from an experiment with German speaking Broca's aphasics (using the same elicitation paradigm as Ullman et al. 1997) which show the opposite pattern: the majority of their subjects showed a selective impairment for irregular past participles. Similar results were obtained with Dutch speaking Broca's aphasics.

Penke and Westermann (2006) present a neural network model which can be trained to produce the past participle from the infinitive of regular and irregular German verbs. When lesions are simulated in this model (by removing units from the hidden layer) performance declines in similar fashion to that seen in German speaking Broca's aphasics. The significance of this result is not so much the plausibility of the network as model of neural architecture, but that the state of the model prior to lesioning (and hence the effect of lesioning) depends crucially on the statistical distribution of forms it received as input (which Penke and Westermann derived from the CELEX database). Thus, if we follow Pinker and Ullman (2002) in interpreting the effects on past-tense production in English speaking Broca's aphasics as demonstrating that Broca's area performs morphological manipulations in English speakers, this functional localisation appears to be experience dependent.

3.2.2 *Early damage to the classical language areas*

The effects of early damage to the classical language areas have long been seen as a source of evidence for or against innately given neural language machinery (Woods 1983; Bates et al. 1999). Difficulties associated with early lesions (e.g. inferring the location of damage, or complicating factors such as seizures) have meant that historically the evidence has been unclear (Woods and Teuber 1978; Carter et al. 1982; Vargha-Khadem et al. 1985; Bates et al. 1999; Vargha-Khadem et al. 1994). However, studies restricted to pre- and peri-natal lesions (reviewed by Bates and Goodman 1997) which exclude cases suffering seizures and other medical complications and for which lesion locations were relatively well known, suggest that the linguistic abilities of individuals with congenital unilateral brain damage develop to be within the normal range regardless of lesion side, size or site. While, as a group, children with damage only in the LH display language performance that is significantly below normal controls, children with congenital damage to the classical language areas of the left hemisphere do not individually show abnormally poor language development and would not be classed as aphasic. However, Bates and Goodman (1997) note that a language specific role for the LH is not entirely ruled out by the data: when children with damage to the left temporal cortex (the presumed site of Wernicke's area) are compared to children with any other damage, they show delays in expressive language across the period from 10 to 60 months of age. This deficit affects performance on tests measuring lexical abilities and grammatical measures. This appears to be the only effect of lesion side, and it doesn't seem to last beyond 5–7 years (i.e., the linguistic deficits experienced after this age show no effect of lesion site or side). Whatever the specific details of how damage to the left temporal cortex effects early expressive language, the brains of these children undergo sufficient reorganisation to bring their abilities into the same range as children with damage to other parts of the brain, demonstrating that the language functions performed by the classical language areas need not be hardwired.

On the basis of these and other results, Bates et al. (1999) suggest a compromise position between "equipotentiality" (the view that at birth the two hemispheres are equally capable of mediating language functions) and "irreversible determinism" (the view that left-lateralisation of language functions is innate and irreversible, a view compatible with the notion that language functions are genetically fully specified and so can straightforwardly be projected onto genetically similar ancestors). They describe their view as "constrained plasticity" and posit "soft constraints" which are not deterministic and can be overcome. These constraints, however they are instantiated, are supposed to be far less direct than language functions being "hard-wired" in the LH from birth. Examples of possible differences between the hemispheres given by Bates et al. (1999) include variations in processing speed, neural packing density and types of neural transmitters. Their suggestion is that language functions do not need to be wired in before language learning can take place, but rather develop from the interaction of linguistic experience and low-level differences between the hemispheres which favour development of language related activation in the left hemisphere over the right.

One possible "soft constraint" may be simply the physical asymmetry of the brain. Dehaene-Lambertz and Dehaene (1994) found that as a group, the responses of infant brains to the first syllable in a sequence and to deviant syllables showed hemispheric asymmetries in the same direction as those found in adults. However, they noted that these differences may result from either functional lateralisation or differences in brain morphology: for instance, differences in orientation of the left and right perisylvian fissures. They also noted that subjects showed a considerable degree of variability in size and in some cases direction of activation asymmetries. These results suggest a moderate LH advantage for language sounds rather than a sharp lateralisation of function.

The constrained plasticity view of language lateralisation suggests that whatever the language related functions performed by the classical language areas of undamaged brains are, they develop through individuals' experience rather than being "hardwired". In different circumstances (offering different experiences) these neural processes may develop differently. If we generalise this view to putative externalising mechanisms of contemporary humans, then even if such mechanisms are granted the inference that they would drive the usage generalisations of chapter 2 in the absence of experience of such a language is not sound.

3.2.3 *Why are Broca's and Wernicke's areas where they are?*

The soft constraints suggested by Bates et al. are intended to account for the consistent lateralisation of neural structures dealing with language while allowing that these functions may, to a large degree, develop through experience. However, lateralisation is not the only consistent feature of the organisation of normal language functions. If these language functions develop through individuals' experience, mightn't we expect to observe greater variability in their location in the LH?

Pulvermüller (2002) offers an explanation for why Broca's and Wernicke's areas should be crucially involved in mature language functions. Pulvermüller's explanation appeals to some very general principles of neural organisation with an additional *ad hoc* principle that language processes are lateralised to the LH (Pulvermüller adds this latter principle, analogous to Bates et al.'s "soft constraints", as an acknowledgement that the specific causes of language lateralisation are presently unknown). Pulvermüller's suggestion relies on the Hebbian principle (Hebb 1949) that neurons that fire together strengthen their mutual connections while neurons that fire independently of each other weaken their mutual connections. During speech, neurons controlling the movements of the articulators are active. This activity can spread to adjacent areas in the inferior frontal lobes. Provided there is no damage to the auditory system, activity will also be present in the auditory cortex due to stimulation by the sounds produced by speech, and this activity may spread to adjacent areas in the superior temporal lobe. Because there are long distance connections between the superior temporal lobe and inferior frontal lobe, the correlated pattern of activity set up by a normally hearing person speaking can strengthen the connections between the inferior frontal lobe (which includes Broca's area) and the superior temporal lobe (which includes Wernicke's area) (Pulvermüller 2002, p. 37). While this explanation is too vague to predict precisely what effects local damage to these areas should have on mature language users, it does offer an account for the location of language areas in terms of general principles of neurophysiology. Pulvermüller's (2002) explanation may be able to account for the commonality of the classical language areas across individuals without having to see these as specified prior to experience. His explanation may also predict some differences between individuals based on their language.

3.2.3.1 *Language areas of deaf signers*

Pulvermüller's explanation of the locations of the classical language areas suggests that the brains of deaf people who use sign language should show different patterns of language related brain organisation. If the location of Broca's area is explained in part by the location of the areas controlling movement of the face and mouth, then the functionally comparable area of the brain of deaf sign-language users might be expected to be closer to the parts of motor cortex which correspond to movement of the hands and arms. If this were the case then deaf signers would be expected to have a frontal language centre dorsal to that seen in spoken language users. Evidence related to this prediction, however, is unfortunately not particularly clear: reports of frontal lesions leading to aphasia in sign language users (e.g. Poizner et al. 1990) are difficult to interpret in this respect as cases in the literature tend to involve rather large lesions involving *both* the area anatomically corresponding to the frontal language area in spoken language users and the more dorsal areas that might be predicted to be involved in sign language (Pulvermüller 2002).

Corina et al. (1999) provide more direct evidence for the location of the functional equivalent of Broca's area in sign language users. They report data from a cortical stimulation mapping (CSM) procedure performed on a deaf user of American Sign Language (ASL) undergoing a surgical procedure on the left hemisphere for treatment of a seizure disorder. The CSM procedure infers the function of parts of the cortex by observing the effects of application of a localised electric current at specific cortical sites. Corina et al. (1999) found that stimulation of the posterior portion of Broca's area resulted in laxed and imperfect formation of signs for both sign repetition and object naming tasks. Non-sign repetition was also affected by stimulation of this area, ruling out effects of semantic content disruption. Unfortunately, Corina et al. (1999) did not test the effects of stimulation to regions dorsal to Broca's area which Pulvermüller's explanation suggests may be more implicated in sign language than spoken language.

Corina et al. (1999) take the proximity of the area they did test (and for which they found an effect) to the site of motor representations of the mouth as suggesting that the function of this area cannot be accounted for in terms of proximity to the motor representation of the parts of the body used in producing language. However, there are two factors which complicate predictions based on Pulvermüller's suggestions. First, the repeated use of hands and arms rather than face and mouth in communication may lead to cortical reorganisation (Buonomano and Merzenich 1998), in which case areas of motor cortex normally involved in movements involved in speech may be recruited for sign-language movement (though it should be noted that none of the areas Corina et al. tested resulted in hand or arm movement). Second, Pulvermüller's suggestion relies on the availability of neuroanatomical links between areas, and these links may also constrain the location of frontal language areas.

Evidence concerning the equivalent of Wernicke's area in sign language users is more suggestive of an anatomical difference. Pulvermüller's suggestion appeals to the location of auditory areas to account for the location of posterior language areas. Because sign language is perceived primarily visually (as opposed to primarily auditorially for speech[3]) Pulvermüller's account would predict a different location for the functional equivalent of Wernicke's area in deaf signers due to the different locations of visual and auditory cortex. There does appear to be some evidence for this difference. Corina et al. (1999) found that stimulation of the

[3] Auditory perception does not have sole responsibility for speech perception, as demonstrated by the role of visual perception in the McGurk effect (McGurk and MacDonald 1976)

supramarginal gyrus (SMG), did not impair the subject's ability to repeat signs and non-signs, but did induce formation and semantic errors in an object naming task. The SMG is an area in the parietal lobe which lies above the part of the temporal lobe identified as Wernicke's area. Corina et al. (1999) compare their result to MEG results which implicate Wernicke's area and the tempo-parietal junction in spoken language picture naming (Levelt et al. 1998), and note "the differences in anatomical location implicated across these two studies (i.e., SMG versus Wernicke's area) [may] reflect normal variation, methodological differences, or language modality differences," (Corina et al. 1999, p. 580). The latter possibility is supported by evidence from sign-language aphasias. In their review of the literature on sign-language aphasias, Poizner, Bellugi, and Klima (1990) noted that a number of cases of sign language aphasia resulted from lesions to the parietal lobe, and that corresponding lesions suffered by hearing individuals would not be predicted to result in aphasia.

Likewise, imaging results find greater language modality dependent variability in the more posterior language processing regions. Sakai et al. (2005) compared activation in a number of sentence comprehension conditions involving aurally presented Japanese and visually presented Japanese sign language (JSL). The only area that did not show a main effect of language modality was Broca's area. The locations of posterior language processing regions were dependent on language modality. The prediction of differences in language processing areas derived from Pulvermüller's (2002) explanation appear to be borne out for posterior language structures. Thus predictions from an experience-driven account of the locations of the classical language areas appear to be supported for the functional equivalent of Wernicke's area.

3.2.4 *Summary: plasticity in the development of neuro-linguistic functions*

This section has reviewed evidence that functions attributed to the classical language areas on the basis of lesion studies with mature language users develop through experience with language. The view that soft constraints on experience–driven processes rather than "hardwiring" are responsible for the commonalities seen across individuals' language processing brain regions fits well with considerations of the role of plasticity in the development of non-linguistic processes. West-Eberhard (2003) criticises the general idea that genes constitute a set of instructions for the construction of an organism:

> The complete-instruction metaphors are particularly problematic because they reinforce the misconception held by many that the genome is a complete set of instructions for making an organism. The image has little resemblance to the decentralized way that responsive structure is made during ontogeny, with the form and function of tissues and organs depending importantly on their circumstances and uses as they are being formed. (West-Eberhard 2003, p. 15)

Experience seems to play an important role in determining neural organisation for brain functions other than language. Buonomano and Merzenich (1998) reviewed a great deal of evidence for what they take as the modern consensus view that "cortical maps are dynamic constructs that are remodeled in detail by behaviorally important experiences throughout life" (p. 150). Buonomano and Merzenich cite Hebbian plasticity as playing an important role in both cortical development (e.g. the development of somatotopically arranged cortical maps) and cortical reorganisation in adult animals in response to either physical changes to the body or to training. One example Buonomano and Merzenich (1998) cite to demonstrate cortical rearrangement due to a physical change was a study by Clark, Allard, Jenkins, and Merzenich (1988) which found that the surgical connection of two fingers of adult owl monkeys can erase somatotopic boundaries between digit representations. Prior to surgery, the zones representing different digits showed "striking discontinuity" (Clark et al. 1988, p. 444). This surgical manipulation increases the correlation between inputs received by the brain from the two fingers, and so also increases the correlation between input received by one finger and activity which had previously been only responsive to the other finger. Through Hebbian plasticity, this new correlation induced brain cells which had previously been responsive to only one finger to become responsive to both, even (shortly) after the fingers were surgically re-separated. Buonomano and Merzenich (1998) suggest that similar processes (relying on correlated input from adjacent areas) are important for the development of somatotopic cortical maps in ontogeny. Training can also induce changes to cortical maps: Recanzone, Schreiner, and Merzenich (1993) trained monkeys on an auditory frequency discrimination task, and found increases in the size of the cortical areas responsive to the trained frequencies.

The importance of plasticity and experience-driven development suggest that even if a mechanism responsible for externalising the internal representations of modern language users is assumed, there is little support for the assumption that *the same* mechanisms would be found in the brains of our ancestors at an earlier stage in the evolution of language when experience with other individuals who already know a language was not available. That is, even if such mechanisms are assumed for modern language users, they do not demonstrate that the externalisation window has the property of warrantedness.

3.3 First language acquisition

The final potential source of evidence that the externalisation window has the property of warrantedness considered in this chapter is first language acquisition. If children learning their first language appear to spontaneously make the kinds of usage generalisations assumed by language evolution theories (chapter 2) this would support the contention that our ancestors would have made the same generalisations at an earlier stage in the evolution of language.

3.3.1 Word learning "strategies"

Estimating the size of an individual's vocabulary is a difficult task (in part because of the vagueness of what should count as an individual word). Bloom (2000, p. 25) offers what he sees as a conservative estimate that by school leaving age, the average American or British person has learned 60,000 words which corresponds to about 10 words learned per day (or one word every waking 90 minutes) of an individual's early life. In certain experimental situations, children can learn the meaning of a word on the basis of very little exposure (Carey and Bartlett 1978; Tomasello and Akhtar 1995; Markson and Bloom 1997). Quine's (1960) *gavagai* thought experiment is often cited to demonstrate just how impressive this feat of word learning is (Bloom 2000; Clark 2003; Masur 1997; Gentner 1982; Landau, Smith, and Jones 1988). Quine imagined a field linguist trying to find the correct English translation of a word in an unknown language (*gavagai*) uttered by a native speaker as a rabbit runs past. In Quine's thought experiment the field linguist relies on a number of observations of the use of *gavagai* in order to rule out certain translation possibilities (e.g. *white*, or *furry*).[4] In contrast

[4] The problems of induction from limited evidence are the less interesting facets of Quine's thought experiment. It is often overlooked in discussions of children's acquisition of new words that Quine's point was that even with unlimited behavioural evidence the field linguist could assign radically different meanings to a term (e.g. *rabbit* vs *all–and–sundry–undetached–rabbit–parts* or *brief–temporal–slice–of–rabbit* for *gavagai*).The problem is not a paucity of evidence as no amount of evidence could decide these issues. Rather, the thought experiment is an attack on the notion that there is such a thing as a correct translation/definitive meaning assignment. This attack does

children seem able to infer the appropriate meaning (as evidenced by their abilities to appropriately use and respond to the word) without such cross-situational learning.

This ability to quickly learn the meanings of new words has been interpreted in terms of initial or default hypotheses brought by children to the word learning task (e.g. Landau, Smith, and Jones 1988). These hypotheses bias the child towards interpreting novel words in certain ways, and making certain generalisations of use. Thus a child in the place of Quine's field linguist might, by default, infer from the native speaker's single use of *gavagai* that the word refers to what the English word *rabbit* refers to (and not, for example, what the English word *white* or the English noun phrase *Himalayan spotted rabbit* refer to).

3.3.1.1 Noun bias

If children are biased towards assuming certain kinds of meaning for words, they might be expected to acquire words with those kinds of meaning faster than other words. Children's early vocabularies are often reported to be dominated by nouns (Gentner 1978; Gentner 1982; Caselli et al. 1995). Gentner (1982) provided the first cross linguistic comparison of this phenomenon, and noted that it seemed to occur in languages that might be thought of as syntactically disfavouring nouns. English has appeared to some researchers (e.g. Choi and Gopnik 1995) to favour noun acquisition over verb acquisition more than some other languages for a number of reasons: subjects are obligatory in declarative sentences (even those which semantically do not seem to have a subject, such as *it is raining*) whereas some languages allow null-subjects (as in Italian, *Piove* – "(It) is raining", Caselli et al. 1995) thus English presents a child with fewer verb–only utterances than other languages; English word order is relatively strict and objects are placed at the end of utterances, a position of relatively high salience; and English nouns are morphologically less complex than English verbs. However, Gentner (1982) argued that in languages in which these factors varied relative to English, the noun bias in early vocabulary was still observed.

not seem to be answerable by appealing to data from language acquisition as the attack depends on the bloody-mindedness of an imagined field linguist, not on any difficulties *children* encounter when learning new words (if it did, Quine's conclusion would be that children find it *impossible* to learn new words, a palpably false proposition). Nonetheless, much research has pitted such data against Quine's *gavagai* thought experiment. Bloom (2000, p. 12) provides a good example of a misunderstanding of the point: "If a dog jumps onto a stove and gets burned, it is likely to infer that stoves are hot — not that undetached stove parts are hot." Of course, if stoves are hot, then so are undetached stove parts, and vice versa.

It should be noted that the universality of the noun bias is a matter of some controversy. Some studies have presented evidence that the early vocabularies of children learning some languages (such as Korean: Choi and Gopnik 1995; and Mandarin: Tardif 1996) do not show a bias towards nouns. However, studies which have reported no bias in favour of nouns were based on samples of children's spontaneous speech. Obtaining an accurate estimate of a child's vocabulary using this method is problematic as the chances that a child does not produce a word they know are quite high. Tardif, Gelman, and Xu (1999) estimated early vocabularies of English- and Mandarin-speaking children from spontaneous samples, and found that the proportions of nouns and verbs discovered by this method varied with the context in which recordings were made: regardless of the language spoken, children's vocabularies appeared dominated by nouns when they were engaged in book reading, but not when they were playing with toys.

An alternative way of assessing a children's vocabularies is to ask their parents. Parents are given a list of words and asked whether the child (a) understands the word, (b) understands and produces the word. When this method of vocabulary assessment is employed, a noun advantage is generally found for children's early productive and receptive vocabularies, even for "verb-friendly" languages such as Italian (Caselli et al. 1995) and Mandarin (Tardif, Gelman, and Xu 1999).

Parental checklists are not free of problems, some of which may bias the noun/verb ratios found: checklists can discourage reporting of proper nouns; success depends on having an inclusive, language appropriate list of words; checklists commonly ask for nouns first, which could possibly lead to fatigue factors in reporting verbs; phrases used as wholes may be underestimated; and for heavily morphologised languages it can be difficult to decide how to count words (Gentner 1982, p. 238). An additional difficulty with checklists was identified by Tardif, Gelman, and Xu (1999) who found that for both English and Mandarin, verbs that the child had been observed producing during experimental sessions were more likely to be omitted when parents filled out checklists than were nouns. In spite of these difficulties, evidence on early vocabularies obtained by the checklist method appears to be favoured by first language acquisition researchers, and is interpreted as showing that children do have a bias towards noun learning, albeit one that is modulated by the degree to which the target language is "verb-friendly" (Tomasello 2003; Gentner and Boroditsky 2001; Caselli, Casadio, and Bates 1999).

Gentner (1982) proposed that a noun advantage is (at least in part) a product of the way children perceive and cognize the world around them. Her "Natural Partitions hypothesis" holds that the linguistic distinction between nouns and verbs (seen as universal) is based on a pre-existing perceptual–conceptual distinction between concrete concepts (e.g. persons and things) and predicative concepts (e.g. activity, change–of–state, causal relations). According the the Natural Partitions hypothesis, the internal representation of the meanings of concrete nouns are less complex than the representations of verb meaning, and this accounts for the early noun advantage (Gentner 1982; Gentner and Boroditsky 2001).

3.3.1.2 Comparison of noun and verb usage consistencies

Sandhofer, Smith, and Luo (2000) analysed distributional features of the language directed at children by English speaking caregivers. In common with analyses of adult–adult (written) English, they found that adult–child English speech had roughly equal token frequencies for nouns and verbs, but a greater number of nouns than verbs on a type–count: English speaking adults use fewer verbs than nouns when speaking to children, and on average repeat verbs more frequently than nouns. Gentner (1982) had taken this difference between nouns and verbs as a fact that should favour verb learning as individual verbs are heard more frequently than individual nouns.

However, when Sandhofer et al. (2000) analysed the statistical structure of the noun and verb classes, they found a more even frequency distribution for nouns than for verbs: the most frequently produced verbs constituted a far higher proportion of the total number of verb tokens than the most frequently produced nouns (as a proportion of the total number of noun tokens). While the fifth most frequent verb was approximately three times as frequent as the fifth most frequent noun, the frequencies of lower ranking nouns and verbs were roughly equal (i.e. the frequency of the 25th most frequent verb was roughly the same as the frequency of the 25th most frequent noun, and likewise for the 50th and 100th most frequent words). Thus the advantage that a higher token frequency should bestow is only relevant to a small number of verbs.

A more important difference between nouns and verbs was discovered when the semantics of the most frequent nouns and verbs were considered. Sandhofer et al. (2000) judged each word according to a set of features. In terms of these features, the most frequently used nouns were far more consistent than the most frequently used verbs: 72% of nouns were count nouns referring to solid objects which shared the same shape. In contrast, while 76% of verbs referred to motion, 20% were ACTOR-ONLY motion verbs, 52% were ACTOR-AND-PATIENT verbs and 14% were ACTOR-OR-PATIENT verbs.[5] Verbs displayed a greater diversity of meanings according to this analysis. When the speech of Mandarin–speaking caregivers to children was analysed, similar results were found.

Sandhofer et al. (2000) suggest that the relative consistency of noun referents (for the nouns children are exposed to) may aid their acquisition. Children can make generalisations across nouns in terms of their extensions which will aid the acquisition of other nouns. In contrast, verbs present a less unified structure for which it is harder to identify a generalisation, and for which any generalisation made will be less informative than noun meaning generalisations. Thus, in contrast to the Natural Partitions hypothesis, Sandhofer, Smith, and Luo's results suggest that it is the consistency of noun meanings relative to verb meanings rather than the (presumed) structure of cognitive representation that accounts for the noun bias. If the patterns of usage in the language being learned account for the noun advantage, the noun advantage is the product rather than evidence for the cognitive cause (via externalisation) of these usage patterns. As such, the noun advantage does not constitute evidence that reference to concrete objects is a basic assumption made by children, and does not show that the externalisation window is warranted in attributing concrete-object-reference to earlier stages in the evolution of language.

3.3.1.3 Noun-extension generalisation strategies

Sandhofer et al.'s (2000) findings suggest that the noun advantage results from children's exploitation of regularities in the extensions of nouns in their languages. A number of studies have experimentally investigated young children's expectations about the relationships between objects which can be referred to with a newly encountered noun. In these experiments, children are presented with a target object (or picture of an object) along with a nonce word used to refer to the object ("this is a dax"). Generalisation strategies can then be assessed by presenting children with test objects which differ from the target object in specified ways and testing with various response paradigms (e.g. forced choice "which one is a dax?", yes/no "is this a dax?", etc.). Children's responses can be

[5] For some reason, 14% of verbs seem not to have been classified in terms of actor-patient relationships.

compared with chance, with responses in “no-label” conditions (in which target objects are not named and children given an instruction such as “find another one the same as this”), and with adult responses. The most firmly established generalisation strategy found using these techniques is generalisation of a word used to refer to a solid object on the basis of shape (ignoring dimensions such as texture, colour and size, Landau, Smith, and Jones 1988). This “shape bias” corresponds to the use of the majority of nouns used by adults when talking to children (Sandhofer et al. 2000).

Shape is not the only organising principle along which children generalise novel words. Other generalisation strategies have been found when target objects have been more than simply novel solid objects. For example, Jones, Smith, and Landau (1991) repeated the experiments of Landau et al. (1988), but added toy eyes to the objects. They found that this addition (which they took as an animacy cue) prompted children to generalise labels to objects with similar shape and texture. Jones and Smith (2002) found a similar effect, and related this to the words in young children’s vocabularies: adults were given lists of names for animals and non animal objects that children learn early on, and asked to judge whether each name referred to a class of things with similar shape, similar colours, and/or were made of similar material. Animals were judged to be similar in terms of shape and material, whereas non animal objects were generally rated as being most similar in terms of shape. Thus the generalisations children make, which are affected by animacy cues, are appropriate to the categories of object they learn the names for early in life.

Booth and Waxman (2002) and Yoshida and Smith (2003) found that the linguistic context within which a novel count noun was presented affected the way children generalised. Booth and Waxman (2002) presented novel names for novel objects embedded in short vignettes that suggested the object was either an animal or a tool. Children generalised the name to objects with the same shape in the tool condition, and to objects with both the same shape and texture as the original object in the animal condition. The strength of this effect seemed to override the effect of adding eyes as animacy cues as Jones et al. (1991) had done: when objects with eyes were presented in stories suggesting the object was a tool, children generalised on the basis of shape, rather than the combination of shape and texture. Yoshida and Smith (2003) introduced Japanese children to novel words as labels for objects designed to have ambiguous animacy cues. Japanese makes a distinction between animate and inanimate things in locative constructions. Whereas English would use the same verb for both animate and inanimate things ("There is a cup" and "There is a dog"), Japanese (to a rough approximation) uses the verb *iru* for animate things and *aru* for inanimate things. Yoshida and Smith (2003) found that when a novel word is presented to Japanese children as the name for an object using the verb *iru*, they generalised on the basis of shape and texture combined. In contrast, when words were introduced using the verb *aru*, generalisation was on the basis of shape.

Other cues that alter the generalisations children make are solidity (labels for non-solids tend to be generalised on the basis of material rather than shape, Soja, Carey, and Spelke 1991; Imai and Gentner 1997) and apparent intended function (for example, children do not appear to show the shape bias when similar shaped objects are known to have different functions, such as one being the container for the other, Diesendruck, Markson, and Bloom (2003)).

The possibility that noun generalisation strategies reveal innate ontological categories which children naturally expect to be labelled has been raised by several researchers (Markman and Hutchinson 1984; Soja, Carey, and Spelke 1991; Imai and Gentner 1997). Innate generalisation biases would have obvious significance for the warrantedness of the externalisation window. However, there are a number of reasons for thinking that the generalisations children make reflect their experience. A number of studies have noted that the various biases children display when generalising novel labels become more pronounced with age (Landau et al. 1988; Jones et al. 1991; Imai and Gentner 1997; Samuelson and Smith 1999; Jones and Smith 2002). Increasing bias strength with age is not incompatible with the idea that biases are innate as they may take time to develop, but this evidence does not support the interpretation of biases as innate over an experience dependent interpretation. Stronger evidence for experience dependence comes from a longitudinal study of eight children, which found that the shape bias increased over time, and was most strongly linked to the number of nouns in each child's productive vocabulary (more strongly than to either time in the experiment or to age, Gershkoff-Stowe and Smith 2004). If the number of nouns in a child's vocabulary is taken as a reflection of their experience, this result supports the view that the shape bias is driven by experience with nouns whose extensions can be grouped by shape.

Interestingly, cross-linguistic/cross-cultural variability has been found in these experiments: Imai and Gentner (1997) used a forced choice paradigm to compare shape and material based generalisation. When children were given labels for complex objects, American and Japanese children generalised on the basis of shape, as did American and Japanese adults; for simple objects, however, American children and adults showed a shape bias but Japanese children and adults did not (Japanese adults commonly generalised on the basis of material while children's responses were not different to chance); and for non-solid targets (e.g. Nivea, hair-setting gel, etc.) American children and adults generally did not show a preference while Japanese subjects generalised in the basis of material. Imai and Gentner's choice of native English and Japanese speakers was designed to test whether the English count/mass distinction is responsible for young American children's tendency to generalise solid targets on the basis of shape and non-solid targets on the basis of material. Japanese (according to Imai and Gentner's analysis) lacks a linguistic count/mass distinction. The fact that both groups of children generalised on the basis of shape for solid targets to a greater extent than they did for non-solid targets was taken by Imai and Gentner as supporting the view that word learning is guided by pre-linguistic ontological categories since the difference could not be attributed to count/mass syntax in Japanese. However, the similarities and differences in children's responses across the two language groups mirrored the similarities and differences in adult responses, leaving open the possibility that in both groups, children's responses simply reflected what they had learned about noun extensions from the adults around them.

Training studies have shown that experience with language *can* induce generalisation biases. Jones and Smith (2002) conducted an experiment in which 22-month-old children were introduced to four novel categories of object through play with an experimenter over a period of seven weeks. Categories were created with two exemplars and one non-instance. Two of the object categories were designed to be non-animals (i.e., had no animacy cues), and the exemplars within each of these categories matched in shape only; the other two categories were given eyes (animacy cues) and were grouped by shape and texture. The distinction between exemplars and non-exemplars was presented to subjects through play ("Look, here's a dax," "Oh, that's not a dax"). When tested with new categories without eyes, trained children showed a shape bias similar to that observed for untrained children of the same age. For objects with eyes, however, untrained subjects generalised on the basis of shape alone whereas trained subjects generalised on the basis of shape and texture combined. Thus training with just four categories with two exemplars each had induced in these young children generalisations characteristic of the generalisations older children make for nouns referring to animals.

Yoshida and Smith (2005) conducted a similar training study with Japanese children to test whether a correlated linguistic cue could induce differential generalisation strategies (just as the non-linguistic animacy cue had in Jones and Smith's experiment). During training, shape based object categories were presented to children using one classifier (*hitotsu*) while material based substance categories were presented using a different classifier (*sukoshi*). Half the subjects (two-year-old Japanese children) received training with these correlated cues, and half did not. After training, Yoshida and Smith (2005) tested subjects generalisation of a different set of novel words for different novel objects. "Correct" responses were defined as generalising on the basis of shape for solid objects and generalising on the basis of material for non-solid objects. Subjects who had been trained with correlated linguistic cues outperformed subjects trained without cues, both when cues were present and absent during testing. Thus linguistic cues that correlate with lexical category structures help induce biases based on those structures.

These studies demonstrate that experience with language *can* induce these biases. Investigations of the early vocabularies of young children also show that the nouns they learn have the right sort of structural properties to induce the clearest of children's generalisation biases, the shape bias. When the words in English speaking children's early vocabularies are rated by adult participants, the majority of nouns are count nouns referring to solid things organised by shape (Samuelson and Smith 1999; Gershkoff-Stowe and Smith 2004). And, as mentioned above, the noun vocabularies used by mothers in interaction with their children are also dominated by nouns referring to solid objects with similar shapes (Sandhofer, Smith, and Luo 2000).

While these results are consistent with the interpretation of generalisation strategies as being experience driven, the relationship between generalisation bias and early vocabulary can be interpreted as consistent with biases being innate: if noun generalisation strategies aid language acquisition, the link between vocabulary and shape bias could be due to this acquisition boost (Gershkoff-Stowe and Smith 2004). However, Smith et al. (2002) provide evidence that generalisation strategies can be experimentally induced in children (using similar procedures to those of Jones and Smith 2002) and that these result in increased rates of vocabulary growth (outside the laboratory) compared with untrained children. Thus, while innate biases are a theoretical possibility, the evidence from studies of children's noun generalisation strategies suggest they are unnecessary in an account of language acquisition, and as such do not serve as evidence that the externalisation window is warranted.

3.3.2 Earliest word uses

Section 3.3.1 above focused on which words children acquire early. However, these issues provide only indirect evidence about the *use* children make of words early on during language acquisition. Words in children's early vocabularies are generally categorised according to their roles in adult language (Caselli et al. 1995). While it seems likely that these categorisations would correspond to differences in the ways words are used, there has been little systematic research into the issue of how children use the words they know, and how these usage patterns develop.

A number of studies have reported that children's earliest uses of some of their first words are tied to specific contexts. For example, Bloom (1973) noted that when her daughter began to use the word *car* (at 9 months), she did so only when looking out of the window at cars moving on the street below. She did not use the word to refer to stationary cars, to pictures of cars, or while she was sitting in a car. Tomasello (2003) describes this kind of restricted usage as being not "truly symbolic". For a child to demonstrate that a truly symbolic grasp of a word the word must be "flexibly used across an array of appropriate referential situations" (Tomasello 2003, p. 52).

Harris et al. (1988) investigated the degree to which the first words produced by children were context bound. They recorded interactions between four mother/child pairs between the ages of 0;6 and 2;0, and used a combination of maternal diary reports and experimenter judgement to identify words produced by the children. Using specific selection criteria criteria, the first ten words of the four children were recorded along with the behavioural contexts in which those words were used. Of these 40 words, Harris et al. (1988) described 22 as being CONTEXT-BOUND ("use in only one highly specific behavioural context"), 14 as being NOMINAL ("used to refer to one or more objects in at least two different behavioural contexts"), and 4 as NON-NOMINAL (not used to refer, but used in at least two different behavioural contexts).

Harris et al. (1988) also coded mothers' uses of their children's first words during the videotaped sessions. For 37 out of the 40 words, the child's initial use of a word and their mother's uses of the same word were judged to share a "strong resemblance". For the 22 words that were used in context-bound ways, 18 were found to be related to the most frequently observed maternal uses. For example, one child (James) only used the word *mummy* when handing a toy to his mother, and his mother's most frequently observed use of this word was when she held out her hand to take a toy from James. The words that were used by children in more than one behavioural context were also judged to have been used across more than one behavioural context by the mother, and for the majority of these (12/14) a common element could be identified across contexts of both child and mother's uses.

While these results argue against the notion that children's first words are *never* referential, they do not unequivocally suggest that usage generalisations are spontaneously produced by children. The dependence of the CONTEXT-BOUND words on maternal usage may be taken as suggesting that NOMINAL child uses were also dependent on maternal uses, but that the variety of maternal uses prompted a variety of child uses. According to Tomasello (2003, p. 53), the possibility remains open that "children's early event-bound words are simply those relatively few words that they have heard and attended to in only a single communicative context or with a single communicative purpose, with no opportunity (for whatever reason) to observe adult generalizations to other contexts." An extrapolation from this possibility to earlier stages in the evolution of language would suggest that observation of (or interaction with) different uses of external gestures would have been necessary for modern-adult-like usage generalisations to occur.

Investigation of the first uses of some of the words that enter a child's vocabulary relatively late (compared to first nouns) suggests that children's use of these words develops gradually, and mirrors the usages they have been exposed to (if somewhat imperfectly). For example, Levy and Nelson (1994) found that one particular child used *because* in a way that was restricted to the topics her father had used the word for, emulated the sentence constructions of his that she had heard, but displayed poor understanding of the function of this word (for example, she would often omit a clause that a *because* clause was to explain). Her control of this word improved with time, as did a number of temporal words (*afternoon*, *tomorrow*, *pretty-soon*).

3.3.2.1 Adult construction of child meaning

Investigations into children's first uses of words face a number of methodological and conceptual difficulties. Methodologically, observation of first uses presents a challenge as the child has to be under a high level of observation for a sustained period of time. Conceptual difficulties attach to decisions experimenters have to make concerning what a child has said. For example, there is a general acceptance that babbling continues for some time after the appearance of a child's first words and there are strong similarities between the phonetic sequences in babbles and early words (Oller et al. 1976); decisions have to be made by investigators as to whether a child has tried to say a word or has produced a babble that sounds like the word. For example, in an experiment designed to test the ways children initially use (adult) nouns, Huttenlocher and Smiley (1987, p. 71) noted that "some speech sounds produced by the youngest children were not treated as utterances. We excluded … apparently random sounds when a child's attention did not appear to be focused on anything." It is possible that the extent to which children appear to use language as adults do is in part a function of a selection filter imposed by adults who dismiss vocalisations they cannot understand as babbling. Clark (2003) summarises the problem:

> Children also try out some words in ways that are hard to link to any identifiable adult use. The target word itself may not be identifiable, and the general lack of adult comprehension typically leads to the word's being abandoned. Such mismatches, though, perhaps because of their complete failure to communicate anything, have rarely been reported. (Clark 2003, p. 92)

This is a conceptual difficulty in language acquisition research, as it isn't clear how the distinction between non-language babbling and poorly articulated attempts to use language in ways not conforming to adult usage is to be made. The fact that adults can impose interpretive structures on children's vocalisations raises difficulties for discussions of children's earliest uses of language, but it may also form an important aspect of the language learning process. Unfortunately, the role of adult interpretation of child utterances in language acquisition has been relatively neglected:

> The popularity of the idea that babies are innately gifted to learn language has tended to misdirect investigative attention away from the possible role of parents and other caretakers in systematic nurturance of infants during critical periods of learning.... [I]t may be important how parents react to infant behavior that is deemed communicative. (Oller, Eilers, and Basinger 2001, p. 49)

Oller et al. (2001) suggest that since babbling sequences sound very much like words, parents may entrain infants to associate word-like utterances with objects or events.

> For example, a parent, having heard an infant say [baba], may immediately pick up an object that might be so-named in nursery usage (e.g. a bottle), and present it to the baby, reiterating the term [baba] to highlight the potential association. (Oller, Eilers, and Basinger 2001, p. 50)

Expressing a similar idea, Holzman (1972, p. 312) suggested that "[t]he child finds out by responses of adults what he is assumed to mean by what he is saying, the illocutionary act he is responded to as having performed." There is some evidence that parental attitudes to children's vocal productions change from an initial willing acceptance of burps and kicks as conversational turns to stricter conditions for acceptance (Clark 2003, p. 35). Ninio and Bruner (1978) observed interactions between a mother and her child (Richard) in the context of book reading over the period of time when Richard's first recognisable lexical vocalisations occurred. The responses of Richard's mother to his vocalisations changed over this period. Prior to any recognisable lexical vocalisations, Richard's mother treated *all* of his vocalisations as if he were naming pictures in the book ("yes, it's an X"). The following is an example of such a dialogue between Richard at age 1;1.1 and his mother:

> Mother: Look!
> Richard: (touches pictures)
> Mother: What are those?
> Richard: (vocalises a babble string and smiles)
> Mother: Yes, there are rabbits
> Richard: (vocalises, smiles, looks up at mother)
> Mother: (laughs) Yes, rabbit

Richard:	(vocalises, smiles)	
Mother:	Yes. (laughs)	(taken from Clark 2003, p. 35)

Later, once Richard had started producing recognisable lexical vocalisations, his mother responded to non-lexical vocalisations as if he had named a picture far less frequently. Prior to developing the ability to label pictures in a book, Richard's mother imposed this interpretation on his vocalisations, giving him the opportunity to learn the structure of this social interaction. If adults impose interpretations on children's vocalisations through their responses (i.e. Richard's mother imposing the interpretation that Richard's vocalisations are attempts to name pictures), then there is no *need* to assume that children would spontaneously develop these uses without adult structuring from which to learn.

An adult noun bias Gillette et al. (1999) conducted a study designed to investigate whether nouns or verbs are easier to learn from observation of contexts in which they are used. They presented muted videotaped interactions between mothers and children to undergraduate students. Subjects were asked to guess the word used by the mother at a particular point during the recording, identified by an auditory cue (a beep). Subjects were presented with six consecutive examples of the same word. Gillette et al. (1999) found that the nouns used by mothers were easier to identify than verbs. They (and Gentner and Boroditsky 2001) interpret this as significant to the noun bias in children's early vocabulary (section 3.3.1.1): analogising the responses of undergraduates to children's attempts to infer the meaning of adult utterances, these results may be taken as suggesting it is easier for children to pick up noun meanings. A different interpretation of the significance of this result is in terms of caregiver interpretations of child vocalisations. If adults more readily project a noun-like interpretation onto an unclear word than a verb-like interpretation, their responses to infant vocalisations may be more likely to be noun-appropriate than verb-appropriate. Thus the environment an infant finds itself in may make it easier for them to learn noun–usage–like interactive effects than verb–usage–like ones.

Early entrainment Linguistic interactions between adults and children are rather complex, making it difficult to tease apart the extent to which interpretive structures are imposed on the child's utterance. However, there is evidence that at an earlier age (when interactions are simpler and easier to study) that adults selectively reinforce infant behaviours that can be taken as communicative. Masataka

(2003) describes a series of intriguing experiments concerning this issue. As infants develop, morphological changes alter the range of vocalisations they are able to produce. Prior to the age of about three months, infants are only able to produce rather nasal sounds referred to as "vocalic sounds". At about three months, a number of simultaneous changes to the vocal tract occur which allow the infant to produce sounds with greater oral resonance and pitch variation in addition to the nasal vocalic sounds. Masataka refers to these later, more speech-like sounds as "syllabic sounds".[6]

The proportion of syllabic vocalisations is higher when infants receive caregiver stimulation contingent on their own behaviour (i.e. "conversational" turn taking) than when stimulation is absent or independent of the infant's behaviour (Bloom, Russell, and Wassenberg 1987; Bloom 1988; Masataka 1993; Masataka 1995). Normal caregiver stimulation typically involves speech which should be classified as syllabic (according to Masataka's use). However, the increased rate of syllabic vocalisations does not appear to be due to infant imitation of the noises made by the caregiver as vocalic contingent stimulation still increases the rate of infant syllabic vocalisation (Bloom et al. 1987).

The fact that infants syllabic responses to contingent stimulation are not imitations of that stimulation suggests that these responses may be the result of selective reinforcement, and there is evidence that caregivers do respond differently to infant vocalisations of different quality. Masataka and Bloom (1994) conducted acoustical analyses on the vocalisations of twelve three–month–old infants who were interacting with their mothers. Vocalisations were categorised according to whether the mother responded to them. The vocalisations that were not responded to were generally more nasal than those that were responded to, indicating that these mothers selectively reinforced the more speech like syllabic sounds of their children. Masataka and Bloom (1994) went on to obtain

[6] Unfortunately, Masataka does not use the terms "vocalic" and "syllabic" in the same way as phoneticians (who would not contrast the two, and who would admit some nasal sounds as being "syllabic"). Masataka's use of the terms corresponds to a perceptible (and measurable) difference between infant vocalisations:

> [Syllabic] sounds have greater oral resonance and pitch variation. They are often produced towards the front of the mouth with the mouth open and moving. To adults, these vocalizations sound relaxed and controlled. [...T]he vocalizations produced before the onset of syllabic sounds [are referred to] as "vocalic sounds." In contrast to syllabic sounds, vocalic sounds have greater nasal resonance and are produced towards the back of the mouth. Vocalic sounds are perceived as more uniform in pitch and more effortful than syllabic sounds. (Masataka 2003, p. 66)

Here I follow Masataka's use of these terms.

adult judgements for the responded–to and not–responded–to vocalisations. 100 adults were played audio recordings of the infants vocalising and asked to rate each one in terms of the statement, "this infant is pleasant, fun, friendly, likeable, cuddly," on a five point scale. Vocalisations that had been classified as responded–to were judged significantly more favourably on this scale than those that had been classified as not–responded–to.

Masataka (2003) suggests that adult preferences for infant syllabic sounds is related to the degree to which adults can construe infant behaviour as social participation. Syllabic sounds, being more speech like, make the child seem more like a communicative partner. This preference would depend on what cues an adult attends to when assessing whether others (infants, children or adults) are communicative partners. Support for this idea comes from a comparison between Japanese and Canadian judgements of infant vocalisations. Masataka (2003) suggests that in Japanese culture, the face is attended to less than in Western cultures.[7] On this basis, Bloom and Masataka (1996) predicted that facial visual cues would have a smaller effect on Japanese judgements about infants' sociability than on Canadian judgements. To test this, they videotaped three–month–old infant vocalisations and played them to Japanese and Canadian adults in three conditions (audio–only, video–only, and audio–and–video). Adults judged each video clip in terms of the infant's communicative intent and their social favourability. Vocalisations were classified according to whether the infant's mouth moved, and it was found that mouth movements produced greater judgements of communicative intent and social favourability in the absence of audio cues for adults from both nations. However, the effects were significantly weaker for Japanese adults than for Canadians, as predicted.

Masataka (2003) summarises:

> When something similar to the adult's characteristics occurs in the infant, adults show a strong tendency to over-assimilate them to their own characteristics. The infant also learns to perceive affordances in the adult's responses through such interactions and to match the quality of its own action with the quality of the social situation — as the adult assesses it. Consequently the infant develops the use of

[7] In support of this suggestion, he mentions cultural differences in the strength of the McGurk effect, the role facial cues play in Japanese and American sign languages, and anthropological speculations that the Japanese may reduce arousal by avoiding face–to–face visual inspection during verbal interactions

> behaviors with the characteristics that underlie intentional communication in the true sense, and uses them with adults who then interpret those behaviours as intentional. This could be a first step for the infant in developing intentional communication and in developing cultural variation in their communicative behaviors. (Masataka 2003, p. 88)

Masataka's work suggests that from an early age, modern human infants receive contingent stimulation from adults in relation to how well the infant's behaviour fits the adult's expectations of how the infant would behave were it a communicative partner. However, as Lieven (1994) notes, not all cultures treat infants as if they were communicative partners. While there may be some cultures in which "adults do not even speak to children until the children are using at least some words in a meaningful manner," (Bloom 2000, p. 8), cross cultural studies suggest that language acquisition is affected by social expectations, albeit expectations that vary across cultures (Schieffelin and Ochs 1986). For example, the Kaluli of Papua New Guinea are reported to not speak to pre-linguistic children, but they do speak "for" them (Lieven 1994). From an early age (Schieffelin and Ochs 1986 give examples with a three–month–old infant) Kaluli mothers hold their babies facing outwards so that they can see, and be seen, by other members of the group. Older children address the infant, and the mother responds in a high-pitched, nasalised voice "for" the child. Masataka (2003) compares this register with infant vocalic sounds, and suggests that Kaluli maternal imitation may selectively reinforce aspects of infant vocalisations, albeit aspects other than the syllabic sounds initially reinforced by the Canadian and Japanese adults he studied.

Thus adult attitudes toward infants and their vocalisations may play an important role in early development. This kind of influence may continue into early childhood as adults treat children's utterances as if they had adult–like meaning, thus providing one way in which children can discover the ways they can use words with others. If adult structuring of infant behaviour is important for infants' linguistic development, then children's apparently adult-like usages cannot simply be projected back onto a stage in the evolution of language at which there were no language-proficient conspecifics.

3.4 Summary

Modern humans are immersed in an environment dominated by language users, and may be affected by this environment even before birth (DeCasper et al. 1994). It is thus exceedingly difficult to tease apart the aspects of human linguistic behaviour that can be thought of as independent of the influence of other language users. This chapter has explored the possibility that neuroscience (sections 3.1 and 3.2) or studies of first language acquisition (section 3.3) could provide evidence relevant to inferring communicative systems at earlier stages in the evolution of language, and to showing that the externalisation window has the property of warrantedness. Generally, the conclusion is negative. Our current understanding of the neuroscientific processes which underpin meaningful language use is too poor to construct convincing arguments about the kinds of uses individuals would spontaneously create on the basis of a connection between an internal representation and an external gesture. The evident role that plasticity plays in the development of neural structures disrupts attempts to project back the patterns of usage characteristic of modern languages' meaningful units: the neural structures that underpin such uses may develop through experience of an environment which crucially contains language users. Indeed, the evidence from first language acquisition suggests that interaction with modern humans may play a far greater role than simply the presentation of a series of labels for children to attach to internal representations: both the structure of the language and the expectations of caregivers, formed through interaction with other language users, appear to guide the child's development of the skills of using language.

The externalisation window therefore does not appear to be warranted, and the theories of language evolution discussed in chapter 2 for which the window is crucial rest on unwarranted assumptions. However, if we drop the assumption that the patterns of usage characteristic of modern languages are the product of an externalisation process that can be assumed in language evolution theories, how are we to go about theorising the evolution of language? The following two chapters address this issue.

CHAPTER 4

A Wittgensteinian perspective on language

> That linguistics should continue to be the prerogative of a few specialists would be unthinkable — everyone is concerned with it in one way or another. But — and this is a paradoxical consequence of the interest that is fixed on linguistics — there is no other field in which so many absurd notions, prejudices, mirages, and fictions have sprung up. From the psychological viewpoint these errors are of interest, but the task of the linguist is, above all else, to condemn them and to dispel them as best he can. (de Saussure 1966, p. 7)

4.1 Introduction

If we are to think about language evolution without resting the account on internal representations of meaning, an alternative approach to language and human linguistic behaviour is called for. In this chapter I discuss the approach to language taken by Wittgenstein in his later philosophy. This approach focuses on what we do with language, the activities into which it is woven, rather than seeing the goal of a description of language as showing the relationship between external public forms and internal private entities. By thinking about language as a practice, a number of characteristics of language (which often conflict with the *requirements* imposed on language by linguistic theories) can be enumerated; most importantly for this thesis, Wittgenstein's radical version of linguistic arbitrariness (section 4.2.4.2). Wittgenstein's approach is useful here both as the foundation for the descriptions of language evolution developed in chapter 5, and as a means to clear up certain confusions that can be found in linguistic theorising and which may spill over into theorising about the evolution of language. For example, it might be objected that in spite of the evidence reviewed in the previous chapter, language evolution theories are still justified in using representations of meaning to drive their accounts since we are sure that modern humans *do* know the meanings of the words in their vocabularies, and it is *because* they know these meanings that they are able to communicate with each other. The Wittgensteinian approach will be used in this chapter to examine this kind of statement and hopefully to illuminate the place of such statements in language evolution theories.

4.2 Language games

4.2.1 The Augustinian picture of language

Wittgenstein's *Philosophical Investigations* open with a quotation from Augustine's *Confessions*:

> When they (my elders) named some object, and accordingly moved towards something, I saw this and grasped that the thing was called by the sound they uttered when they meant to point it out. [. . .] Thus, as I heard words repeatedly used in their proper places in various sentences, I gradually learnt to understand what objects they signified; and after I had trained my mouth to form these signs, I used them to express my own desires. (*PI* §1)

It will be useful to refer to this fledgling idea about language and language learning (which mirrors some of the assumptions of language evolution theories discussed in chapter 2) as the *Augustinian picture* of language. This picture presents a language as being a system in which the words are names for objects. While the inadequacies of such a restriction may appear obvious ("jump" for example doesn't name an object) the picture may be generalised to the idea that all words are correlated with a *meaning*, the meaning taking the place of object for which the word stands. (Harris 1998b refers to such an idea as *surrogationalism*: words being surrogates for other things, either objects, events and actions in the world or meanings in the head.)

The Augustinian picture is somewhat nebulous, accommodating different conceptions of what it is that a word might stand for (e.g. a generalised notion of *object* which incorporates e.g. colours, not normally thought of as objects; or psychological entities which combine to form thoughts, etc.). This vagueness is a virtue in the present discussion, allowing for a general contrast between perspectives underpinned by an Augustinian picture which see language as a system of stand–ins, and Wittgensteinian (and Integrationalist, Harris 1981) perspectives which focus on the roles of language in the lives of its users.

One aspect of Augustine's account is that he forms a general picture from the observation of specific cases. The relationship between e.g. the word *table* and the object is blown up into a general account. The craving for generality is a factor which can obscure a clear view of what it is that we call language. We should be wary of this impulse, for taken in its simplest form (i.e. that there is a physical object corresponding to every word *in the same way* as tables correspond to "table"), the Augustinian picture clearly descends into nonsense.

The Augustinian picture is not hard to detect in contemporary linguistics:

> A language is a system for translating meanings into signals, and vice-versa. Thus language is anchored in non-language at two ends, the end of 'meanings' and the end of signals. (Hurford 2007, p. 3)

> [...] a language is a set of semantically interpreted well-formed formulas. A formula is semantically interpreted by being put into systematic correspondence with other objects: for example, with the formulas of another language, with states of the user of the language, or with possible states of the world. (Sperber and Wilson 1995, pp. 172–173)

> It has been recognized for thousands of years that language is, fundamentally, a system of sound-meaning connections... (Hauser, Chomsky, and Fitch 2002, p. 1571)

> What is it that words denote, or **symbolize** as cognitive linguists usually put it? A simple assumption that has guided much research in semantics is that words denote **concepts**, units of meaning. (Croft and Cruse 2004, p. 7, original emphasis)

4.2.2 *Wittgenstein's alternative*

In contrast to the Augustinian picture, a Wittgensteinian perspective focuses on the use of language, its integral role in the various activities human beings engage in. The idea of *language games* helps with this change of focus, a language game being something like a kind of activity in which the use of language is embedded (or: the practices of using language). The idea of language games is not supposed to be a replacement cognitively-flavoured Augustinian idea; for example, that words are translated into internal representations of games. The purpose is to help us *look at* what we do with language, before diving in with assumptions about what must be the case. In particular, Wittgenstein's thinking in terms of language games was directed toward achieving a clear view what we do with the language in which philosophical puzzles (including puzzles about meaning, the mind and the brain) are couched (*PI* §§119–133).

The idea of a language game can serve two distinguishable functions. Artificial, simple language games can be imagined and described which have the virtue of eliminating many of the complexities about our own language that confuse us. These simple language games, Wittgenstein insisted, are not models of the way our language *really* works (or the way it *ought to* work) but are set up as objects of comparison "meant to throw light on the facts of our language by way not only of similarities, but also of dissimilarities" (*PI* §130). The second way the idea may be employed is to approach fragments of our ordinary linguistic practice, bringing to the fore aspects of our language use which may escape attention because of their familiarity, and generally focusing on language use as an activity. Here, our ordinary use of language is thought about through an analogy with games: it would be a mistake to interpret the idea of the language-game played with the word "pain", (*PI* §300) as a claim that such language use is *really* just a game. While some activities in which we employ language clearly are games (e.g. setting and solving riddles), not all of them are and it would be a mistake to inflate the concept of a language game beyond being a tool for thinking about language.

4.2.3 *Simple language games*

The first imaginary language Wittgenstein offers is employed by a builder (A) and an assistant (B). During the course of their building activities, B is to pass A building stones (blocks, pillars, slabs and beams) in the order A needs them. To achieve this, they use a language consisting of the words *block*, *pillar*, *slab*, and *beam*. A calls a word, B fetches a corresponding stone.

In this language there is an association between the word *slab* and a certain kind of building stone. Indeed Wittgenstein's idea was to set up a game which conformed to the Augustinian picture in order to contrast it with our language. But what does it mean to say there is this association? In *PI* §6 Wittgenstein notes that one may mean various things by this, one of which is that the word calls to the imagination a picture of the object (one could add in more contemporary parlance something about the internal representation of the object being "activated"). There is no need to deny that this *may* be the purpose of a word in some situations, but in the language game under discussion it is *not* the purpose of the words to evoke images (or activate representations). And indeed, the evocation of an image is not necessary for someone to understand this language (though we might find that it helps): we would say that B understands *slab* in this language if he fetches slabs when told to (whether or not the image of a slab comes into his mind). That is, what we appeal to to justify the assertion that B understands the order in this simple language is not the image that comes to his mind, nor the presence of an internal representation, but that he generally goes and fetches the right building stone. We will return to the concept of *understanding* later (section 4.3.2), but for now it will do to note that we can justify talk about understanding without having to mention a *mechanism* of understanding by which B manages to bring the right stone or A manages to produce the right call. In this language game, when we say there is an association between the word *slab* and a building stone the association is a feature of A and B's activity, not of whatever mental images or representations they may have.

This language is designed such that all the words operate in one particular way in conformity with the Augustinian picture. However, our language is more diverse in its operation than this. In *PI* §8 Wittgenstein introduces some more words into the language on analogy with some words in our languages. The language now contains words which A uses to tell B how many of a certain kind of building stone to bring, the word "there" used in conjunction with pointing to indicate where B is to put the building materials and colour patches which A uses to show B what colour of building stone he is to bring. These different pieces of the language function in different ways. For example, the words used to indicate how many building stones B is to bring are the letters of our alphabet and are used in the following way: if A says *d-slab*, B goes to the pile of slabs, says the letters of the alphabet up to *d* and for each one takes a slab. He then takes all four slabs to A.

In describing this language game, we may talk about "how this word works" to describe the practices in which that word is involved.[1] This is not to say that when A says *d-slab*, B always fetches four slabs; B may, for example, make a mistake and bring only three slabs (in which case, we may imagine A displays signs of annoyance). The description of the language game is a normative description of what A and B do when they use the language *correctly*.

The normativity of the language game suggests another handle with which to emphasise differences in the use of the words in the simple language game, namely the teaching of words. Words for building stones might be taught by directing the pupil's attention to a building stone and saying the appropriate word. This may be done by pointing at the building stone while saying the word. In contrast, an aspect of training with the 'numerals' may involve learning to say them in the correct order by heart (a technique which has no analogue in the case of the words for building stones). When teaching the 'numerals', there isn't an analogue to pointing to the right kind of object as *any* group of objects can be used for training (e.g. three pillars, three slabs and three blocks could be used to test the pupil on the technique of using the numerals up to *i*). The word "there" is used in this imagined language with the gesture of pointing and so cannot be taught in the same way as "slab" ("there" is not the name of the place to which one points when teaching in the way that "slab" is the name for the kind of stone the teacher points to). Practices of teaching and learning are not here intended to be realistic (this is not child psychology) but serve to emphasise the differences in the techniques for using these words correctly.

4.2.3.1 *The assimilation of expressions*

As noted above, the Augustinian picture can be seen as arising from a temptation to achieve an all encompassing general account of language. The description of the builders' language game above focused on the differences between the use of words, and in *PI* §10 Wittgenstein considers the temptation to think that these differences mask an underlying similarity. The temptation considered arises from the possibility of using the same formula ("X signifies . . . ") with each

[1] This may be contrasted with another way of talking about "how this word works": a description A or B as a physical mechanism and a demonstration of e.g. how B's physical actions follow from the physical input of the order in particular cases.

word. We can say (perhaps to someone who is confused about what the names of each kind of building stone are) "*slab* signifies this building stone, *pillar* signifies that one, etc." And we can also tell someone that the sign "a" signifies a number (we might do this if they asked which building stone "a" signified). Or we might say that "a" signifies the number one (perhaps if someone mistakenly thinks the series goes b, a, c, d,...).

Does the fact that we can assimilate expressions in this manner reflect something about this simple language? The uses of the different kinds of word are (in virtue of the way the language is set up) very different, and clearly the fact that we can talk about what each word signifies (in various ways) doesn't alter the ways the words are used. But might the fact that for each of these words we can talk about what it signifies (or what its meaning is) show that there must be something (perhaps hidden) that is the same for all these words? Wittgenstein considers an analogous move in the description of tools:

> Imagine someone's saying: "*All* tools serve to modify something. Thus the hammer modifies the position of the nail, the saw the shape of the board, and so on." — And what is modified by the rule, the glue-pot, the nails? — "Our knowledge of a thing's length, the temperature of the glue, and the solidity of the box." —— Would anything be gained by this assimilation of expressions? — (*PI* §14)

The *decision* to adopt the formula "X serves to modify ..." as a notational technique for the description of tools clearly does not *reveal* something about tools, nor is it based on an insight into tool use. Similarly, the fact that in English one can use the formula "X signifies ..." to describe the words in the builders' language game does not reveal that there is really something common to the use of each of these words which is perhaps hidden from view.

The importance of this is to remove the temptation to generalise from e.g. the case of words for building stones to e.g. the 'numerals'. The fact that we can talk about "what these words signify" or "what these words mean" may lead us to the erroneous conclusion that details of the use of one kind of word can be transferred over to the other. In this way, for example, we may think that there *ought to be* some kind of thing for each of *a*, *b*, *c* in the way that there is some kind of thing for *slab*, *pillar* and *block*. However, as nothing in the use or the training of numerals so corresponds, this insistence generates the myth that what the numerals stand for is unobservable numbers (perhaps in some ethereal numerical realm, or representations in the mind). Indeed, Jackendoff's idea that words refer to representations in the mind (section 2.2.2.4) can be seen as deriving from the insistence that all words should uniformly bear a particular relationship with a "referent" coupled with the fact that they do not.

In Wittgenstein's view, philosophical problems arise when we take our observations or ideas about how one part of our language works and apply them where they do not belong. We "lay down rules, a technique, for a game, and then when we follow the rules things don't turn out as we had assumed," (*PI* §125). Philosophical problems are dissolved by achieving a clear view of the uses of the words that cause us problems (e.g. "sensation", "meaning", "time", "number" etc.) and accepting that we use them in various different ways.

> One cannot guess how a word functions. One has to *look at* its use and learn from that.
> But the difficulty is to remove the prejudice which stands in the way of doing this. It is not a *stupid* prejudice. (*PI* §340)

4.2.4 Depth grammar

The bulk of the *Philosophical Investigations* consists of carefully chosen reminders our everyday use of various words which have become the focus of philosophical problems and myths (McGinn 1997). Wittgenstein characterised these investigations as "grammatical" (*PI* §90) to emphasise that they were notes on what we do when we use language *correctly* (just as the description of the builders' game is a description of what they do when they do not, e.g., make a mistake). Wittgenstein's "grammar" is more far reaching than traditional notions of grammar in linguistics, encompassing all aspects of language games, not just the construction of sentences. For example, how a particular proposition may be verified is an aspect of that proposition's grammar (*PI* §353).

Many propositions are used in such a way that they could be right or wrong (e.g. "this key unlocks this door"). Whether such a proposition is true is an empirical matter and one can find out e.g. whether this key really does unlock this door. Grammatical propositions, on the other hand, express the rules for correct play in a language game. The contravention of such rules results in nonsense. For example, "every rod has a length" is a grammatical proposition about the words "rod" and "length" (*PI* §251): if one has an object for which there is nothing we call "the length" (e.g. a sphere) then what one has is not "a rod"; conversely, if what one has is correctly called "a rod" (that is, it *is* "a rod") then it has "a length". The truth of "every rod has a length" lies not in an empirical investigation into rods and lengths, but in the rules for using these words. "Some rods do not have a length" does not express a possibility that actual rods in the world do not conform to, but is at most a rule for using these words which we do not follow.

4.2.4.1 Wittgensteinian grammar and grammar in linguistics

Talking about explanations of meaning as giving the "grammar" of expressions may jar with linguists accustomed to a theoretical division of language into various parts (semantics, syntax, pragmatics, morphology) only some of which are governed by "grammar". However, if we relinquish the Augustinian model of language which sees meaning as entities at one end of a chain, we also relinquish a principle on which the division between e.g. syntax and semantics may be based (Baker and Hacker 1984). Waismann (1965, pp. 135–136) offers the following list of rules of English which progress from what would normally be taken to be "grammatical" rules to rules which are less familiar as "grammatical" rules.

(1) The verb "to see" is an irregular verb.
(2) In Latin, the preposition *cum* takes the ablative.
(3) The verb "to master" may only be used transitively.
(4) A proper name cannot be the predicate of a sentence.
(5) The adjective "identical" cannot form a comparative or superlative.
(6) The word "north-east" should not be used in the contexts "north-east of the North Pole" or "north-east of the South Pole".
(7) The words "it is true that ... " should not be used with an adverb of time.[2]

A division on this list between grammatical and non-grammatical rules may be defensible on pragmatic grounds. If one is writing a grammar for people who already speak a language (perhaps one is providing a standard of correctness for their native language, or a pedagogical tool for second language learning) many rules will be unnecessary: a French speaker learning English, for example, will not ordinarily need to be told that the English translation of "nord-est" should be used with a noun designating a place, nor that that noun should not be the

[2] Waismann (1965, pp. 135–136), quoted in Baker and Hacker (1985, p. 58)

English translation of "le pôle Nord" (if a French speaker *did* need to be told these things, we would doubt whether they understood that "north-east" is the English translation of "nord-est", or whether they knew what "nord-est" means).

There are a variety of ways one can choose to draw a boundary through rules of language use to separate those which belong to linguists' grammar from Wittgensteinian grammatical rules. However, failure to see that this is an arbitrary or purpose relative decision may blind us to the kinship between rules of linguists' grammar and other rules, and this can lead to philosophical delusion (taking rules for the use of words to express "metaphysical" truths, such as "it is impossible to travel north east of the north pole" c.f. *BB* p. 55–56) or muddled theories of language such as the following:

> [V]erbs obligatorily assign arguments (one to three), the number of which is predictable from the verb itself. What this means is that if you know the semantics of a verb, you know how many arguments will be obligatorily represented (for "sleep" one, for "break" two, for "give" three and so on). This is true for whatever language you choose: there is no language in which the verb that means "sleep" takes two obligatory arguments, the verb that means "break" takes three, but the verb that means "give" takes only one. (Bickerton 2000, pp. 269–270)

By "obligatory arguments", Bickerton doesn't mean that the arguments have to be explicitly expressed in a clause ("I give blood" doesn't violate any rule) but that they must be, in some sense implicit, or explicitly represented in some structure other than the uttered words. Thus "I give blood" may be paraphrased as "I give blood to the NHS" as it is implicit (in most cases) that it is to the body that receives blood donations I give blood. In cases where it may not be obvious to whom I give blood, I may be asked to expand by providing the third argument ("to whom?"), in response to which (if I am compliant) I will say something like "I give blood to X". In contrast "I broke the chair" does not admit of a similar expanded paraphrase. We could express (part of) this rule perspicuously as: it is right to ask "to whom?" when someone says "I gave x" but not when someone says "I broke x".

The criteria by which we would decide that a foreign word took three obligatory arguments would be part of the *reason* for translating it as "give". If *no* foreign sentence involving the word, when translated into English had three arguments, and attempts to translate English sentences with three explicit arguments into the foreign language produced sentences the native speakers didn't understand, these would be grounds for denying that "give" was a good translation of the foreign word (and so that it meant what the English word "give" means). This too may be expressed as a perspicuous rule: if a foreign word takes only one obligatory argument, then "give" is not a good translation of it. Certainly Bickerton is thinking something like this in the last sentence of the above quotation, as his sweeping statement is not made on the basis of investigation of all the verbs in all the world's languages (past, present and future).

Bickerton's adoption of a semantics/syntax distinction, however, masks these rules by making it look like the number of arguments (syntax) is somehow the *consequence* of semantics. This way of expressing the rule is misleading as it makes it seem as though the semantics of the verb are dissociable from the number of obligatory arguments it takes, but that throughout all languages there is something which prevents the word with the semantics of "sleep" taking two obligatory arguments (perhaps some "innate" or "universal" knowledge or internal representation which prevents this from happening).

4.2.4.2 *Arbitrariness*

The rules of a game may be said to be arbitrary in the sense that they are not justified by anything outside the game. In chess, it is a rule that only the king may be checked. If we were to play a game with different rules (say, either the king or the queen may be checked) the rules would not be inherently *wrong*,[3] rather we would be playing a different game (which might be thought of as a variant of chess). That it is the king that we check in chess is not a fact which is *justified* by mentioning some other property of the chess king from which the rule *follows*. Rather, the fact that it is the king (and only the king) which is checked is (partially) constitutive of the game of chess. In contrast, what we might call rules of cookery (e.g. a recipe) are beholden to the properties of what is being cooked: if one follows the rule, "boil the egg for one second," one cooks badly and the rule is in this sense "wrong". The purpose of cooking is not determined by the cookery rules but by the outcome of cooking, and it is against this that cookery rules are judged good or bad (Z §320). Some games may be judged bad in the

[3] There is a sense in which the rule would be wrong, namely if we presented it as a rule of chess.

sense that they are not entertaining or are too simple, etc., but this does not mean that the rules of these games are poor rules *for those games* in the way that bad rules for cooking an egg are poor rules *for cooking an egg*.

An important aspect of Wittgenstein's concept of language–games is that he thought of grammatical rules as having the same kind of arbitrariness as rules for a game ("The question, 'What is a word really?' is analogous to 'What is a piece in chess?' " *PI* §108). If one imagines rules for the use of a word or expression being different to what they are, one doesn't imagine an incorrect language (or bit of language) but a *different* (bit of) language. For example, in *PI* §556, Wittgenstein imagines a language with two words for negation, "X" and "Y". Doubling "X" yields an affirmative and doubling "Y" yields a strengthened negative, but in all other respects the words are used alike. These differences do not, however, make one word "right" and the other "wrong" in themselves (though if someone were to argue that these words were equivalent to "$\neg$" in our standard logical notations he would be right about "X" and wrong about "Y"). The use of doubled "X" to express an affirmative is not justified by anything outside the conventions of its use (such as the "nature" of negation) but is instead an arbitrary grammatical rule.

As grammatical rules are rules for the use of words, the concept of justification applicable to empirical statements has no place. For example, it is no justification to say "But every rod really *does* have a length." Similarly, it is no justification to say "But there really *are* four primary colours" (*Z* §331) as if one could point to samples of red, green, yellow and blue to verify this statement. For if someone were to object and, pointing to a pink sample, say "surely this is the fifth primary colour" how could we respond? If we conceded pink as a primary colour, we would have changed the concept of "primary colour". But if we were to deny that pink is a primary colour we would have no grounds other than that the only colours we call "a primary colour" are red, green, yellow and blue. We could not appeal to something "similar" which we perceive in the four primary colours as this similarity is nothing more than that these four colours are the "primary colours".[4]

[4] It is no objection here to say that actually there are three primary colours as colour matching experiments only require three different coloured lights (Wyszecki and Stiles 1982). For this is just to use the words "primary colours" according to different rules (e.g. "the primary colours are the colours of the lights used in TV screens"). One might (though it is doubtful anyone does) use the rule: "the number of primary colours is the minimum number of different spectral distributions with which one can match the visual appearance of any spectral distribution in a colour matching experiment" in which case it would be an empirical fact that for human beings

A recurrent example of the misinterpretation of grammatical rules for empirical statements is the interpretation of colour exclusion as a reflection of the properties of colours:

> Relations of hue similarity also have an opponent structure. Red is opposed to green in the sense that no reddish shade is greenish, and vice versa; similarly for yellow and blue. (Byrne and Hilbert 2003, p. 13)

In contrast to the opposition between red and green, orange is an example of a colour that is both red and yellow, purple as a colour that is both blue and red, etc. Thus (8) and (9) are correct, but (10) is incorrect and, moreover, no colour can take place on the right hand side of this colour equation.

(8) 🟥 + 🟦 = 🟪
(9) 🟥 + 🟨 = 🟧
(10) 🟥 + 🟩 = 🟫

Byrne and Hilbert (2003) attempt to account for red/green exclusion by hypothesising a certain mathematical form of colour representation in the brain. However, this seems redundant as there is no difficulty in imagining a language (or a group of people) in which (10) was just as acceptable as (8) and (9). Indeed, a fellow PhD student was surprised to hear that "nothing can be both greenish and reddish" and claimed that something like (10) was correct. *Perhaps* this individual is biologically unusual and has an abnormal visual system (though he isn't colour blind and doesn't seem unusual in any way other than his acceptance of something being both greenish and reddish). However, there is no reason to suppose this. The colours he allows as being "both reddish and greenish" cover a range of shades which he could point to and we could learn. There is no difficulty in assuming that we would be able to learn to identify the colours "he calls 'both reddish and greenish'," and if this is so, there is no difficulty in assuming we (with ordinary human visual systems) could learn to call those colours "both reddish and greenish". The exclusion of colours that are "both reddish and greenish" is a grammatical feature of our colour terms.

there are three primary colours. Obviously, in this usage it is not an empirical fact that primary colours are connected to colour matching experiments. (Incidentally, it is an empirical fact that linking primary colours to colour matching experiments in this way doesn't *determine* a set of three spectral distributions which should count as the primary colours as there are many different sets of three spectral distributions which can be used in colour matching experiments.)

Failure to note the grammatical presuppositions attendant on one's own language can lead to philosophical confusion (or the mistaken belief that citing a grammatical rule constitutes a discovery about the nature of things). For example, the "law of contradiction" ("Nothing can both be and not be," Russell 1912, p. 137) may seem to present an insight into the metaphysical structure of the world: "[w]hat is important is not the fact that we think in accordance with [the law of contradiction], but the fact that things behave in accordance with [it]; in other words, the fact that when we think in accordance with [it] we think truly," (ibid., p. 138). Human languages, it might seem, are constrained to obey the law of contradiction for it is unimaginable that there could be a language in which one can simultaneously assert that something is the case and that it is also not the case; and this is *because* reality accords with the law. However, this is a misleading way of expressing some of the rules for what we call a proposition: "to say that a proposition is whatever can be true or false amounts to saying: we call something a proposition when *in our language* we apply the calculus of truth functions to it," (*PI* §136). Were we presented with a language of which some linguist claimed that it permitted violations of the law of contradiction, we would question whether what he was talking about should be called *propositions* in that language, or whether he was right in saying that one utterance asserted something to be "true" and another something "false". We would investigate the uses of what our linguist claimed to be "propositions", and in giving a description of this use we would, of course, be bound by the law of contradiction. Viewing the law of contradiction as a metaphysical insight into some general fact about the world which constrains our use of language is analogous to viewing the rule "only the king may be checked" as a metaphysical insight into some property of chess from which the rule follows. Of course, a language need not have "propositions" or the concepts "true" and "false" (the builder's slab-language in *PI* §2 does not have this apparatus) just as a game need not have the rule "only the king may be checked" (in which case the game is not chess). The law of non-contradiction applies universally not because reality metaphysically accords with it, but because a language which does not conform to the law is not a language which contains propositions.

In *PI* §497, Wittgenstein mentions an objection that might be raised: "If our language had not this grammar, it could not express these facts." The concern here is that the expression of such facts is something that language somehow ought to do, but without the particular grammar it has this would be impossible. However, "it should be asked what '*could*' means here," (*PI* §497). If the absence of a grammatical rule means that something cannot be expressed, then a language without that grammatical rule is a language in which that particular something *is not* expressed. Consider, for example, Bickerton's protolanguage in which thematic roles are not expressed (section 2.2.2.3). In this language thematic roles such as agent and patient are not expressible, which amounts to saying that in the language games played by the protolanguage users there is nothing which corresponds to our contrast between an agent and a patient. (If there were something analogous, this would be something we could describe, and what we would describe would constitute the difference between expressing agent and patient in the language.) It is important to note that this means we would have no way of justifying an assertion that when ProtoJohn said the words "I saw Og take Ug meat" (Calvin and Bickerton 2000, p. 141) he *meant* what we would mean by saying "I saw Og take Ug meat" in English and not "I saw Ug take Og meat" as all the criteria by which we would judge this (e.g. ProtoJohn's correcting ProtoMary's interpretation) are ruled out by hypothesis (*if* ProtoJohn is able to correct Mary's mistake, he is able to express thematic roles as his correction *is* that expression). (It is no argument to point out that ProtoJohn actually saw Og taking Ug some meat, for describing what someone means is not the same as describing what has happened to them. I do not mean "it's raining and cloudy" when I say "it's raining" on a cloudy day. Had Calvin and Bickerton invented words rather than use the English "saw" and "take" we would have no reason for saying that what ProtoJohn saw could differentiate between the two meanings.)

4.2.4.3 *Justification of arbitrary rules*

Having identified arbitrariness (in a more restricted form than Wittgenstein's) as a fundamental aspect of language, de Saussure went on to state:

> Everything that relates to language as a system must, I am convinced, be approached from this viewpoint, which has scarcely received the attention of linguists: the limiting of arbitrariness. This is the best possible basis for approaching the study of language as a system. In fact, the whole system of language is based on the irrational principle of the arbitrariness of the sign, which would lead to the worst sort of complication if applied without restriction. But the mind contrives to introduce a principle of order and regularity into certain parts of the mass of signs [. . .] (de Saussure 1966, p. 133)

Joseph (2000) argues that the limiting of arbitrariness has been a central theme in much thought about language since Plato's *Cratylus*. The determination to produce justifications for aspects of language (and thereby render them non-arbitrary) has led language theorists to embrace and contribute to various myths from which facts of language supposedly follow naturally. For example, in the *Cratylus*, Socrates discusses (though admittedly does not endorse) the idea that, while words are learned as conventions, they have their origins in an attempt by a lawgiver to correctly put the nature of things into sounds and syllables. Contemporary theorists commonly see (Wittgensteinian) grammatical facts as naturally flowing from some principle in the mind (along with de Saussure in the above quote). For example, Kemmerer (2006) explains the patterns of correctness and incorrectness in examples (11) and (12) by appealing to various kinds of "encoding" in human cognition

(11) (a) Sam sprayed water on the flowers
(b) Sam dripped water on the flowers
(c) * Sam drenched water on the flowers

(12) (a) Sam sprayed the flowers with water
(b) * Sam dripped the flowers with water
(c) Sam drenched the flowers with water

> Human cognition is remarkably flexible, and we are able to take multiple perspectives on events by allocating our attention to the entities in various ways [...]. If I see Sam spraying water on some flowers, I can conceptualize the water as being most affected, since it changes location from being in a container to being on the flowers, or I can conceptualize the flowers as being most affected, since they change state from being dry to being wet. The construction in [(11)] captures the first kind of perspective since it has the schematic meaning "X causes Y to go to Z in some manner," whereas the construction in [(12)] captures the second kind of perspective since it has the schematic meaning "X causes Z to change state in some way by adding Y." Semiotically, the two constructions signal these different perspectives by taking advantage of a general principle that guides the mapping between syntax and semantics, namely the "affectedness principle," which states that the entity that is syntactically expressed as the direct object is interpreted as being most affected by the action [...]. *Spray* can occur in both constructions because it encodes

> not only a particular manner of motion (a substance moves in a mist) but also a particular change of state (a surface becomes covered with a substance). However, *drip* and *drench* are in complementary distribution, largely because each constructional meaning is associated with a network of more restricted meanings that are essentially generalizations over verb classes [. . .]. One of the narrow-range meanings of the first construction is "X enables a mass Y to go to Z via the force of gravity," and this licenses expressions like *drip/dribble/pour/spill water on the flowers* and excludes expressions like **drench water on the flowers*. Similarly, one of the narrow-range meanings of the second construction is "X causes a solid or layer-like medium Z to have a mass Y distributed throughout it," and this licenses expressions like *drench/douse/soak/saturate the flowers with water* and excludes expressions like **drip the flowers with water*.

Kemmerer's attempt to show that (11) and (12) are *not* arbitrary sees the justification in the "meanings" of the two constructions and the meanings of the words. However, this is unconvincing as no way of transcending usage to disclose meaning is provided, and without such an independent handle the description in terms of the meanings of expressions merely reiterates the fact that in English it is wrong to say either (11c) or (12b). Of a language like English which differed only in that any word used in construction (11) could also be used in (12), Kemmerer would not make the distinction between the meanings of the two constructions. Or of a language in which "spray" could only be used in one of the two constructions, Kemmerer would presumably not say that it encoded a manner of motion and a change of state. We do not need to invent such a language as the German words "sprühen" and "besprühen" (which may both be translated as "spray") appear to behave this way:

(13) *Sam sprühte Wasser (auf die Blumen)*
Sam sprayed water (on the flowers)

(14) * *Sam sprühte die Blumen (mit Wasser)*
Sam sprayed the flowers (with water)

(15) * *Sam besprühte Wasser (auf die Blumen)*
Sam sprayed water (on the flowers)

(16) *Sam besprühte die Blumen (mit Wasser)*
Sam sprayed the flowers (with water)

Kemmerer's "meanings" which purport to "explain" (11–12) rely on the very rules they are designed to account for. The explanation is rather like saying "the king is the only piece which is checked *because* the game is chess". If it is objected that German "sprühen" and "besprühen" do not have precisely the same meaning as English "spray" the reply is that this only demonstrates what one is willing to say about whether words in other languages have the same meaning as "spray". Importantly, this objection gives no reason for thinking that other languages will or will not have a word that has "precisely the same meaning as 'spray' ". This attempt to justify rules of grammar by reference to something hidden fails to constrain the possibilities for grammatical rules.

4.2.4.4 Family resemblances and indeterminacy

One aspect of Wittgensteinian arbitrariness is that various different practices of using a word can coexist. That is, in some contexts or situations we may operate with the word in a particular way, while elsewhere we proceed in some other way. Wittgenstein's famous example is the word "game" (*PI* §66): investigation into the various proceedings that are called games reveals a variety of similarities between certain games, overlapping and criss-crossing, but with nothing emerging as *the* essence of "games" or as *the* reason for calling all of these proceedings "games". Wittgenstein characterised the relationships among "games" as "family resemblances", analogous to the various ways in which members of a family can resemble each other (build, features, colour of eyes, gait, temperament, etc., *PI* §67). One can *impose* on the word "game" a formula which appears to reveal unifying essence (e.g., "the logical sum of everything called a 'game' ", *PI* §68, c.f. section 4.2.3.1), but this is not to report a discovered fact, but to simply reiterate the preconceived Augustinian idea that there *must* be some*thing* to which the word "game" corresponds.

The absence of a unifying rationale behind our use of particular words can be appreciated by considering the development of new uses for a word. The term "markedness" as it is used in linguistics relates to various different aspects of languages, related by family resemblances (Haspelmath 2006). The term was coined by Trubetzkoy in 1930 in a letter to Jakobson discussing patterns of phonetic contrast. Joseph (2000) describes Jakobson's exuberant inflation of the concept to cover such diverse areas as the relationships between sin and virtue, between life and death, and between holidays and work days. While Jakobson's expansion of the concept did not generally take hold, various linguists extended its usage in different ways, and the absence of a coherent emergent rationale behind the uses of "markedness" can be seen as resulting from the diversity of individuals adopting the term and the questions they applied "markedness" to (Haspelmath 2006).

While Wittgenstein's observations about "games" and family resemblance concepts are sometimes acknowledged by linguists (e.g. Jackendoff 2002, p. 288), linguists often ignore the fact that these observations apply to the language they use. Indeed, Wittgenstein raised the case of "games" to help clarify a point he was making about "language", namely that there is "no one thing common to [all the various activities we call *language*] which makes us use the same word for all," though there are various relations between them (*PI* §65). In *PI* §23, Wittgenstein instructs us to

> Review the multiplicity of language-games in the following examples, and in others:
>
> Giving orders, and obeying them —
> Describing the appearance of an object, or giving its measurements —
> Constructing an object from a description —
> Reporting an event —
> Speculating about an event —
> Forming and testing a hypothesis —
> Presenting the results of an experiment in tables and diagrams —
> Making up a story; and reading it —
> Play-acting —
> Singing catches —
> Guessing riddles —
> Making a joke; telling it —
> Solving a problem in practical arithmetic —
> Translating from one language into another —
> Asking, thanking, cursing, greeting, praying. (*PI* §23)

Noting that our language contains family resemblance terms (and appreciating how such a situation might come about) should make us suspicious of the *assumption* that something hidden in these activities unites them all on the basis of the fact that we call them all "language". That is, we need not assume that all the various activities we engage in are actually different manifestations of the same activity in each case (such as translating internal representations into external ones). It is possible to adopt a convention that makes it look as though a common thread holds all of these activities together: one can, for example, adopt the convention that every activity we call language is "the operation of an I-language mechanism," but this is not to report a discovery, rather it is to give a rule for the expression "I-language". In the context of theorising about the evolution of language, such terminological stipulations masks the diversity of activities we call "language" and lead the theorist to assume that what we do with language flows naturally from some internal faculty ignoring the possibility that these various activities developed in piecemeal fashion.

Indeterminacy Family resemblance concepts resist analysis into a set of universally applicable conditions for correct usage of a term. In different circumstances we make different appeals to justify our application of a term. A related aspect of language use is that in some circumstances it is not clear what counts as the correct use of a term. For example,

> I say "There is a chair". What if I go up to it, meaning to fetch it, and it suddenly disappears from sight? — "So it wasn't a chair, but some kind of illusion". — But in a few moments we see it again and are able to touch it and so on. — "So the chair was there after all and its disappearance was some kind of illusion". — But suppose that after a time it disappears again — or seems to disappear. What are we to say now? (*PI* §80)

Here we do not have rules for the use of the word "chair". This case is recognisable as one in which we do not know what is correct to say. When we tell someone the meaning of the word "chair" we do not include provisos about cases such as this (and if we did, it would be possible to construct some other kind of case for which our practices of explaining the meaning of the word "chair" did not settle the question). But this is not a shortcoming of the word "chair" (nor of the word "illusion") for the fact that our use of it is not determined in every conceivable case does not impact on its utility in the cases in which it actually is used.

The use of an expression is indeterminate where the rules of grammar do not determine a use. However, it is not always immediately obvious that the rules of grammar have given out. For example, the expression “five o’clock on the sun” (*PI* §350) may at first glance appear unproblematic (I know what “five o’clock in Paris” means, and I know what “on the sun means”) but reflection (or an attempt to use the expression) reveals that while various things can be thought of in connection with the expression, none of them is conventionally *correct*. In section 4.4.3 I will argue that a similar indeterminacy affects linguists’ attempts to produce descriptions of the workings of the brain in producing language: while it may not be immediately obvious, statements such as “the computational mechanism of recursion is recently evolved and unique to our species” (Hauser, Chomsky, and Fitch 2002, p. 1573) are indeterminate as there are no rules governing the description of the brain in computational terms.

4.3 Psychological concepts and language games in linguistics

A mechanistic interpretation of psychological concepts is crucial to the mechanistic interpretations of linguistic theories. For example, in the latest incarnation of Chomsky’s scheme, I-language has to “interface” with the “conceptual-intentional system” which imposes constraints on the forms language can take (Chomsky 1995, p. 2). *Relevance Theory* (Sperber and Wilson 1995) is conceived of by its authors as an account of how internal representations of “assumptions” are generated, manipulated and appraised. Cognitive Linguistics declares that “the representation of linguistic knowledge is essentially the same as the representation of other conceptual structures,” (Croft and Cruse 2004, p. 2). Such mechanistic interpretations are crucial for the externalisation assumptions dealt with in chapters 2 and 3, as it is these supposed internal structures, thought of as corresponding to “beliefs”, “ideas”, “concepts”, etc., which may be thought of as forming one end of the externalisation chain, driving the evolution of language.

This section looks at the language games in which these psychological concepts and some other concepts basic to linguistics operate. The goal is to shed light on the differences between the grammar of these concepts and concepts to which there undoubtedly do correspond physical objects or structures, in order to assuage the feeling that these psychological and linguistics concepts can form part of a mechanistic account of the evolution of language. In addition, looking at the grammars of these concepts will hopefully provide a fuller picture of language. When Augustine described how he learned language, he focused only on words that stood for the objects around him, ignoring the variety of ways in which words operate. When thinking about the evolution of language, there is an analogous danger that we consider only language used to describe concrete situations, thinking that the rest of language will somehow fall into place along the same model.

4.3.1 Folk psychology

One way in which psychological concepts (such as "belief" and "desire") may seem to function like labels for things is through the interpretation of their role as theoretical terms in an explanatory theory of human behaviour (sometimes dubbed "folk psychology"). Churchland (1981) is a prominent example of a philosopher who draws a parallel between the terminology of folk psychology and the terminology of explanatory theories. The motivation for this seems to be a parallelism between statements relating to beliefs and desires (and other "propositional attitudes") on the one hand, and theoretical statements in physics on the other. Drawing the parallel begins with noting that physics has many predicate forming expressions (or at least, you can look at it this way) which have place holders in which numbers belong. Examples Churchland (1981, p. 70) gives are "...has a mass$_{\text{kg}}$ of n", "...has a velocity of n"[5], "...has a temperature$_{\text{K}}$ of n". These predicate forming expressions (Churchland calls them 'numerical attitudes') allow the formation of generalisations concerning law-like relations that hold between these various expressions. Churchland's example is

(17) $(x)(y)(z)$ $[((x$ has mass of $m)\&(x$ suffers a net force of $f))$
$\supset (x$ accelerates at $f/m)]$

Churchland provides a similar notation for propositional attitudes in the form "...believes that p", "...desires that p", "...fears that p", "...is happy that p", etc. Propositions, rather than numbers, are substituted for p. Using this notation, Churchland describes the "laws" of folk psychology as including (18 – 21).

(18) $(x)(p)$ $[(x$ fears that $p) \supset (x$ desires that $\neg p)]$

(19) $(x)(p)$ $[(x$ hopes that $p)\&(x$ discovers that $p)$
$\supset (x$ is pleased that $p)]$

(20) $(x)(p)(q)$ $[((x$ believes that $p)\&(x$ believes that (if p then $q)))$
$\supset$ (barring confusion, distraction, etc., x believes that $q)]$

[5] Churchland's (1981) physics doesn't seem up to much as his omission of the units of velocity makes clear. Velocity is a vector (i.e. has a magnitude *and* direction). He should have used the (far less useful) property *speed* instead.

(21) $(x)(p)(q)$ $[((x$ desires that $p)\&(x$ believes that (if p then $q))$
$\&(x$ is able to bring it about that $q))$
$\supset$ (barring conflicting desires or preferred strategies,
x brings it about that $q)]$

The structural correspondences between (17) and (18 – 21) prompt Churchland (1981, p. 71) to the statement that the "structural features of folk psychology parallel perfectly those of mathematical physics; the only difference lies in the respective domain of abstract entities they exploit — numbers in the case of physics, and propositions in the case of psychology." However, the appeal to the structural similarity between (17) and (18 – 21) does not demonstrate that the language games in which "beliefs" figure parallel those in which "mass" figures any more than the possibility of using the formula "X signifies ..." for each of the words in the builders' language game demonstrates that their uses are the same (section 4.2.3.1).

An obvious failure of parallelism between the cases rests on the exclusion of alternatives in (17) which does not apply to (18 – 21). In (classical) physics, an object only has one mass, one net force and one net acceleration at any one time (in a given frame of reference). The 'numerical attitudes' do not allow multiple simultaneous application of the same attitude to different numbers for a given object (in contrast a person can simultaneously believe many different propositions). The structural similarity between (17) and (18 – 21) disguises this difference[6].

A more significant difference between the use of (18 – 21) and statements of empirical hypotheses lies in the possibilities of confirmation and disconfirmation. The difference can be brought out by considering different ways of using (17): in classical physics, this *can be* taken as an empirical proposition open to disconfirmation, or as a conceptual truth. As an empirical proposition, it would rely on independent notions of mass, force and acceleration: if one was sure that an object had a certain mass (because one had weighed it on the scales prior to the experiment), that the object experienced a certain net force (which one took as being proportional to the extension of the spring to which the mass was attached) and underwent a particular acceleration (which one measured using a stopwatch, ruler and calculator) then one could check whether (17) actually

[6] The difference could be expressed in something like the notation Churchland uses as $(x)(m)(n)[\neg((x$ has mass $m)\&(x$ has mass $n)\&(m \neq n))]$. Would this add anything? Would the absence of a parallel statement for beliefs in folk psychology damage the "perfect" structural parallelism?

holds. Alternatively, one can hold (17) to be true and check one's system of measurements against it (e.g. one might doubt that the mass experienced the force which had been calculated on the basis of the spring's extension). If one trusted one's measurements, (17) could *still* be held true in light of its failure to match what was measured in a particular experiment by appealing to some hither-to unknown force acting on the object (and if it consistently happens, this could even constitute the discovery of a new force in nature).

A simple difference between these two uses of (17) can be extracted and compared to (18 – 21): when Newton's second law is taken as falsifiable, this is because *other* procedures for determining mass, acceleration and net force applied are employed (and held as at least *less* disconfirmable than (17)); in contrast, when the law is taken as an unfalsifiable truth, it can be used as part of the procedure for determining the values of the quantities it relates. Thus the net force acting on a particular body can be measured by finding what force in (17) would produce the observed acceleration of that body (as long as the mass is known).[7]

The way we use words like *belief* and *desire* has more in common with the attitude to (17) which takes it as unfalsifiable. Take, for example, (20) (if someone believes a conditional and its antecedent, they will believe its consequent). If we judge that someone believes a conditional (e.g. "If I leave the soup to boil for more than fifteen minutes it will taste disgusting") and the antecedent ("I have left the soup to boil for more than fifteen minutes") and we judge that the person is neither distracted nor confused, then if we judge that they do not believe the consequent (instead they believe "the soup will be delicious") then we do not take this as disproving (20), but rather as a reason for doubting our judgement.[8] Our judgements of what people believe and desire are defeasible (Bennet and Hacker 2003), so if for example (21) (which relates belief, desire and action) fails to hold in a given situation, this is taken as a reason for questioning whether the person really believed or desired what we thought they did, and *not* whether (21) holds or not.

[7] One could say that the two ways of taking (17) correspond to, or allow for, different meanings of *force*.

[8] There are different points in the chain of reasoning at which we could insert such a doubt, though in this case I imagine it would be most natural to say that the person *was* distracted or confused, or that they hadn't realised, and so didn't believe, that they had left the soup boiling for so long

Whatever grounds are cited in defence of the attribution of a belief or desire to someone, they are generally defeasible. We do not have a technique of determining a person's beliefs and desires, or whether they suffered distraction or confusion, that may trump such relationships as (18 – 21). *In this respect*, the language games involving beliefs and desires is similar to the language games played by the classical physicist who takes (17) as an incontrovertible relationship (for example, expressing the *definition* of a force). The relationships expressed in (18 – 21) are not relationships that have been *found* to hold among people's beliefs and desires (for which we have a different system of determination), nor are they conjectures concerning things (or events, or processes) which we know how to identify but are unsure how to relate.[9] The relationships expressed in (18 – 21) are grammatical rules (as opposed to empirical hypotheses) about the concepts of folk psychology. Knowing that they generally hold is part of knowing the meanings of such terms as *belief* and *desire*; failure to adhere to these rules results in a failure to use these words *correctly*.

Our language games involving the word *belief* do not rest on the existence of some *thing* which is labelled "belief that X", just as the builder's language does not rely on their being something that is labelled "c" for that sign to play the role it does in their activities. Wittgenstein's famous "beetle in the box" example offers a language game to help see that the presence of a *something* as the referent of a noun is not essential to a language game:

> [. . .] Suppose everyone had a box with something in it: we call it a "beetle". No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at *his* beetle. — Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. — But suppose the word "beetle" had a use in these people's language? — If so it would not be used as the name of a thing. The thing in the box has no place in the language-game; not even as a *something*: for the box might even be empty. — No, one can 'divide through' by the thing in the box; it cancels out, whatever it is. [. . .] (*PI* §293)

[9] There is such a thing as uncertainty about what someone else believes, but this is not resolved by doing neuroscience but by various different investigations into e.g. the person's situation, the ways they respond to various events, how they behave when they think they are not being watched, etc. all of which relate to what the person believes by rules like (18 – 21)

In this language game, what is hidden plays no role. The language games with belief do not involve the formation of a hypothesis about the internal state of the believer. Language games which do involve hypothesising about the internal state of something are played by people who know about the internal states of the objects they are looking at, and how they relate to the evidence being cited. For example, a car mechanic might listen to an engine and form a hypothesis about the cause of the car's frequent stalling ("that sounds like a spark plug has gone"). However, since we are generally ignorant of the workings of the brain, yet use the words "belief" and "desire" in rule governed ways, the correct use of those words must be independent of whatever happens in the brain of someone who believes something. For example if someone were to say "I believe John is in the room, but he is not" we would question whether he understood the words they were saying. Were he to reply, "the neurons in my brain are in the configuration that corresponds to the belief 'John is in the room', but actually he isn't in the room" this wouldn't show that what he had said wasn't nonsense (at most it would show that he was using the word *believe* in a different way to us, namely to describe a brain state).[10]

There are, often, differences between telling someone that something is true and telling them that I believe it is true, but the difference isn't one of subject matter: "I believe p" can be used to hedge an assertion that p is the case, or to acknowledge that one's grounds for believing p are not sufficient, or that one's interlocutor may not share the belief (Bennet and Hacker 2003).

We do form hypotheses about other people's beliefs (but not our own) in the sense that the criteria we use for judging someone else's beliefs are defeasible. However, if someone truthfully informs us of his belief (which implies that he understands the sentence he uses to so inform us) then that is what he believes. If someone tells us he believes something (and we believe him), but later it transpires that he doesn't believe what he told us (say, his action is inexplicable given his belief), this simply means that he was not being truthful when he told us, or that his belief has changed in the intervening period, or that he forgot what he believed. It does not mean that the state of his brain was different to, or changed from what we thought: being ignorant of how the brain works with the body, we

[10] One can say "it appears to me that John is in the room, but he isn't" as, for example, the report of a visual illusion. However, in this case, the speaker believes what he reports to be the case (namely that John is not in the room) *in spite of* the impression that what he sees gives him. The use of the distinction between appearance and fact here shows that the speaker doesn't believe all that he sees.

had no idea about what the state of his brain was other than that it was the brain of a person who believed a certain proposition. Wittgenstein (*PI* p. 190) considers the temptation to suppose that when one says "I believe *p*," one is doing something analogous to describing a photograph in order to describe what it is a photograph of (i.e. I describe my belief, conceived of as my internal state, in order to describe what my belief is about). But the analogy disintegrates immediately: I must be able to say that the photograph is a good photograph. The analogue of this in the case of belief just produces nonsense: "I believe it is raining and my belief is reliable, so I have confidence in it."

Conceived of as a tool for use in language games, the word *belief* has rules which, if violated, betoken a failure to understand the concept. When I say "I believe *p*" in many cases this comes to the same as my simply asserting that *p*. Someone who asserts the truth of something contradictory (e.g. "it is windy and raining but it isn't windy") doesn't understand the words he is using (or else is making a joke, or is using a word (e.g. "windy") with different senses). We do not respond by saying "somehow he manages to state something which we cannot think." Similarly, if someone asserts the truth of a contradictory statement, prefixing it with "I believe" this betokens their failure to understand what they are saying, not that their unknown internal belief state somehow represents a contradiction.

4.3.2 Understanding

In the previous section, I made use of the concept of understanding. If someone gives a description of a person using the terms of folk psychology in violation of certain rules (e.g. (18 – 21)) this betokens a failure to understand the relevant concepts. But what about understanding? Isn't *this* a hidden internal process? How do we use this word?

To get a clearer view of the concept of understanding, Wittgenstein (*PI* §151 ff.) introduces a simple case in which we may say there is understanding. In this language game, A writes a series of numbers down, and B is to say "Now I can go on!" when he understands the sequence and could continue it himself. Suppose A writes down the numbers 1, 5, 11, 19, 29; there are a number of different things which might happen with B when he understands. Perhaps as A was writing the numbers, B struggled with various different algebraic expressions (in his head)[11]; after A wrote the number 19 B tried the formula $a_n = n^2 + n - 1$, and the next number confirmed the hypothesis. Alternatively, B might not have thought of algebraic expressions: instead various vague thoughts go through his head, he has a feeling of tension, until finally he asks himself, "what is the series of differences?" He finds the series 4, 6, 8, 10 and says "Now I can go on!". Yet another possibility is that he simply recognises the series, says to himself "oh, its *that* series," and continues it just as if the series had been 1, 3, 5, 7, 9. Or perhaps he says nothing at all and simply continues the series.

Are these processes *understanding*? Take the first case: to say "B understands the principle of the series" is not the same as to say "the formula $a_n = n^2 + n - 1$ occurred to B" as it is easy to imagine that the formula occurred to B and still he didn't understand (perhaps he doesn't understand the formula). If we try the–numbers–B–writes–down as the understanding, we find that they are just the *manifestation* of the understanding, and do not seem criterial: if B writes 41 this might be because he understood, or he might just have been lucky. If he writes 41, 54, still he might just have been lucky (perhaps he *always* writes 41, 54 after any series and in this case it just happens to coincide with what someone who understands would do). In a similar vein: if someone understands me when I tell him to stand perfectly still, he may well stand perfectly still, but you wouldn't say that my table understands me just because it stands perfectly still when I tell it to. There may be characteristic manifestations and accompaniments of understanding, but if we try to pin *understanding* onto any of these we find that we can think them compatible with his not understanding.

The problem is dissolved by thinking of the circumstances in which we say that someone understands something, rather than fixing on the idea that understanding is some process or state which has some characteristic which we spot when we say someone understands. Wittgenstein (*PI* §156) offers an analogy with reading to help here.[12] Consider people learning to read: the beginner may read

[11] "In his head" is an expression we use when talking about calculation to contrast with calculation on paper. It is not based on an observation some physical process that happens in the brain when one calculates without paper, and its adoption tells us nothing about this process. Had we adopted the expression "paperlessly" rather than "in his head" to describe this case, the impression that there must be something equivalent to what happens on paper going on in the brain would not arise. C.f. "Imagine a language in which, instead of 'I found nobody in the room', one said 'I found Mr. Nobody in the room'. Imagine the philosophical problems which would arise out of such a convention," (*BB* p. 69).

[12] This discussion is not about 'reading' as a kind of understanding what is read as in "my reading of Wittgenstein is different to yours." Rather, it is to be thought of as "the activity of rendering out loud what is written or printed; and also of writing from dictation, writing out something printed, playing from a score, and so on," (*PI* §156).

some words by laboriously spelling them out, others he guesses from the context, and in some other cases he already partly knows the passage by heart. In these cases, his teacher will say that he is not really *reading*, and indeed that (in some cases) he is only *pretending* to read.

However, if we consider just *one* printed word, the pupil who is 'pretending' to read may do exactly the same thing as a well practiced reader, and may even have the same thoughts and feelings. Yet still we say there is a difference, that the practiced reader really reads the word while the learner is only pretending. (Suggesting that the difference is a difference in internal machinery is a non-starter here: we know nothing about this internal machinery and yet we still draw the distinction between the pupil's pretending to read and the practiced reader's *actual* reading, and the basis for *this* is what is in question.) What makes the difference (and why it is difficult to see) is illuminated by considering an imaginary case of learning to read from close quarters. Suppose someone is training a pupil to read (*PI* §157). Early in the training the pupil sometimes produces a sound when shown a word, and sometimes this sound happens to be right. If some third person were to overhear such an incident he might say "the pupil is reading" and the trainer would reply "No, he isn't reading; that was just an accident". But suppose that the pupil now goes on to make the right sounds when presented with the next words that the teacher gives him. After a while of this happening, the teacher says, "Now he can read!" But does this mean that he *did* read the first word which the teacher said he didn't? Was that word the last word the pupil guessed or the first word the pupil read? We could stipulate a definition for the first word a pupil actually reads (say, 'the first word of the first series of fifty he reads correctly') but changing the concept here doesn't help clear up the problem (which is about *our* language, not some language invented to make the problem disappear).

If we had *built* a reading machine, we could talk about its design and say in a given case that it only began to read when, for example, a certain lever had been connected to another part. But in the case of the pupil, we have no such language game based on the design of him as a reading-machine (attempts to produce such a way of thinking about him would be derivative on our decision whether he is or isn't reading the word, and so cannot clear up the problem). The change that takes place when a pupil starts to read is a change in the kind of thing he is able to do (when he can read, he is generally able to produce the right sound when presented with a word). This criterion is diffuse, relating to an indeterminate stretch of time (and because of this it makes no sense to talk of "a first word in his new state," *PI* §157).

Returning to understanding: there are many diverse and diffuse criteria which make it right to say of B that he understood the series. If he says the formula, *and has learned algebra, can use formulae to derive series, etc.* we can say he understood even if he then doesn't go on to continue the series. Indeed, he might forget the formula or how to use it in between displaying his understanding (by saying the formula) and the attempt to continue the series. Someone doesn't have an ability *only* when they are exercising it, and someone doesn't understand something *only* when their understanding is manifested in their behaviour (any more than they have a name only when responding to it). This is how we use the words *understanding* and *ability*. In *PI* §182, Wittgenstein sets his reader some exercises to bring to light the complexities of the criteria we rely on for expressions like *fitting*, *being able to* and *understanding*:

> (1) When is a cylinder C said to fit into a hollow cylinder H? Only while C is stuck into H? (2) Sometimes we say that C ceased to fit into H at such–and–such a time. What criteria are used in such cases for its having happened at that time? (3) What does one regard as criteria for a body having changed its weight at a particular time if it was not actually on the balance at that time? (4) Yesterday I knew the poem by heart: today I no longer know it. In what kind of cases does it make sense to ask: "When did I stop knowing it?" (5) Someone asks me "Can you lift this weight?" I answer "Yes". Now he says "Do it!" — and I can't. In what kind of circumstances would it count as a justification to say "When I answered 'yes' I *could* do it, only now I can't"? (*PI* §182)

The comparison of the difficult concept of *understanding* with the more prosaic example of cylinders fitting into each other, helps to alleviate the worry that *understanding* is somehow a mystical peerless concept, referring to something very special and very hidden inside the person who understands. Rather than relating to hidden internal states, the rules for the correct use of "understanding" relate to what is spread over an indeterminate stretch of time, a grammatical feature that also characterises a number of linguistics concepts. The complicated role of *understanding* in our lives means that the fact that someone understands something has many ramifications: someone who understands a particular sentence is different to someone who merely responds appropriately without understanding in that the one who understands may, *inter alia* paraphrase the sentence, explain it to someone else or accept someone else's explanation. As Wittgenstein puts it, "To understand a sentence means to understand a language. To understand a language means to be master of a technique," (*PI* §199).

A central "problem" of much contemporary linguistics concerns understanding: "Empirical linguistics takes the most general problem of the study of language to be that of accounting for the fluent speaker's ability to produce freely and understand readily all utterances of his language, including wholly novel ones," (Fodor and Katz 1971, p. 281, quoted in Baker and Hacker 1984, p. 321). The idea that an individual's capacity to understand sentences he has never heard before requires a special explanation contains a mistake: it is presented as if it could be the case that a person understood only the sentences they had heard before, and that the human deviation from this constraint requires an explanation. However, a person (or a machine) whose responses were confined to a set of sentences they had previously heard (or been taught, or programmed with) would be an example of someone who didn't understand *those* sentences. This would rather be a case of someone who had learned these sentences by rote. The mistake is supported by the notion that understanding a sentence means translating it into an internal representation. If we are clear about how we use the word *understanding* (which is how it is used when it is pointed out that humans can understand sentences they have never heard before) we see that understanding a sentence implies far more than simply that a process takes place when the sentence is heard. It is meaningless to say that a holistic protolanguage could have a unit which means 'five minutes ago I saw a buck rabbit behind the stone at the top of the hill' (Wray 2002a, p. 120) just like that, because whether someone (either the speaker or hearer) understands something as meaning this is a far more involved fact than that certain of his neurons fire in a certain pattern. Without the supporting edifice of a language (in which one could, for example, ask 'wasn't it actually ten minutes ago? You've been with me for the last six minutes') there is no sense in the idea that a holistic utterance could be so understood, and hence no sense in the idea that it could have that meaning.

4.3.3 Meaning in context

It is useful here to note the connections between "meaning" and "understanding". Someone who doesn't know the meaning of a word (and cannot guess) is someone who is unable to understand someone else when they use that word. Explanations of the meaning of a word are given to people who do not understand, and are judged successful when they bring the person to be able to use the word appropriately. As we saw above (section 4.3.2) whether someone understands something on a particular occasion is a more involved matter than simply that something comes before their mind or that they do what someone who does understand would do on that occasion. The criteria we accept for saying someone "understands a word" or "knows a word's meaning" are more diverse and diffuse than an Augustinian picture would lead us to believe, but they are not ineffable. We only say of someone that they understand the meaning of a word on a particular occasion if there are other occasions on which they do (or would) understand the meaning of the word (and these might be various different kinds of occasion including explaining to someone else the meaning of the word). "Meaning" is connected to the *practice* of using a word, not solely to an isolated instance.

> It is, of course, imaginable that two people belonging to a tribe unacquainted with games should sit at a chess-board and go through the moves of a game of chess; and even with all the appropriate mental accompaniments. And if *we* were to see it we should say they were playing chess. But now imagine a game of chess translated according to certain rules into a series of actions which we do not ordinarily associate with a *game* — say into yells and stamping of feet. And now suppose those two people to yell and stamp instead of playing the form of chess that we are used to; and this in such a way that their procedure is translatable by suitable rules into a game of chess. Should we still be inclined to say they were playing a game? What right would one have to say so? (*PI* §200)

The tribesmen sitting at a chess board are not playing a game of chess any more than they would be if they were stamping and yelling their feet. When Wittgenstein notes that "if *we* were to see it we should say they were playing chess" he is noting a mistake we would make. Given that the tribe is unacquainted with games, the two 'players' will e.g. not be able to challenge the legitimacy of each others' moves or cite some standard (perhaps a rule book) to resolve such a dispute. Even if the tribesmen were to say to themselves something like "I'll check him in three moves" or "I'm playing chess" this would not show they were playing a game: how could these words in their mouths mean what they mean in English? To what would we appeal in defence of the translation of the tribesmen's word "playing" into the English "playing" given that they know nothing about games, so the word cannot play an analogous role in their lives as the role it plays in the lives of English speakers. To suppose that they *are* playing chess on the grounds of the similarity between their actions and those of people who really are playing chess is analogous to the supposition that their stamps and yells constitute a game of chess, or that when one scribbles on a piece of paper one is writing because it is conceivable that a writing system might exist somewhere in which that scribble stands for a word. Outside the context of a practice of playing games or of a writing system, what the tribesmen do, feel and say to themselves (and scribble) does not constitute playing a game of chess (or writing a word).

Communication and intention The point of emphasising the dependence on surrounding factors for the language games involving the word "meaning" is to remove the temptation to think that what someone means by what they say is determined purely by some internal mental act or process, as, for example, Chomsky[13] suggests. However, a popular analysis of linguistic meaning (or communication), following Grice (1957), is to identify certain kinds of intention speakers have which hearers recognise when they understand what is said. Sperber and Wilson (1995) dub these *informative intention* ("to make manifest or more manifest to the audience a set of assumptions") and *communicative intention* ("to make it mutually manifest to audience and communicator that the communicator has this informative intention"). What the attribution of these intentions amounts to is, however, rarely discussed. If they are conceived of as some sort of internal state or process, it seems that one might be able to appeal to these intentions as a way of defending the notion that the meaning of what one says is determined by an internal state or process. In *PI* §205 Wittgenstein considers an analogous move:

[13] "[B]ehavior and corpora [are] not the object of inquiry, as in the behavioral sciences and structural linguistics, but merely data, and not necessarily the best data, for discovery of the properties of the real object of inquiry: The internal mechanisms that generate linguistic expressions and determine their sound and meaning."(Chomsky 2007, p. 12)

> "But it is just the queer thing about *intention*, about the mental process, that the existence of a custom, of a technique, is not necessary to it. That, for example, it is imaginable that two people should play chess in a world in which otherwise no games existed; and even that they should begin a game of chess — and then be interrupted."
>
> But isn't chess defined by its rules? And how are these rules present in the mind of the person who is intending to play chess? (*PI* §205)

Resting on the notion of "intention" as an account of how meaning could be divorced from the practice is only convincing as long as the notion of an "intention" is not examined. Even if, as Sperber and Wilson (1995) seem to imagine[14], there were some neural code that corresponded to "I want to play chess" found inside the head (and to remove the occult character of this idea we might as well imagine it written in English on the inside of the skull) this would not show that the subject intended to play chess, because this code could be variously interpreted. Just as the criteria for saying that someone is reading a word aloud on a particular occasion, and the criteria for saying that someone understands a formula on a particular occasion are more involved than simply that they make the right noise or that they produce the right series of numbers on that occasion, so the criteria that someone has a particular intention on a particular occasion are more involved than that they do something that accords with that intention on that occasion. We would not say of a child playing around with chess pieces that they intended to take the queen with the knight even if they moved the pieces in exactly the same way as a practised chess player who did have that intention (*BB* p. 147), and while we may be inclined to describe this as a difference in what is in the mind, the criteria we appeal to in justification of our describing the two cases differently (and hence the criteria for saying that the child and the chess player have some difference in their minds) relate to the circumstances in which the pieces are actually moved.

Another of Wittgenstein's remarks which perspicuously brings out the dependence of an intention (i.e. the dependence of the language games we play with the word "intention") on the circumstances in which the intention is embedded, concerns the invention of a game:

[14] "Let us assume that there is a basic memory store with the following property: any representation stored in it is treated by the mind as a true description of the actual world, a fact. What this means is that a fundamental propositional attitude of belief or assumption is pre-wired into the very architecture of the mind." (Sperber and Wilson 1995, p. 74)

> As things are I can, for example, invent a game that is never played by anyone. — But would the following be possible too: mankind has never played any games; once, however, someone invented a game — which no one ever played? (*PI* §204)

For what I do to count as "inventing a game", it is not necessary (in the circumstances that exist) that anybody should ever play the game I invent, yet in different circumstances these same actions would not constitute inventing a game (just as in different languages the same vocal sounds may be the realisation of different phonemes). Similarly, while it is true that the fact that a person has the intention to play a game on a particular occasion does not necessarily consist solely in their going on to play the game (they need not: their intention may be thwarted), the criteria we appeal to in saying that a person does have this intention may include such things as that they say *and understand* "I want to play chess", and the criteria that they understand this may include such things as that they have learned how to play chess, have played it (or can play it) on other occasions, know of other people who can play whom they may ask, etc. In the absence of the backdrop of a practice of game playing, the utterance "I want to play chess" cannot justifiably be called the expression of an intention, for how would one justify the interpretation of this, or anything the subject did, or any characterisation of his state as relating to chess? The only kind of justification available is of the sort that would identify the tribesmen's yelling and stamping as a game of chess (namely, no justification at all).

If, rather than taking intentions as *sui generis* we remind ourselves that "intention" is an English word which we know how to use, and which we justify our use of in various ways, we see that an intention is no less embedded in a context of practice than the meaning of a word is.

> Language, I should like to say, relates to a way of living.
>
> In order to describe the phenomenon of language, one must describe a practice, not something that happens once, no matter of what kind.
>
> It is very hard to realize this. (*RFM* pp. 335–6)

4.3.4 The reality of the phoneme

The Augustinian picture is important in linguistics not only by characterising many views of what language is, but also by characterising the assumptions linguists seem to make about their own language. The status of the *phoneme* in the 1930s is an illustrative case. The concept of phonemes had been used by linguists trying to describe the sound systems employed by different languages: while the physiology of the human vocal tract can be thought of as providing a universal inventory of sounds/articulatory gestures (phonetic units) languages differ in terms of which phonetic contrasts are employed to contrast different words. Roughly, a phoneme is a grouping of phonetic units together on the basis of the fact that those units are (a) similar and (b) are not used to contrast different words in the language. Twaddell (1935) noted that there were various interpretations of "what a phoneme is" in the writings of linguists at the time which could be broadly lumped into two categories: linguists of the school of Baudouin de Courtenay and the members of the Cercle linguistique de Prague generally thought of the phoneme as having a "mental reality" while Bloomfield, levelling behaviourist criticisms at the utility of such "mental realities" interpreted the phoneme as having a "physical reality". Twaddell found neither of these accounts convincing: evidence cited in support of phonemes being mental realities of the speakers of a language was too easy to interpret in simpler terms, and attempts to account for the physical reality of phonemes seemed premature given that no non-arbitrary way of identifying phonemes on the basis of their physical characteristics had been identified. His solution was to postpone the question of whether phonemes had either mental or physical reality, and instead opt for a kind of instrumentalism[15]: he outlined a method by which a linguist could produce a phoneme inventory for a language, and suggested that this process would produce something useful to linguists irrespective of their opinions about its mental or physical reality. The method he proposed seems to be an idealised version of what linguists actually did when constructing phoneme inventories for languages: find groups of words with small and consistent differences between them (e.g. *beet, bait, bit, bat*) and try to correlate these differences across different groups of words (e.g. *seat, sate, sit, sat*).

[15] Twaddell described his account as "Phoneme as a fiction". However, it isn't clear that "fiction" is an appropriate term. Conducting a linguistic discussion about phonemes is not a discussion of something that is *not the case* in the way that conducting a discussion about unicorns is.

Twaddell's (1935) method makes no claims about the physical nor mental reality of phonemes, but rather attempts to summarise numerous observations concerning significant phonetic contrasts within a language. Attention to the procedures by which phonemes are identified alleviates the pressure to find just *what thing* it is in (whatever kind of) reality that the term corresponds to, for no such entities need be assumed for the operation of Twaddell's method, and for the use of the concept of *phonemes* to have a place in linguists' activities (e.g. presenting a summary of various observations/teaching a language to other trained individuals/contributing to speculation about the historical relations between groups of people). A phonemic description of a language may produce many interesting results (such as that adults show different patterns of categorical perception of speech sounds based on the significant contrasts in their language). One may sensibly pose questions about why a language has the various phonetic contrasts which lead to its characterisation by a set of phonemes, but this does not mean that what one is asking is "what is a phoneme really?" The procedures by which one generates a set of phonemes make it clear that, in spite of the fact that *phoneme* is a noun, it does not play the same role in the language of linguistics as a noun like *table* plays in ordinary life. One can walk into a room one has never been to and point to the table (if there is one), but one cannot take a word in a language one has never heard and say what the phonemes are (though a phonetician can say what the phonetic units are): phonemes depend on the sounds that occur in the rest of the language and without these (and the judgements that two sequences of sounds are either different words or the same word) the concept of a phoneme is meaningless[16] (until we introduce rules for the use of "phoneme" in connection with such a situation).

4.3.5 Sentence structure

Words Linguistic theories in the generative tradition purport to give abstract computational descriptions of the mechanisms which produce grammatical sentences. A basic feature of such models is the conception of a sentence as being an assembly of parts (either words, or some technical idealisation of the concept

[16] As Twaddell's (1935) method relies on choosing sets of words with minimal differences between them it wouldn't escape the fact that there is usually more than one way to 'solve' the problem of coming up with a phoneme inventory for a language (Chao 1934): for it isn't clear how one should decide what these sets are as there are several ways one could operationalise the notion of "minimal difference" none of which is clearly more 'correct' (in a purpose neutral way) than any other. And even once a notion of minimal difference has been decided upon, it isn't clear that the words of a language will fall neatly into coherent sets which can be correlated in a unique manner.

of words, such as morphemes or lexical items). While it is undoubtedly true that sentences are made up of words, what is the grammar of the concept “word”, and is it appropriate for a mechanistic description?

In Wittgenstein’s imagined building–stone language game, A called out *Slab!* and achieved a similar effect to the English *Bring me a slab!* . We could imagine that rather than using the sounds *Slab!* for this purpose, A would say *Bring me a slab!* when B was to bring a slab (but for the other building stones the situation is just as described, with *Pillar!* used to request a pillar, etc.). But this would clearly not make it the case that A said four words when he said *Bring me a slab!* (any more than it would mean we use only one) as what constitutes a word in an utterance is not determined solely by the sounds produced (c.f. *PI* §20). In the language of A and B, *Bring me a slab!* does not contrast with *Bring me a pillar!*, *Bring him a slab!*, etc. and contrasts of this kind are important to the concept of a *word* (though they need not exhaustively account for the way we use this concept: historical accident and preservation through writing also no doubt play a role). The grammar of “word” is spread over a stretch of language, not simply what happens when one utters a single sentence. This can be seen most clearly by considering the fact that the number of words in a sentence could change because of changes to other parts of the language rather than changes to that sentence.

Chomsky (1986) seems to take a different view on the concept:

> Suppose, for example, that a Martian with a quite different kind of mind/brain were to produce and to understand sentences of English as we do, but as investigation would show, using quite different elements and rules — say, without words, the smallest units being memorized phrases, and with a totally different rule system and UG. (Chomsky 1986, p. 28)

What is Chomsky imagining here? Does the Martian “produce and understand sentences of English as we do” or does he rely on memorised phrases? We can imagine what someone (or something) which relied on memorised phrases would be like, but what we imagine when we do this is someone who *doesn’t* produce and understand sentences of English as we do (we could imagine, for example, that we could say a phrase of English to them that they had never encountered and find that they do not understand, or that when we teach them a new word they are only able to repeat the word as part of the phrase in which we presented it). If Chomsky means that the Martian generally produces and understands the sentences we produce and understand (that is, is behaviourally indistinguishable from a native speaker), then it isn't clear what investigation is going to reveal that he is "using quite different elements and rules". Perhaps Chomsky thinks that an investigation into the physical processes by which the Martian produces language will reveal that the Martian doesn't use words. Such an investigation would have to rely on physical tests, conducted on the Martian, which were taken as showing whether or not words were part of the Martian's constitution. However, no such conventionally criterial physical tests exist (as we are largely ignorant of the physical mechanisms by which human beings produce words). Chomsky could arbitrarily stipulate certain physical tests as criterial, but he might as well arbitrarily stipulate that the Martian does not use words. In addition, arbitrarily stipulated tests would open the possibility that some (or even all) English speaking humans did not "use words", and this possibility shows that these tests would not be using the concept of "words" as it is ordinarily used.

If this objection is thought to be based on a misunderstanding of Chomsky's technical use of the word "word", then we have a right to ask what that technical usage is. Saying that "operating with words" doesn't amount to the *ordinary* use of the word "word", raises the question of what it *does* amount to. The question is: how do we know that humans operate with words given that no linguistic theory has been matched up to a brain process? And the answer is that we do know humans operate with words, so the question must be mistaken: this judgement is not based on some fact about a hidden mechanism (but this was obvious just from looking at the way the word "word" operates).

Surprisingly for someone so influential in the field of linguistics, Chomsky has put his name to some rather strange ideas about "words". For example: "Sentences are built up of discrete units: There are 6-word sentences and 7-word sentences, but no 6.5 word sentences," (Hauser, Chomsky, and Fitch 2002, p. 1571). What is strange about this is that it is presented as if the authors know what a 6.5 word sentence would be like, and have discovered that there is actually no such thing, and that this empirical fact characterises language. While it is obviously true that an utterance may terminate mid way through a word, this would not count as the utterance of a sentence with a non-integer number of words as it would not count as the utterance of a sentence. The statement that "there are no 6.5 word sentences" is not an observation about the relationship between independently identified entities known as words and sentences, but is a grammatical rule for the use of the words “word” and “sentence”. Thus following the rule, we should say that (22) has four words, (24) has five words and (23) has either four words or five words (depending on the other details of how one counts words in a sentence) but not four and a half words.

(22) This potato is mouldy
(23) This potato isn’t mouldy
(24) This potato is not mouldy

In every day use of the word “word” the question of how many words (23) contains does not often crop up, and it is obvious to most people that the answer is to ask what is meant by “word” in this context (and sometimes people might allow that (23) contains four and a half words). Adopting a technical sense of “word” (or using the technical term “morpheme”) may make it true that there are no sentences with non-integer numbers of words, but this observation on the (technical) concepts “word” and “sentence” does not constitute a discovery but a stipulation.

Syntactic categories It is a common observation that linguists deal with syntactic categories by looking at the distribution of words in correctly formed strings (i.e. sentences) of a language (e.g. Dixon 1994). Words which can be used in the same environments have the same syntactic category. The environments in which words of different categories appear may overlap without this forcing the analysis that the two categories are really one. I.e. examples (25) and (26) do not lead to the conclusion that *praised* and *clever* are of the same syntactic category, and in defence of the interpretation that they are different one can cite examples (27) and (28).[17]

(25) He was praised
(26) He was clever
(27) He was praised by his wife
(28) * He was clever by his wife

Were it the case that within a language linguists could identify a number of criterial environments which could be used to clearly distinguish classes of words,

[17] There is, of course, far more distributional evidence than this that one can call on to argue that *praised* and *clever* are of different syntactic categories.

assigning words to syntactic categories would be straightforward. However, languages do not seem to present such coherent evidence. Gross (1979) attempted to apply the distributional method to French, aiming at a large scale coverage of the language (as opposed to a small selection of hand-picked examples). He identified 600 rules which he (and his colleagues) tested on 12,000 lexical items to see whether the application of the rule could result in a grammatical sentence of French. He found that no two rules had exactly the same domain of application (lexical items) and no two lexical items had exactly the same distribution (rules).

Croft (2001) suggests the lack of consistent distributional patterns (which would make the assignation of syntactic categories uncontroversial) is a common feature of different languages, and (he argues) leads to *methodological opportunism* among linguists trying to extract systems of categories and rules to describe the grammatical sentences of a language. In the absence of any agreed criteria for what is to count as evidence for syntactic categories (and syntactic analyses) linguists (Croft charges) *choose* what evidence is to count on the basis of whether it supports or undermines their model, and that many disagreements among linguists do not concern the facts about what it is acceptable to say in a language, but how these facts are to be described.

The absence of agreed criteria for what is to count as distributional evidence causes a difficulty for the idea of analysis of a language into 'the true' syntactic categories. This compounds difficulties attendant on the interpretation of these categories relating to some kind of unit in the brain. The same problem infects attempts to analyse different languages using the same categories. If a linguist appeals to certain kinds of evidence in favour of a certain cross linguistic identification of categories, then the absence of that evidence in a language presents a problem. For example, parts of speech are commonly identified in different languages on the basis of inflectional criteria (inflections for number, case and gender are used to identify nouns and inflections for tense, aspect, agreement and mood are used to identify verbs). However, Vietnamese lacks all morphological inflection, so inflection cannot be used to identify parts of speech in Vietnamese (Croft 2001). The parallel problem of a language in which the basis for a system of classification is identifiable, but the resulting classifications bear little resemblance to other languages, is also highlighted by Croft. His examples is Makah, which may be analysed as having inflections marking Person-Aspect-Mood on words which, for other (e.g. semantic) reasons, may be likened to English verbs, nouns and adjectives. Cross linguistic syntactic analyses are, thus, also prone to methodological opportunism, with linguists adopting different criteria for different analyses of different languages.

These features of the language game of syntactic analysis and the communication systems it is applied to should not be taken as warranting a wholesale dismissal of the idea of describing the syntax of a language. Such a description may achieve its purpose in spite of the fact that it cannot be wholly consistently applied to a language without some arbitrary decision on the part of the theorist. The point here, though, is the purpose of the analysis. If, for example, one wants to create a tool to help people learn a new language, the fact that the rules one gives for the word categories one operates with may in some cases be incorrect does not detract from the fact that such an analysis may help the pupil master (much) of the language faster than if this system were not available. Similarly, as Haspelmath (2007) argues, the absence of cross linguistic coherence in syntactic categories does not mean that nothing interesting can be said when comparing languages, nor that we might not see commonalities across languages which we may find puzzling and hope to produce a plausible account for. However, the differences between languages and the inconsistencies within languages do undermine the notion that there is any evidence of a monolithic underlying cause of syntactic structure. Thus, I think we should read the following statement by Chomsky as expressing a determination to *create* a descriptive system into which different languages can be forced rather than as reflecting any kind of discovery:

> The primary [task] is to show that the apparent richness and diversity of linguistic phenomena is illusory and epiphenomenal, the result of interaction of fixed principles under slightly varying conditions. (Chomsky 1995, p. 8)

The criteria for a sentence being composed of certain words, and those words having particular syntactic categories, and consequently the "structure" of the sentence relate to the other sentences in the language. That is how this game proceeds, not by reference to some hidden structure. The interpretation of sentence structure as a description of the physical mechanisms operable when a person speaks a sentence is an *additional* assumption, orthogonal to the criteria in the language game of describing sentence structure. Whether such a move can be used to make significant statements about the human–body–as–language–mechanism and whether such theories can inform language evolution theorising will be the subject of section 4.4.3.

4.4 Abstract descriptions

4.4.1 *Meaning* in the mind*

The interpretation of our ordinary use of psychological concepts, or linguists' talk of the structure of a sentence as statements about hidden internal states or processes is a grammatical confusion: the grammar of these concepts is independent of such hidden entities. However, it may be objected that while it might be mistaken to identify the meaning of a word something internal, surely someone who knows the meaning of a word (or knows the rules of various language games) must have something internally guiding their action. If so, perhaps it is a valid move to use the term "meaning" in a technical sense (*meaning**) to refer to this mind-internal factor: Someone who understands the meaning of a word has the word's *meaning** which is the source of their use and understanding of that word.

While the details of what this kind of *meaning** might be are here intentionally left vague[18], various linguistic theories which purport to be abstract descriptions of the mind/brain employ such entities. Chapter 3 reviewed evidence pertinent to the use of such entities in language evolution and concluded that there was no evidence that anything in the brain would play the simple role assumed by theories of language evolution which invoke *meanings**. The argument there was that no coherent neural structures emerge from studies aimed at identifying the neural correlates of meaning, and that there was no evidence to suggest that in the absence of interaction with language–proficient humans, modern–like language use would spring forth from individual psychology. However, mightn't we risk throwing the baby out with the bathwater if we deny a role for *meanings** in our account of the evolution of language on the basis of their complexity and the fact that they *alone* are unable to produce meaningful language in an evolutionary sense? Can't we be sure that proficient speakers have these *meanings** even if we don't know exactly what they are (just as European explorers in Africa were sure that the Nile and the Congo had sources and that these were different places in spite of their location being unknown, Hurford 2007, pp. 15–16), and so aren't we justified in building models in which they play *some* role?

[18] They could be unitary or composite, neural structures or neural activity, specific to language or involved in various cognitive activities, etc. etc.

The idea behind the *meaning** of a word is that it makes me use the word as I do. But *how* does this internal *meaning** manage this? In *PI* §139 Wittgenstein considers the word "cube":

> What really comes before our mind when we *understand* a word? — Isn't it something like a picture? Can't it *be* a picture?
>
> Well, suppose that a picture does come before your mind when you hear the word "cube", say the drawing of a cube. In what sense can this picture fit or fail to fit a use of the word "cube"? — Perhaps you will say: "It's quite simple; — if that picture occurs to me and I point to a triangular prism for instance, and say it is a cube, then this use of the word doesn't fit the picture." — But doesn't it fit? I have purposefully so chosen the example that it is quite easy to imagine a *method of projection* according to which the picture does fit after all.
>
> The picture of the cube did indeed *suggest* a certain use to us, but it was possible for me to use it differently. (*PI* §139)

The idea that some kind of picture in the mind underpins individuals' use of a word is not uncommon in linguistics (e.g. "image schemas" in cognitive linguistics, Lakoff and Johnson 1999; Evans and Green 2006). However, the point that a *meaning** can be variously interpreted and so admit various uses of the word to which it is attached is not restricted to the somewhat homuncular vision of me *comparing* my internal *meaning** with an application. The point applies also to a mechanistic interpretation of a *meaning**: if a *meaning** is supposed to be some kind of physical structure (say, a pattern of synaptic weights) then it will only manage to produce the right use of the word in the context of the rest of the system (" 'I set up the brake by connecting up rod and lever' — Yes, given the whole rest of the mechanism. Only in conjunction with that is it a brake-lever, and separated from its support it is not even a lever; it may be anything, or nothing," *PI* §6).

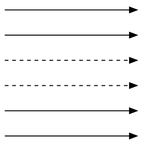
In *PI* §141, Wittgenstein considers a possible response to this kind of qualm. Perhaps it is not only the *meaning** (the picture of the cube) which comes before my mind, but also some rule for its use. For example, the method of projection by which the picture is a picture of a cube. But if we flesh this out, we see it gets us no further: for if we imagine a schema for the method of projection comes before someone's mind (say, in the form of a picture of a picture of a cube connected to a cube by lines of projection) we can also imagine various different applications of *this* image.

For each rule we give for the application of the *meaning** of a word, if we conceive of the rule as something appearing in the mind (on the model of the conception of *meaning** as something appearing in the mind), it will admit of different interpretations. Positing a *meaning** to guide my application of a word does not solve the problem of how it is that I use the word, but defers it to the problem of how it is that I use the *meaning**.

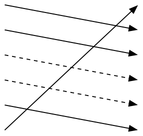
Consider the case of a table showing the names of colours. Such a table might look like this:

red	🟥
blue	🟦
green	🟩
yellow	🟨
purple	🟪

Part of the technique of using this table may involve hearing a word, running one's finger along to the colour, and then holding the colour patch against an object to see if they are the same (or similar) in colour. That is, the use of this table is governed in part by a rule which we could express like this



This expresses the rule according to which the table is used, and contrasts with a different way of using the table which we could express like this:



However, in this case we are far less tempted to suppose that something like the first set of arrows *has to* come before the mind of the person who uses the table (and, of course, if we were tempted to think this we might also want to add something which shows the user how to apply the arrows that come before his mind). This game of giving a rule formulation which directs the use of something else clearly has to come to an end at some point, and at that point rather than giving another interpretation which comes before one's mind we will have to say something like "this is just what we do". That is, even if we are inclined to posit something in the mind which directs our actions (e.g. our use of a word) there will come a point in our explanation at which we *stop* giving such entities to explain the use of the last one.

This can be seen in terms of the grammatical relationship between "following a rule" and "having a rule formulation": while a rule formulation may sometimes be involved in following a rule, following a rule does not necessarily imply that one has a rule formulation. That is, one can follow a rule *without* one's actions being guided by a rule formulation and so it *doesn't* follow from the fact that "when one uses language correctly one is engaged in rule-governed behaviour" that one *must* be guided by some rule formulation (mentally represented or otherwise).

The term *meaning** is a neologism to which we can give whatever grammar we choose. We can choose to say (expressing a rule) "someone who knows the meaning of a word has the *meaning** in their mind when they use it." But this gets us no closer to understanding the physical processes by which the human machine produces language: compare our ignorance before introducing the formulation with our ignorance afterwards. Crucially for this thesis, such a *decision* cannot help establish the character of communication at earlier stages in the evolution of language. (Of course, it is conceivable that we might *discover* that the brain contains entities which drive meaningful language use in the way *meanings** are supposed to, but the evidence reviewed in chapter 3 suggests that this unlikely to happen).

4.4.2 Cognitive prerequisites as grammatical rules

> We are treating here of cases in which, as one might roughly put it, the grammar of a word seems to suggest the 'necessity' of a certain intermediary step, although in fact the word is used in cases in which there is no intermediary step. Thus we are inclined to say: "A man *must* understand an order before he obeys it", "He must know where

> his pain is before he can point to it", "He must know the tune before he can sing it", and suchlike. (*BB* p. 130)

A number of language theorists claim to have identified certain cognitive pre-requisites for language (e.g. possession of a linguistic theory, Chomsky 1965, or powerful forms of intention reading and cultural learning, Tomasello et al. 2005). As pre-requisites, it might be thought that their presence must be assumed prior to the emergence of language in our species, and so can be used to give insight into the cognition, and thereby behaviour, of our ancestors at distant stages in the evolution of language. However, in this section I argue that these cognitive prerequisites for language are, like *meaning** as a prerequisite for using a word, grammatical features of an adopted set of descriptive practices. The grammatical nature of the relationships between these prerequisites and language use is suggested by the certainty with which they are *known* by their proponents to be true in spite of our poor understanding of humans as physical language producing mechanisms.

For example, Damasio et al. (2004) see an unassailable connection between naming and concept retrieval. Describing their method for scoring patients responses to pictures presented to them they say:

> [If] the stimulus was named correctly, the item was scored as a correct recognition and naming. In other words, we accepted correct naming as unequivocal evidence of correct recognition. (Damasio et al. 2004, p. 185)

Subjects who did not name the stimulus correctly could be scored as having recognised it if their responses allowed a third person to identify the stimulus. The decision to accept correct naming as an instance of correct recognition (that is, to use "naming" and "recognition" in these ways) excludes the possibility of a subject being attributed with correct naming but not recognition (this grammatical decision reflects the ordinary grammar of these terms). However, Damasio et al. (2004) go on to present the decision to exclude the possibility of naming without recognition as if it were a pragmatic response to the empirical fact that no subjects ever named a stimulus but didn't recognise it:

> Our rationale for this approach is that we have never found a subject who would produce a correct name, and then fail to recognize the

> stimulus that was named. In background work for this set of studies we explored in a subset of subjects with correct naming responses whether they had retrieved the concept for an item prior to retrieving its name, and, as we expected, they had. In other words, we never encountered a patient who would, for example, name Jane Fonda when presented with her picture and then say "who is Jane Fonda?", or see a broom, call it a broom, and not know what a broom is, or what it is made of, or what it is used for. (Damasio et al. 2004, p. 185)

This looks like an empirical claim as they give a description of what naming without recognition would be like. However, as the quotation continues, the reason why this was never observed becomes apparent:

> In fact, we do not believe it is possible for someone to name, accurately and reliably, an unrecognized item, even in the extreme instance of patients with Alzheimer's disease who may on occasion appear to do just that. A severely inattentive or demented patient may produce a correct naming response and, by the time the response comes under scrutiny, may have lost from working memory the material recalled during concept retrieval. It may appear that the patient has 'named but not recognized', but this is an artifact of the attentional/working memory defect, and we remain convinced that concept retrieval is a prerequisite for accurate naming. (Damasio et al. 2004, p. 185)

Thus even if a patient were to correctly name Jane Fonda and then ask "who is Jane Fonda?" this would not be a case of naming without recognition, but "evidence" of the volatility of that patient's short term memory. That is, the relationship between recognition and naming in Damasio et al.'s study lies in the rules for using those words, not in any empirical discovery.

4.4.2.1 Recognising intentions

In an influential paper, Grice (1957) presented a distinction between natural and non-natural meaning. He introduced these kinds of meaning with a number of examples: "Those spots mean measles," and "The recent budget means that we shall have a hard year," are examples of natural meaning while non-natural meaning is exemplified by "Those three rings of the bell (on the bus) mean that the 'bus is full'," and "That remark 'Smith couldn't get on without his trouble and strife,' meant that Smith found his wife indispensable." Grice proposed that the hidden rationale behind the intuitive difference he detected related to the intentions of individuals who non-naturally meant something. Grice did not consider it sufficient that A intends to produce an effect on (or in) B by an utterance: Grice might leave a handkerchief near the scene of a murder in order to make the detective believe that C was the murderer yet the handkerchief does not non-naturally mean anything. Nor will it be enough that A intended that B recognise A's intention to have an effect on B: when Herod presents Salome with the head of John the Baptist on a charger, he no doubt intended Salome to recognise that he intended her to believe that John was dead, yet the severed head did not non-naturally mean that John was dead.

To clear up the problem, Grice contrasts two cases: (1) Grice shows Mr. X a photograph of Mr. Y displaying undue familiarity to Mrs. X and (2) Grice draws a picture of Mr. Y behaving in this manner and show it to Mr. X. Grice (adding to our knowledge of what he means by non-natural meaning) tells us "I find that I want to deny that in (1) the photograph (or my showing it to Mr. X) non-naturally meant anything at all; while I want to assert that in (2) the picture (or my drawing and showing it) non-naturally meant something".[19] The difference, according to Grice, is that in the first case the photograph would have had the effect on Mr. X whether or not he recognised the intention to produce the effect (Grice might as well have left the photograph lying around) while in the latter case what intention Mr. X thinks Grice has (e.g. whether it is to inform or just to doodle on some paper) makes a difference to whether the effect is produced. Thus for something to non-naturally mean something for Grice, the recognition of the intention to produce the effect by means of the signal must be a necessary condition for the production of the effect.

However, Harris (1996) complains that this is a strange conclusion to reach for some of Grice's introductory examples:

[19] Grice's (1957) paper has a peculiar structure: the natural/non-natural distinction is a neologism of Grice's, so his audience has no authority to question his word on whether a particular example is natural or non-natural meaning. (There isn't an existing regular practice or acceptable explanation of meaning one can use to appraise Grice's decisions.) If one were to disagree with Grice, and perhaps say that his leaving a handkerchief for the detective *was* a case of non-natural meaning (perhaps *because* Grice intended the detective to reach a particular conclusion) then Grice could simply respond that the objector was thinking of some other distinction than the natural/non-natural distinction. The price paid for this authority is that Grice can simply stipulate that his analysis is correct by using *it* to determine whether cases are natural or non-natural meaning.

> If [...] I tell someone that the conductor, by ringing the bell three times, meant that the bus was full, in no way do I mean or intend to imply that the driver's recognition of the conductor's communicational intention was a necessary condition for its fulfilment. The reason for denying this is simple: I do not in fact think it was necessary. Nor do I think that the bus conductor would think so, unless confused by having taken an evening course in philosophical semantics. (Harris 1996, p. 54)

Grice (1957, p. 385) explicitly states that "for x to have non-natural meaning, the intended effect must be something that [...] in some sense of 'reason' the recognition of the intention behind x is for the audience a reason." The question is, in *what* sense of "reason"? Harris's complaint seems to focus on the fact that the bus driver would not say "I recognised that in ringing the bell three times the conductor intended to make me believe that the bus was full" when asked for the reason why he believed the bus was full, and nor would he think to himself "conductor intends me to believe..." when hearing the three bell rings. (More likely, when asked he would say something like "the conductor told me the bus was full" or "the conductor rang the bell three times, and ringing the bell three times means the bus is full", neither of which can be paraphrased in terms of the conductor's intention unless one is *already* committed to Grice's analysis.)

However, a Gricean could respond by pointing out that if, for example, the bus driver had thought that the conductor was having a joke, he would not have believed that the bus was full. And so when he understands the three bell rings he recognises the intention in the sense that he doesn't think the conductor has any *other* intention. What shows that the driver doesn't think that the conductor has some other intention (i.e., a criterion in this case) is that the driver believes that the bus is full because he heard the three bell rings (the criterion is *not* that he would give an explanation in terms of intentions, or that he would think to himself something about the conductor's intentions). The connection between the recognition of the communicative intention and "understanding when A means (meant) something by x" is thus not a discovered or hypothesised empirical connection, but a grammatical rule (adopted by Grice but rejected by Harris).

A grammatical relationship between meaning and recognition of intentions appears in various theories of language, albeit in slightly different forms (Sperber and Wilson 1995; Clark 1996; Tomasello et al. 2005). While Grice (1957,

p. 386) “disclaim[ed] any intention of peopling all our talking life with armies of complicated psychological occurrences,” many theorists whose work has inherited Grice’s grammar mistake the relationship between meaning and recognition of intentions as a mechanistic process taking place in the brain which requires “[species] unique forms of cognitive representation,” (Tomasello et al. 2005, abstract). This is a misinterpretation which suggests that these mechanisms have been discovered and may be attributed to human ancestors at earlier stages of the evolution of language. However, just as with the decision to adopt a particular grammar for *meaning**, grammatical relationships between meaning and recognition of intentions do not alter our knowledge of what physically happens within a human being when he uses some bit of language.

In terms of the evolution of language, we know (if we adopt such a rule) that language has always been used by individuals who could recognise each others intentions. However, this does not get us any closer to an account of the history of language as this fact is true no matter what took place in the past. Crucially, this grammatical fact does not help us understand how humans or our ancestors would have behaved in environments which lacked modern–language–speaking humans.

4.4.3 Describing the brain abstractly

> You think that after all you must be weaving a piece of cloth: because you are sitting at a loom — even if it is empty — and going through the motions of weaving. (*PI* §414)

This chapter has looked at a number of concepts deployed in linguistic theorising from a Wittgensteinian perspective, and has argued that, grammatically, these concepts do not act as surrogates for hidden entities or processes in the brain. A language evolution theorist might accept that words such as “word”, “concept”, “meaning”, etc. do not straightforwardly function as labels, and that the grammar of following a rule does not force us to describe a rule formulation to be consulted, yet still maintain that linguistic theories have produced empirical hypotheses about the way the brain works (even if those hypotheses sometimes appropriate expressions like “meaning” as labels for internal entities). Successful linguistic theories, it might be argued, represent relatively well confirmed hypotheses from which behaviour at earlier stages of the evolution of language can be deduced.

The idea that linguistics produces mechanistic descriptions of the production of language (whether these are models of the mind, the brain, or their conflation, the mind/brain) is common to a number of differing linguistic theories:

> [The] steady state of knowledge attained and the initial state S_0 are real elements of particular mind/brains, aspects of the physical world, where we understand mental states and representations to be physically encoded in some manner. The I-language is abstracted directly as a component of the state attained. Statements about I-language, about the steady state, and about the initial state S_0 are true or false statements about something real and definite, about actual states of the mind/brain and their components. (Chomsky 1986, p. 27)

> The first hypothesis is that language is not an autonomous cognitive faculty. The basic corollaries of this hypothesis are that the representation of linguistic knowledge is essentially the same as the representation of other conceptual structures, and that the processes in which that knowledge is used are not fundamentally different from the cognitive abilities that human beings use outside the domain of language. (Croft and Cruse 2004, p. 2)

> Language offers a window into cognitive function, providing insights into the nature, structure and organisation of thoughts and ideas. The most important way in which cognitive linguistics differs from other approaches to language [. . .] is that language is assumed to reflect certain fundamental properties and design features of the human mind. [. . . An] important criterion for judging a model of language is whether the model is psychologically plausible. (Evans and Green 2006, p. 5)

> Concepts are the elementary units of reason and linguistic meaning. They are conventional and relatively stable. As such, they must somehow be the result of neural activity in the brain. (Gallese and Lakoff 2005, p. 455)

However, in spite of the ubiquity of this position in linguistics (or perhaps because of it), what it means to say that linguistic theories describe human machinery, or that a linguistic description may be "abstracted directly" from the state of a language users brain, is somewhat opaque. Contemporary popularity for this kind of idea is often traced to Chomsky's (1959) attack on Skinner's (1957) behaviourist description of language, but the rejection of behaviourism does not produce an explanation of what it means to call one's theory a model of the mind or brain (beyond the double negative rejection of the behaviourists' rejection of such models).

This section looks at what the claims that linguistic theories are abstract descriptions of the mind and/or brain can amount to with reference to the evolution of language. Attention will be restricted to the claim that linguistic theories somehow produce an abstract description of the physical mechanisms which lie behind language production and/or reception. A theory which sees itself as a model of the mind, but then is agnostic about the relationship between the mind and the body will not be discussed as this idea leaves obscure the question of how such a model could inform us about the (proto)linguistic practices of our ancestors at earlier stages in the evolution of languages.

4.4.3.1 Instantiation

The notion that a linguistic theory is an abstract description of some physical process may be cashed out in terms of a bridge theory that links statements of the linguistic theory to statements of a physical description of the mechanisms of language production. Fodor (1974) discusses this kind of relationship between a "Special Science" (e.g. psychology or linguistics interpreted as an abstract description of the brain) and physics. Fodor (p. 98) characterises a law of a special science as

(29) $S_1x \rightarrow S_2x$

Which is intended to be read as something like "all S_1 situations bring about S_2 situations" (though in the Special Science the "all" is to be taken with a pinch of salt as Fodor assumes such laws are not exceptionless). S_1 and S_2 are predicates of the Special Science and are not predicates of what Fodor calls "basic physics". Bridge laws are introduced as (30):

(30) (a) $S_1x \rightleftharpoons P_1x$

(b) $S_2x \rightleftharpoons P_2x$

(31) $P_1x \rightarrow P_2x$

P_1 and P_2 are predicates of physics, and (31) a law of physics or a consequence of the laws of physics.

When a linguist claims to have produced an abstract description of the workings of the brain it is unlikely that he will impose a restriction that, e.g., a meaning representation will correspond to a basic element of neuroscience (say, a neuron) or physics (say, an atom). Similarly, a linguistic theory is not generally restricted to the description of only one kind of stuff (brain stuff), but may be thought applicable to anything (perhaps a Martian's innards) which produces language "in the same way" as human beings. Therefore, we may allow the bridge laws to relate linguistic–theory–descriptions to complex physical–description–theories which may be "satisfied" by more than one physical brain state (c.f. Horgan and Woodward 1985).

For example, a linguistic theory statement such as "the lexical item 'square' has the value [-colour]" might correspond to a neuroscientific description that runs like "neuron 2134 is connected to neuron 34 with a connection strength weaker than α and neuron 7723422 is connected to ..."[20]. (N.B. being a *complex* description, this may be satisfied by numerous physical states which might in other contexts be thought of as different.) Within the linguistic theory, such a statement will have certain consequences (for example, "the sentence 'paint it square' has acceptability level 0.1"[21]) which themselves can be related to complex physical descriptions via bridge statements.

These relationships for a particular part of a (toy) linguistic theory are diagrammed in Fig. 4.1. If we have a linguistic theory, a complete set of bridge rules and the ability to calculate what future physical states of the brain will be from descriptions of the current state, then we will be able to show that the linguistic theory makes the right predictions: bridge laws would convert the linguistic theory description on the left into complex physical descriptions; with our physical theory we could calculate the physical state the brain would develop into; with our bridge laws we could convert this physical description

[20] Neuroscience as it stands does not use a numbering scheme to identify individual neurons like this. This is supposed to be a statement from some imaginary future super-neuroscience.

[21] N.B. none of this is supposed to represent any real linguistic theory, but is set up as a toy example.

$$\begin{tikzcd}
\text{Item ‘Square’ has value [-colour]} \arrow[r, phantom, "\&"] \arrow[d, phantom, "\downharpoonleft\upharpoonright"] & \text{“Paint it square” presented aurally} \arrow[r] \arrow[d, phantom, "\downharpoonleft\upharpoonright"] & \text{Acceptability score: 0.1} \arrow[d, phantom, "\downharpoonleft\upharpoonright"] \\
\text{Neuron 2134 connected to neuron 34} \ldots \arrow[r, phantom, "\&"] & \text{Auditory neuron 842 stimulated at frequency } f \ldots \arrow[r] & \text{Neuron 8573 firing rate below } \beta \ldots
\end{tikzcd}$$

Figure 4.1: Linguistic theory and physical descriptions running in parallel

into a linguistic description and verify it as the prediction made by the linguistic theory.

It is worth noting that this account parallels a certain kind of description that is often invoked to explain the idea of an abstract description, namely the description of the workings of a computer. The workings of a computer can be described at a number of different levels: the physical structure of the machine, along with patterns of electrical charge across certain components may be used to work out, e.g., what the distribution of charge in the machine will be in the near future. We may also abstract away from the actual levels of charge in the computer's components to a description in terms of simply high or low charge (or, on vs off, or 1 vs 0) and if we know what the distribution of high and low charges are in the computer's semiconductor devices we will be able to predict the distribution of high and low charges in the near future (computers are designed and built to work in this way). Further abstraction casts the state of the computer in terms of numbers, memory addresses, instructions, etc. Thus, if one knows that the computer is in the physical state that corresponds to the description "the value in memory address 2351 is 362 and the next instruction is 'increment the value in memory address 2351 by 1'," then one can predict that the next physical state the computer will be in will correspond (according to the bridge rules) to "the value in memory address 2351 is 363".

A glaring difference between the relationship between linguistic theories and the brain and the levels of description of a computer is that in the latter case we have bridge laws whereas in the former we do not (it is, perhaps, seen as one of the jobs of neuroscience to find such rules). The absence of conventional bridge rules is crippling for both the meaningfulness of claims that a linguistic theory is an abstract description of how the brain works, and for the utility of such theories to mechanistically deduce what human behaviour would have been like at earlier stages in the evolution of language.

Imagine two people arguing about what a particular computer does in a certain instance. One claims that the computer started with the number six and the instruction to "add one" and produced the number seven. The other claims that it produced the number eight. Standard bridge rules can be called on to resolve this dispute: the physical state of the machine can be compared with the physical descriptions corresponding to the number seven or the number eight being produced. In this sense, the two people manage to say something about the physical state of the computer (e.g. one is right, the other is wrong). In contrast, a dispute between two people cast in terms of abstract descriptions of the brain cannot be so resolved. As there are no conventional bridge laws, both parties to the dispute could claim that the physical state of the brain corresponds to their linguistic theory description (via different bridge rules in each case). Thus, for example, if a linguist claims that "the fish" and "the river the fish swam in" are both represented as the same kind of entity (both noun phrases), and so representations are "recursive", no investigation into the physical structures in the brain will reveal that this is true (as whatever physical structures are found could just as well be translated via *some* bridge rules into a linguistic theory description in which the two strings are not the same kind of entity). In this sense, such claims (e.g. that language contains recursion, Hauser, Chomsky, and Fitch 2002) are not *about* any structure or process in the brain (they are neither statements of fact, nor hypotheses). Here, we might say that it is meaningless to state that a linguistic theory describes processes in the brain (though this is not to say that linguistic theories are meaningless, but that they are a different kind of game).

It is not entirely true to say that we have *no* bridge rules to relate descriptions of linguistic theories to descriptions of the human body. For example, if a linguistic theory makes a statement about the order in which a person will say certain words, we are (on the whole) able to tell cases in which those words are said in that order from cases in which those words are said in a different order (we do not need any special equipment for this, all we have to do is listen). If we have a physical description of such a case, we will be able to say something about the bridge rules (but only that this case corresponds to this particular linguistic theory description). A linguistic theory may be correct in that when this method of linking linguistic to physical descriptions is employed, predictions made by the linguistic theory hold.[22] However, even in such a case, the linguistic theory will be unable to make significant claims about the neurological processes taking place as bridge rules linking the aspects of the linguistic theory which we can recognise in actual situations to physical descriptions of those situations will not determine bridge rules for the aspects of the linguistic theory with which we are not familiar. Suppose two linguistic theories make the same behavioural predictions, but do so using different internal representations and procedures. Just as above, these different descriptions could, in principle, be related to whatever physically happens via different bridge rules.

It might here be objected that the possibility of relating a linguistic theory to physical processes relies on what might be highly gerrymandered, unsystematic, arbitrarily selected bridge rules. This is a valid objection, and were a linguistic theory's proponents to seriously propose such bridge laws, their endeavour may be dismissed by others as a somewhat pointless exercise. However, the objection cannot *make* the relationship between a particular linguistic theory and the physical processes in the brain any more or less gerrymandered. That a linguistic theorist would *like* his theory to relate in a simple and illuminating way to physical processes in the brain has no bearing on whether such a relationship actually holds.

If a linguistic theory can be related to brain processes only via highly gerrymandered bridge rules, this will impact on that theory's utility for predicting human behaviour in unusual (and untested) conditions. Suppose we have a linguistic theory that successfully describes linguistic behaviour, though the physical processes in the brain are unknown. If we want to know what human behaviour would be like in unusual circumstances (such as an earlier stage in the evolution of language in which an individuals' experiences do not contain language-using conspecifics) the linguistic theory would be a poor guide, as we would have no grounds for saying that a particular manipulation to the linguistic theory corresponded to the unusual situation. This is because we would have no bridge laws to link linguistic and physical descriptions of that situation. (In contrast, if a linguistic theory were constructed from an understanding of physical processes of and impingements on the brain, it may be possible to make statements about the physical differences between contemporary humans and those in the unusual situation, and from these compute likely behavioural consequences; in this endeavour, the linguistic theory might offer useful calculation shortcuts.)

[22] I do not claim that such any extant linguistic theory has achieved this.

The number of components in the human brain is “monumental”:

> Within the liter and a half of human brain, stereologic studies estimate that there are approximately 20 billion neocortical neurons, with an average of 7,000 synaptic connections each. The cerebral cortex has about 0.15 quadrillion synapses — or about a trillion synapses per cubic centimeter of cortex. The white matter of the brain contains approximately 150,000 to 180,000 km of myelinated nerve fibers at age 20, connecting all these neuronal elements. (Drachman 2005, p. 2004)

While it is conceivable that a currently available linguistic theory may relate via simple, systematic bridge rules to the physical processes in the brain, this possibility seems remote. Given the failure in chapter 3 to find neural structures reflecting the vary basic assumption (common to most cognitively cast linguistic theories) that to words there correspond internal representations of meaning, we must remain sceptical that linguistic theories can inform us about the evolution of language in virtue of being abstract descriptions of physical processes.

4.5 Summary

The Augustinian picture has a detrimental effect on approaches to language evolution for two reasons. (1) The picture is misleading as an account of what a language evolution theory has to describe. The various ways in which we operate with language are far more diverse than the idea that every word stands in for something else can allow for. (2) Augustinian interpretations of the language used in linguistics generate illusions such as that to a “phoneme” there must correspond some*thing* in the way that to a “table” there corresponds an object, or that someone’s following a rule for the use of a word means that they consult a rule formulation in the mysterious medium of the mind.

Our concepts are embedded in our lives, the activities we engage in. For concepts such as “meaning” and “understanding”, this embedding is far more complex than an Augustinian picture would suggest, the criteria for someone’s “understanding” on an occasion spreading over various aspects of their lives (such as that they *generally* understand similar things). While these words are governed by grammatical rules (in Wittgenstein’s sense), the rules do not cover every possibility, and beyond the context of their familiar application the rules may have no clear analogue. This is important to bear in mind when thinking about animals' communicative behaviours and earlier stages in the evolution of language. For example, we may be misled by a question such as "does the Vervet eagle alarm mean 'There's an eagle!' or 'Climb down from the trees!'?" into thinking that there must be an answer if only we knew what was going on inside the Vervet. The problem in such a question is not a lack of information, but a lack of rules for the expression "meaning" in this context (it is, of course, possible to introduce new rules to cover such cases as, for example, Harms (2004) does, but we do not *learn* anything about Vervets by an arbitrary stipulation). Earlier stages in the evolution of language may well appear to us (if we could observe them) peculiar, resisting clear descriptions in our familiar metalinguistic expressions, only gradually over time becoming more amenable to, e.g., analysis into meanings.

The account of linguistic theories in section 4.4.3.1 was somewhat idealised in order to press the point that constructing a symbolism that can be used to account for linguistic behaviour does not put one in a position to be able to predict linguistic behaviour in different situations, at least not on the basis of the interpretation of that symbolism as an abstract description of the mechanisms which produce said behaviour. The construction of a linguistic theory could in the future involve the construction of bridge rules, and this could be done on the basis of investigation into the physical brain processes involved in language. Such a theory would be explicit about how its parts related to the brain. This kind of theory is not an impossibility, but, given the conclusions of chapter 3, neither does it appear to be a reality.

The idea that the term "language" relates to some monolithic structure (perhaps a computational module in the brain) from which the variety of human activity which depends on language springs forth is not the result of a discovery but an assumption fostered by the Augustinian view. Rather than adopt this assumption, the next chapter considers the evolution of language as the development of a patchwork of variously interrelated language games.

CHAPTER 5

Describing language evolution without internal mechanisms

> If we look at the actual use of a word, what we see is something constantly fluctuating.
>
> In our investigations we set over against this fluctuation something more fixed, just as one paints a stationary picture of the constantly altering face of the landscape.
>
> When we study language we *envisage* it as a game with fixed rules. We compare it with, and measure it against, a game of that kind. (*PG* p. 77)

5.1 Introduction

Previous chapters of this thesis have identified as a common feature of accounts of the evolution of language a reliance on the externalisation of internal representations. The warrantedness of this "externalisation window" has been questioned both on empirical and conceptual grounds and found doubtful. If we reject the theoretical apparatus of abstract descriptions of internal mechanisms in theorising about the evolution of language, can we still develop an account of the evolution of language which is able to shed light on how the expressions of our languages came to have the grammars that they do? This chapter attempts to develop such an account. Specific and general attested semantic changes are used to illustrate the approach taken here and to identify processes which may be generalised to language evolution.

5.2 Communicational affordances

"Human beings are able to use language." In discussing the evolution of language, we may be wondering, "how come this is true?" However, the statement is amenable to two readings: "humans are biologically equipped with a capacity to use language," and "humans are in a position to use language (just as I am able to email people but Shakespeare was not)." This section argues that the evolution of language may be fruitfully approached by considering how the communicative possibilities available to humans might have developed without relying on a mechanistic model of human cognition/behaviour. To refer to the communicative possibilities made available to an individual by, *inter alia*, other individuals, I will adopt J. J. Gibson's term, *affordances*. Gibson's term was a conscious attempt to describe the environment of an organism in objective terms (as opposed to the organism's subjective attitude toward the environment, or what the environment seems like to an organism) while relativising the description to what the organism *can do*. Thus the surface of a pond affords walking on for a pond skater but not for humans.

An aspect of Gibson's work that motivates my adoption of his term here is his emphasis on what structure there is in the environment for an organism (and for the evolution of organisms) to exploit. In Mace's (1977) terms, "ask not what's inside your head, but what your head's inside of." However, another aspect of Gibson's work is not relevant to the present discussion, namely his intervention in the debate concerning the relationship between physical impingements on an organism's sensory surfaces and what the organism perceived. Gibson (1977) claimed that affordances correspond directly to unique configurations of sensory stimulation and so are perceived "directly". When using the term "affordances" I do not share this (problematic, see Sharrock and Coulter 1998) claim.

The communicational affordances offered to an individual include possibilities available to him for interaction with other individuals. What is of interest is how the possibilities for interaction change as other individuals learn about the possibilities for (and hence develop expectations about) such interactions.

5.2.1 An example of affordance dynamics

Tomasello (2000, 2003, in press) often cites the human ability to adopt the role of one's partner as a uniquely human capacity. In his view, this "role reversal imitation" is an important component of the capacity for groups of humans to (continually) develop cultural artifacts (including language). The following is an abstract example of how the landscape of communicational (or interactional) affordances may change, leading to a situation in which role reversal imitation is a viable strategy for an individual. In what follows, the two acts or roles in a dyadic interaction are referred to as X and Y, and individuals are referred to by capital letters beginning at A. The example is presented as a series of discrete stages for ease of exposition.

Stage 1 X and Y do not form part of any individual's interactional repertoire. Were an individual to happen to perform either X or Y, no other individual would respond in any way characteristic of the interactional activity that will eventually develop. Individuals do interact with each other, but not by using X or Y.

Stage 2 An interactional routine becomes ritualised. Suppose A and B regularly interact in a certain way. Perhaps B sometimes carries A on her back, and this is achieved by A climbing up onto B. Because A regularly begins climbing by grabbing B's back and pulling, it is possible for B to tell that A is initiating a carrying episode by the fact that B's back has been grabbed and pulled. Thus B may come to lower her back to help A when feeling her back pulled. This in turn changes the affordances for getting onto B's back available to A who may now find that B's back is lowered when A simply touches it. This is how Tomasello (in press) describes the development (in his terms, ontogenetic ritualisation) of wild chimpanzee's *touch-back* gesture (used by infants with their mothers).

For the purposes of this abstract example, we need not be restricted to the *touch-back* gesture. The structure of extant interactional behaviours may allow B to pre-empt A's behaviour, and so modify B's own behaviour accordingly. This change to the interactional dynamic may also change what A can learn about B's actions which may then lead to changes in A's behaviour. In this way, by learning from each other's responses, roles X and Y may become part of A's and B's repertoire. Let us suppose A plays role X and B plays Y.

At this stage, it is conceivable that A and B could switch roles. However, it is also conceivable that they simply do not. If we think if A and B as machines tuned to respond to events in certain ways, then the development of the X/Y ritualised behaviour has tuned them differently. Even if A attempts to play B's role, there is no need to assume B will respond by playing role X.

Stage 3 That A and B play roles X and Y together changes the affordances for another individual C. Because B has learned to play role Y with A, it becomes possible for C to learn to play role X through interaction with B: if C does something like X with B then B will respond differently to would have been the case had the interaction with A not been ritualised.[1] Similarly, it has become possible for D to learn to play role Y through interaction with A.

This process of learning to exploit the behaviours of others may not only spread the practice through a community, but can also bring A and B to learn each other's roles (A learning Y through interaction with C and B learning X through interaction with D). Thus a situation may emerge in which all members of a group are able to play both roles X and Y.

Stage 4 At the beginning of stage 3, the number of individuals who would participate in the X/Y interaction was limited. When C learned that he could perform X with B, he did not thereby learn that he could perform X with E, F or G. The scope for C engaging with other individuals by performing X may be limited. C may restrict his X performance to B, or even find that engaging with others does not produce the same results leading him to avoid performing X with anyone but B. These effects may retard the rate at which new group members can pick up X or Y, but they need not make learning by exploiting the affordances of A's, B's and C's behaviour impossible. If C is reluctant to perform X with E, for example, he may nonetheless do so. He might, for example, mistake E for B, especially if what E does resembles B's performance of Y.

However, as more members of the group learn the routine, this dynamic alters: by the time G comes to learn to play role X, his environment contains a number of individuals who can play Y to his X. Whereas C couldn't (initially) learn that X can be played with most people in the group, G can. As individuals learn what opportunities they have to interact with each other, they thereby change those possibilities.

If a number of different interactional games develop along the lines laid out above, another change in what can be learned may happen. Suppose a group has acts or roles X_1, X_2, ... X_n paired in interactional games with Y_1, Y_2, ... Y_n respectively. H is a new member of the group, the other members of which have already learned most of the other roles (including both roles

[1] Here we part company with the *touch-back* example as, according to Tomasello (in press) each chimpanzee pair that employs the signal does so on the basis of their own ritualisation of it, not through contact with others.

> for a number of interactions). In this situation, if H can play role X_m with a particular individual, he can also play X_m with (almost) any other member of the group, and he can also play Y_m with (almost) any other group member. These facts are available for H to learn and exploit, and because there are numerous interaction games, *evidence* is available to H that this is the case (that is, H's experience with $X_1, X_2, \ldots X_{n-1}$ and the corresponding acts or roles will be evidence for it being possible to reverse roles for the next interactional game, X_n/Y_n, he learns). Whereas at stage 2 it might not have been true that A could reverse roles and interact successfully with B (and so not something A could learn) the dynamics described above could produce a situation in which it *is* true that roles can be reversed (and in which this is something that *can* be learned).

Tomasello's characterisation of role reversal imitation as a uniquely human capacity suggests that it is an insight brought to interaction learning by human infants. However, the above scenario suggests the possibility that a group of individuals who do not display role reversal imitation may nonetheless produce a situation in which such imitation may be available as something that can be learned. The degree to which genetic changes between humans and our ancestors make us more able to engage in this kind of imitation is difficult to assess (in part because the organisms from whom we learn are different now to those in the contexts of our genetically different ancestors). However, it is worth noting that early in development, contemporary children may not display the insight that roles in a communicative interaction may be reversed: a number of studies taken together show a double dissociation between infants' production and comprehension of pointing (some point for adults before understanding adults' pointing, while others comprehend adult pointing some time before pointing themselves, Tomasello 2003, p. 33).

Another respect in which young children do not seem to view their own communicative productions as a reversal of adult productions concerns differences between adult and child pronunciation. Some children respond to imitation of their own mis-pronunciations as mis-pronunciations, yet do not consider their own production mistaken. Berko and Brown (1960) dubbed this the *fis* phenomenon:

> In imitation of the child's pronunciation, the observer said: "This is your *fis*?" "No," said the child, "my *fis*." He continued to reject the adult's pronunciation until he was told, "This is your fish." "Yes," he

> said, “my *fis*.” (Berko and Brown 1960, p. 531, quotation taken from Clark 2003, p. 71)

The fact that this child does not accept adults mimicking the form he produces, and the fact that he is able to *distinguish* this form from the adult form, indicate that his own productions are not the result of attempts to imitate adults. If role reversal imitation develops gradually, it is possible that this general strategy is the product of experience in an environment in which roles can generally be reversed.

If we restrict our view of unique human behaviours to the capacities of individuals, we may fall into the trap of attributing social phenomena as a direct reflection of individual psychology. While the discussion of role reversal imitation here isn’t sufficient to demonstrate that the reversibility of roles in many interactions is not the result of an experience-independent human propensity to reverse roles, it does show that this unsupported assumption is not necessary. By broadening our perspective beyond hypotheses about individual psychology, we can approach the development of our interactive practices without the theoretical apparatus rejected in earlier chapters.

5.2.2 *Human nature*

The above scenarios focus on the interactional/communicational affordances available to individuals, and how these change as they are exploited. However, that the environment affords a certain interaction does not mean that individuals will necessarily learn to exploit those affordances. What individuals learn depends on their experience, but is not determined by it (even when their experience involves explicit teaching, c.f. *PI* §§143–145, 185). This is a grammatical note on the concepts of “experience” and “learning”. When explaining someone’s behaviour (i.e. when engaging in this language game) we appeal to their experience to a degree, but also leave room for the individual’s “nature” to account for their action. That human beings learn to speak when raised in a human environment but other animals do not is a ground for talking about differences in the nature of humans and other animals.

Human nature is an ineliminable aspect of a full account of the evolution of language. However, this grammatical note does not enable us to answer empirical questions about what human behaviour would be like given different experience. We may say that it is part of H’s nature to engage in role reversal imitation given his experience of interactions in which parties can adopt each others roles; but this does not answer the question whether H would act in this manner in a context in which roles were tied to individuals. This point may be obscured by abstract mechanistic descriptions of human cognition (c.f. section 4.4). Tomasello et al. (2005, abstract) cast the ability to reverse roles in terms of "unique forms of cognitive representation" in which internal representations incorporate representations of other individuals' internal representations. This way of describing role reversal imitation *looks* like the identification of an architectural feature of some physical component of the brain, a consequence of genetic differences between humans and animals. Such an interpretation may suggest that we are in a position to determine the cognitive architecture (and hence behaviour) of humans in unusual developmental circumstances. However, even if we make the *decision* to adopt this notation, we have no idea how it would relate to neural processes, nor how these neural processes develop and so no further clue as to what will happen in different circumstances (c.f. section 4.4.3).

When viewed from the perspective of communicational affordances, an important factor for the sustainability of an interactive practice or language game is that the judgements made by individuals agree (c.f. *PI* §242). If, for example, what A does doesn't conform to B's expectations, we may say the interaction has failed. How A proceeds, and what B expects, are both the result of their experiences and their natures[2] and so whether a particular interactive practice can be sustained by humans is in part a function of human nature. For example, if a system of colour terms is to be supported and used, the members of the community must on the whole agree in such judgements as "this is the same colour as that" (otherwise a command such as "bring me a red object" would produce a reaction other than that intended by the one giving the order, and the problem could not be solved by pointing to a red sample and saying "I meant an object *the same colour* as this"). In part, the possibility of such a language game depends on the nature of the human visual system (which determines, for example, whether two objects with different reflective properties *can* be perceptually differentiated).

We may say that because ordinary language use relies on individuals' agreement in judgements, the form language games can take is influenced by human nature. However, this is not to say that they are a direct product of the nature of individuals. Thinking about language evolution in terms of continually shifting communicative affordances emphasises the fact that an individual learns to

[2] Again, this is a *grammatical* note

engage in the language games that those in his environment have also learned to play. While the pre-linguistic response of the visual system to different combinations of light may be one factor in determining what colour–like language games are possible, this is not the sole determinant of the colour term system in a language (see chapter 6). Other factors may also influence why two objects are seen as "the same colour" by a community, and initiates into this practice may gradually learn these judgements (visual systems permitting). That a language game exists only shows that it is possible for humans to engage in this kind of interaction; it does not show that such interactions come "naturally" to individuals in the sense that they would behave in this way *whatever* their experience. In the above imagined scenario, a situation emerged in which roles were reversible *in spite of* the assumption that individuals would not spontaneously take on a reverse role. Therefore, it is a mistake to take the properties of language as revealing properties of individual psychology which explain the properties of language. However, this is not an uncommon manoeuvre in explanations of language change (section 5.3.4).

5.3 Language change

This section looks at attested semantic changes (that is, changes in the practices of using words), and attempts to view these changes as shifts in communicational affordances as individuals exploit the possibilities open to them and thereby open up new regularities for others to learn as rules. The accounts below do not rely on internal mechanisms, nor on *ad hoc* assumptions about human psychology. Instead, publicly surveyable reasons for new regularities in the use of particular words to emerge are sought. In this, numerous assumptions about the behaviour of human beings are made, and these can be accepted or rejected at face value as they are not spuriously justified by a model of internal processes.

5.3.1 Irregularities in semantic change

In comparison to sound change, semantic change has traditionally been relatively neglected in linguistics. Semantic changes have been felt to be "random, whimsical and irregular; general rules concerning them are nearly impossible to establish," (Sweetser 1990, p. 23). The kind of irregularity alluded to here may be differences across words: e.g. some word changes appear to relate to emotive value, but changes can either be positive (melioration) or negative (pejoration).

Another kind of irregularity is differences in semantic change across different languages. For example, the Dutch word *knap* meaning "able, fit, clever, good looking" melioratively derives from a Germanic word meaning "fitting close, tightly", while the modern German *knapp* ("narrow, hardly sufficient") is a pejoration of the same word (Bartsch 1984, p. 387). Similarly, English *deer* and German *Tier* ("animal, wild or tame") reflect narrowing and broadening respectively of their common Germanic ancestor meaning "wild animal" (Bartsch 1984, p. 385).

While semantic changes may be irregular in the sense that it is difficult to predict what changes will apply to which words and when, they are not incomprehensible. For example, the word *knave* ("unprincipled man") is a pejoration of the Old English word *cnafa* ("boy"). Bréal (1900) took pejoration to be the result of human dispositions either "to veil, to attenuate, to disguise ideas which are disagreeable, wounding or repulsive" (p. 100) or "to take pleasure in looking for a vice or a fault behind a quality" (p. 101). However, if we consider shifts in communicational affordances, we do not need to see the pejoration of *cnafa* as the result of individuals' tendencies to spontaneously use words in more pejorative ways (e.g. to spontaneously call an unprincipled man a *cnafa*). An account of this change in terms of shifting communicative affordances would note that if (as is not uncommon) the cultural conception of young males included their being particularly mischievous and unprincipled, then the term used to refer to young males can be used in certain ways: for example, the mischievous and unprincipled actions of a particular boy may be accounted for in conversation by appeal to this generalisation (e.g. "he lied to you? Well what did you expect, he is a *cnafa*, after all.") Within such a community, certain communicational moves with *cnafa* are made possible (e.g. excusing bad behaviour, warning about likely behaviour, expressing exasperation that a particular boy conforms to a stereotype, etc.). The appearance of new word in the language (*boy*) replaced many of the old uses of *cnafa*, but left those associated with mischief. How might this selective replacement have come about? In some individuals' experiences of these two words *cnafa* may have been used on the whole in relation to boys' mischievousness and *boy* more generally. For example, a particular boy may generally find he is referred to as a *boy*, but that certain individuals with whom he only interacts when he has behaved mischievously refer to him as *cnafa* (this pattern may come about not because these individuals only use *cnafa* in this way, but because they only ever interact with this boy when he has been mischievous). Having learned an association between *cnafa* and mischief-related uses, this boy then in turn presents communicational affordances to others: e.g. he may take being referred to as a *cnafa* but not as a *boy* as a reprimand. It is unlikely that a sole individual's experience of the uses of *cnafa* and *boy* described above would lead to the observed semantic change for *cnafa*, but it is reasonable to suggest that more than one boy would have had such structured experiences (thereby increasing the scope for individuals to learn and use only mischief related uses of *cnafa*).

This interpretation in terms of replacement by *boy* is supported by chronological evidence: the OED has examples of *cnafa*/*knave* dating back to 1000, but the earliest attested use in relation to a lack of principle dates to 1205; examples of *boy* begin in 1300, and the last attested use of *knave* in a simple 'male child' sense is found in 1460. Given that changes in usage may be assumed to spread gradually across the geographically dispersed users of the English language, and given the noisy relationship between speech and written attestation, these dates support the view above.

The replacement by *boy* of some uses of *cnafa* would account for a restriction to notions of mischief, but why did the word's applicability generalise to a greater age range? One consequence of the above replacement would have been a greater indeterminacy in the age at which a male is no longer considered to be a *cnafa*. A word which is generally used in ways similar to contemporary *boy* may resist such a change as its grammar will in part concern various social and biological properties of male youth (e.g. the grammar of the word *boy* allows pre-pubescence as a justification for designating an individual as a *boy*, and on those grounds denying him certain rights). However, if various uses of *cnafa* are lost (by being replaced by another word) such anchoring factors may also be lost (especially if unprincipled behaviour is not restricted to a particular age range). A border-line case may, if accepted, shift the vague boundary of what is acceptable as a *cnafa*.

This account, rather than relying on the tendencies of individuals to change the meanings of words, locates the possibilities for using language in certain ways in structural aspects of human life (e.g. that boys are often perceived as being generally mischievous; that an individual's mischievousness may influence the conditions under which he interacts with certain individuals). The account given is somewhat simplistic and neglects many factors (e.g. the use of *cnafa*/*knave* to refer to male servants) but is intended to illustrate how the dynamics and structuring of language use can affect the affordances associated with a word. It is also worth noting that the suggested pathway of change was not inevitable (e.g. insufficient numbers of individuals may have had their experience of *cnafa* restricted to its mischief-related uses), and this reflects the seemingly irregular nature of semantic change.

5.3.2 Development of English must

An interesting and well studied change in the English language is the development of the contemporary modal verbs (see e.g. Warner 1993; Traugott and Dasher 2002; Goossens 2000). Here I discuss some of the developments leading to modern uses of *must* from the perspective of communicative affordances. The outline of these developments is taken from Traugott and Dasher (2002), though the theoretical frame through which these developments are viewed differs from theirs (see section 5.3.4).

The uses of *must* can be roughly divided in two categories: "deontic" (or "root") uses which express obligation, as in (1) and "epistemic" uses which are used to express certainty, as in (2).

(1) He must do his homework.
(2) He must be there by now.

Traugott and Dasher (2002) identify three stages which are convenient for describing the development of the use of *must*:

Stage 1: ability, permission The preterite-present verb from which *must* derives (Old English *mot-*) can be traced to the use of a past form as a present. This earlier form itself derives from Indo-European **med-*, whose use related to appropriate measures and fittingness (c.f. *medical*, *modal*, *modify*, *commodity*, etc., Traugott and Dasher 2002). The development of the use of a past form of such a verb in a present sense is intelligible if we consider some of the uses to which past forms can be put. Something that happened in the past relating to appropriateness may have relevance to the current situation (e.g. the fact that an individual "learned a foreign language" is a valid reason for sending him to talk to the foreigners). That is, there are some communicative scenarios in which a word which we might gloss as relating to a past process of "fitting" performs the same function as a word whose gloss would relate to current ability, the difference between these two glosses residing in the *other* uses of these words. However, the development of ability uses of *mot-* had already occurred by the Old English period, and little evidence has been collected concerning the development of these uses (Traugott and Dasher 2002).

The use of *mot-* in relation to ability enabled the development of uses of *mot-* in relation to permission. As an example showing how this change may have come about, Traugott and Dasher (2002) cite the following from Beowulf:

(3)

Ic	*hit*	*þe*	*þonne*	*gehate*	*þæt*	*þu*	*on*	*Heorote*	*most*
I	it	you	then	promise	that	you	in	Heorot	will:be:able

sorhleas	*swefan.*
anxiety-free	sleep

"I promise you that you will be able to sleep free from anxiety in Heorot"

(8th century[3])

Traugott and Dasher interpret this example thus:

> In (3) Beowulf promises to be the enabler of sleep. One could think of him as promising to be the external remover of barriers to sleep. In this sense he is also "permitting" sleep. (Traugott and Dasher 2002, pp. 122–123)

As an explanation for how a word relating to ability could come to be used to talk about permission, this explanation seems somewhat strained. Beowulf's promise is that Hrothgar will be able to sleep soundly in Heorot because Beowulf has slain the terrorising monsters, not because Beowulf gives permission. However, if we consider the variety of ways a word relating to ability is used, a link to permission can be discerned. For example, if A needs someone to perform a certain act, then the fact that B has the ability (*mot-*) is a reason for A to approach B. That B is a suitable candidate for performing the act may relate to his physical or mental state, but equally in a society in which some individuals have the power to regulate the behaviour of others, B might be suitable because of the social permission granted to him. In such a society, a number of language games, developed on the basis of inherent abilities (e.g. recommending B to A), map on to instances where permission rather than personal quality is the basis for thinking a person is in a position to do something.

[3] Beowulf (Traugott and Dasher 2002, p. 122)

This overlap in ability and permission uses may have facilitated the development of permission-related grammar for *mot-*, as learning how to engage in ability-related *mot-* language games may ready an individual for aspects of permission-related grammar. However, we need not assume that the development of permission uses was spontaneous, as transitional cases can easily be imagined. Consider the use of *mot-* by an individual who has power to grant permission to another. The powerful individual (A) may deny the other (B) the chance to attempt some task on the grounds that he doubts B's capability to perform the task. In such instances, A's use of *mot-* would be an expression of his denying permission to B. If A regularly gives and denies permission on the grounds of his perception of abilities, then A's use of *mot-* would in many cases play the role of an expression of permission. Other individuals may then learn to exploit the affordances thus offered by A (e.g. by asking whether in A's opinion they have the capability / whether A grants permission to do something). Another aspect of the development of more clearly permission uses may be A's misjudgement of individual's abilities and the granting or denying of permission on that basis (thus, others may learn that when A says that B lacks the ability it may be the case that, if allowed, B would be able to perform).

Stage 2: obligation Once a permission sense of *mot-* was established, it began to take on uses connected with obligation. For example,

(4)	*swa*	*þa*	*lærendum*	*þam*	*preostum*	*se*	*papa*	*geþafode*
	so	then	advising-DAT	those-DAT	priests-DAT	the	pope	granted
	þæt	*Equitus*	*moste*	*beon*	*gelæded*	*to*	*Romebyrig.*	
	that	Equitus	should	be	bought	to	Rome	

"so then the pope granted to those priestly advisors that Equitus should be brought to Rome."

(*c.* 1000[4])

Example (4) exemplifies the fact that the granting of permission can impose an obligation. Had the advisors not tried to bring Equitus to Rome and instead simply remained satisfied that they had permission to do so, the pope would have become understandably annoyed (why had they wasted his time asking for permission if they didn't then want to bring Equitus?). The consequences of not doing what an authority has granted one permission to do (especially if one has

[4] *Bischof Wærferths von Worcester Übersetzung der Dialoge Gregors des Grossen*, (Traugott and Dasher 2002, p. 125)

petitioned for the permission) mean that an expression relating to permission, as *mot-* did, can, in some cases, function as an expression of obligation.

Another related congruence between permission and obligation concerns the citation of one's permissions as reasons for one's actions. For example,

(5)

we	*moton*	*eow*	*secgan*	*eowre*	*sawle*	*þearfe,*	*licige*	*eow*	*ne*
we	must	you-DAT	tell	your	soul-GEN	need	please	you	not

licige	*eow.*
please	you

"we must tell you about your soul's need, whether it please you or not"

(*c.* 1000[5])

Here, that the speaker has been granted permission is offered as a justification for his action in spite of his awareness of reasons not to carry his action through. In this interaction, *mot-* performs the justificatory role of a word expressing obligation. The development of obligation uses of *mot-* from permission senses may be seen as deriving from the relationship between permission and obligation resulting from the social structure of permission granting, a relationship individuals may learn.

A community which has learned to use *mot-* in relation to obligations deriving from permission will afford some uses of *mot-* for similar deontic purposes in cases where permission does not feature. If A uses *mot-* in an obligation sense only in cases where he has been granted permission, B may be ignorant of the role of permission yet still engage appropriately with A in these language games: if A addressed B with example (5) an appropriate response would be for B to refrain from expressing his displeasure since A's use of *mot-* indicates that such expression is superfluous. Thus B may learn to exploit the communicative affordances around him *well enough* without knowing a rule relating *mot-* to permission. By using *mot-* this way, B of course affords this use to others (that is, C can learn that *mot-* indicates to B that an expression of displeasure will be ineffective).

Stage 3: epistemic An obligation makes the obligee do something. An individual can cite his own obligations as justification for his disregard of other reasons for a different course of action. An individual can cite his interlocutors obligations as a way of telling him what to do. The fact that obligations are reasons

[5] *Ælfric's "Catholic Homilies" 1^{st} Series* (Traugott and Dasher 2002, p. 124)

for action means that a non-present third party's obligations can be cited as reasons for thinking that he will do whatever it is that the obligation tells him to (e.g. A: "will he pay his bill?" B: "yes, he must pay his bill"). Thus certain uses of *mot-* in an obligation sense are connected to individuals' beliefs that what is obliged will happen (whether or not individuals really do, on the whole, fulfil their obligations). This regular connection between citing obligations and belief that something will happen is something that individuals could learn. Having learned it they would afford the use of *mot-* to indicate one's belief that something will happen. As above, this could lead to the development of uses in which *mot-* was used to express one's belief that an event will take place without any obligation imposed on the actors.

While this process could produce epistemic–future uses, it may not by itself account for more general epistemic uses (e.g. "the tallest man in Britain must be taller than 4 foot"). The roots of these uses may lie in societal structures of general obligations, imposed on all people and at all times, deriving from the authority of the law and/or God. For example,

(6)	*ho-so*	*hath*	*with*	*him*	*godes*	*grace:*	*is*	*dede*	*mot*	*nede*
	who-so-ever	has	with	him	God's	grace	his	deed	must	necessarily
	beo	*guod.*								
	be	good								

"he who has God's grace necessarily is required to be/we can conclude is good"

(*c.* 1450[6])

In (6), God's timeless imposition on a generalised subject (a general feature of Christian theology) means that in this context *mot-* functions both as an expression of obligation and as a general epistemic expression. Again, individuals can learn about others' use of *mot-*, learning that it can function as an indication that something can be concluded to be the case. In Traugott and Dasher's survey of textual evidence for the development of *must*, wide scope deontic necessity (commonly imposed by God or the law) was the chief source of Old and Middle

[6] *Life of St Edmund* (Traugott and Dasher 2002, p. 128)

English examples which could be read as expressing epistemic necessity (2002, p. 130).[7]

Development of English must: summary The factors involved in the shift of the communicational affordances of *mot-*/*must* have here been addressed in broad outline. More detailed discussion of the modern grammar of *must* and the possibilities for its development are beyond the current scope. The picture painted here with broad strokes is one of the exploitation of the communicative possibilities afforded by the environment, and the shifts in those affordances brought about as affordances are exploited. Structure in patterns of usage, which may be imposed by factors other than the grammar of *mot-* (e.g. valid conversational moves in relation to obligations imposed by God) may become part of the grammar of a word. The development of *mot-* would have been different had, for example, the social structures in Britain been different: if individuals did not regulate each other's behaviour, *mot-* would not have developed a permission use; if the granting of permission did not impose an obligation, the obligation uses of *mot-* may not have developed, etc.

The chronology of the development of *mot-* presented here is taken directly from Traugott and Dasher (2002). However, while Traugott and Dasher appeal to models of cognition to constrain or direct change (c.f. section 5.3.4 below), the account presented here sees usage as structured by various aspects of human affairs, and change the result of individuals learning and exploiting those structures, thereby creating new regularities.

5.3.3 Generalisations over changes

The development of English *must* fits into patterns of development displayed by other English modals. Like *must* both *may* and *can* developed from words relating to specific kinds of abilities inherent in an agent (physical and mental respectively, Bybee et al. 1994). These developed more general kinds of agent–ability uses and subsequently were used to express a more general notion of

[7] Another possibility is the use of *mot-* when teaching a technique (e.g. some aspect of geometry). Here an expression of what the pupil's is obliged to do in order to correctly perform the technique may also function as an expression of epistemic necessity (as in "the internal angles of a triangle *must* add up to 180°"). This congruence relies on the variety of possibilities for who is obliged when *must* is used in a sentence: "The witch must be kissed by every man in the room … or the leader of the coven will demote her to leprechaun / … or they'll all be turned into star-nosed moles," (Sweetser 1990, p. 67). However, the variety in who is obliged by a deontic use of *must* is not an issue dealt with here.

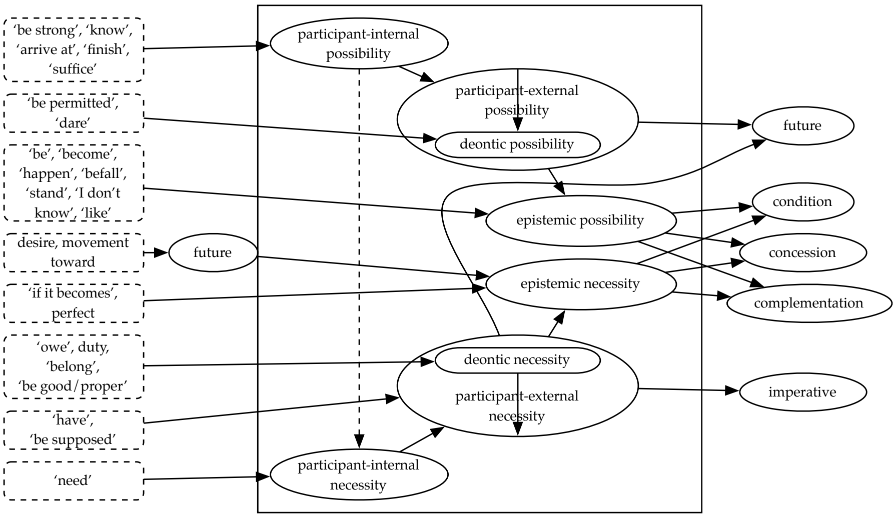

'be strong', 'know', 'arrive at', 'finish', 'suffice'
'be permitted', 'dare'
'be', 'become', 'happen', 'befall', 'stand', 'I don't know', 'like'
desire, movement toward
future
'if it becomes', perfect
'owe', duty, 'belong', 'be good/proper'
'have', 'be supposed'
'need'
participant-internal possibility
participant-external possibility
deontic possibility
epistemic possibility
epistemic necessity
deontic necessity
participant-external necessity
participant-internal necessity
future
condition
concession
complementation
imperative

Figure 5.1: Possibility and necessity paths, from van der Auwera and Plungian (1998, Fig. 14). (Change from participant-internal possibility to participant-internal necessity (dashed arrow) is additional to the original diagram, and taken from Traugott and Dasher 2002, Fig. 3.4.)

ability (i.e. one that need not be thought to inhere in an agent, Bybee et al. 1994). Like *must*, both *may* and *can* developed from ability related uses to permission and eventually epistemic uses. These patterns of development are not restricted to English, but have been detected in a large number of genetically and areally unrelated languages (Bybee et al. 1994).

Cross-linguistic parallels in diachronic developments of usage patterns are not restricted to mood and modality, but have been identified for various domains including tense and aspect (Bybee et al. 1994), discourse markers, and performative verbs (Traugott and Dasher 2002). This regularity has led to the statement of a number of generalisations which apply more or less broadly to diachronic changes. Before discussing these generalisations, it is worth noting that these generalisations represent tendencies or likely changes rather than hard rules to which all changes adhere. Figure 5.1 condenses various attested changes concerning expressions of possibility and necessity. Within a given language, only some of the paths of change will be traversed. The great variety of possibilities is an indication that hard rules do not govern semantic changes.

5.3.3.1 Unidirectionality

An interesting aspect of regular semantic changes is that they seem to be largely unidirectional. Changes represented in Fig. 5.1 follow the direction indicated by the arrow heads, but changes in the opposite direction are less commonly attested or not attested at all. Unidirectionality prompts two questions about change: (1) Why does change happen in this direction? and (2) Why doesn't change happen in the opposite direction?

The first of these questions will be approached from a shifting-affordances perspective below. The second question may be approached (at least for some changes) by considering statistical explanations of the second law of thermodynamics (which relates to irreversibility in physical systems). Informally, the second law states that, over time, the disorder of a system tends to increase. Thus, for example, if a completed jigsaw (i.e. a highly ordered configuration) is put in a box then over time either nothing will change or the system will become more disordered (e.g. shaking the box will move the pieces of the jigsaw out of their position and into a jumbled configuration). The time-directedness of changes summarised by the second law can be accounted for by abstracting away from the forces acting within the system and considering the problem statistically instead. There are many configurations in which the pieces of a jigsaw may be in a box, very few of which are described as the highly ordered state of "being a completed jigsaw". Thus, if we think about the effect of shaking the box as moving the system from one configuration to some other (randomly selected) configuration, ending up in a state characterised as disordered is far more likely than ending up in a state characterised as ordered (even if we assume that any *particular* disordered state is just as likely as any *particular* ordered state, since the argument rests on the far greater number of disordered states than ordered ones). The mechanics governing the motion of jigsaw pieces in a box do not preclude the possibility that when a box containing a jumbled jigsaw is shaken the pieces all fall into place, but this possibility is vanishingly unlikely. Similarly, if a general trend in semantic change goes from type-A states to type-B states but not vice versa, this may simply be a consequence of the fact that there are more patterns of usage which would be characterised as type-B than type-A (such that changes in usage, from either a type-A or a type-B state will probably result in a type-B state).

Consider the generalisation that expressions tend to change from 'concrete' to 'abstract' meanings (Traugott and Dasher 2002). The grammar of a concrete expression is tied to particular physical objects, events or situations. In contrast, the rules for using an abstract term may relate to a greater variety of objects, events or situations, or the rules may even be so general that we should say the word's grammar doesn't relate to anything physical. That abstract terms tend not to give rise to concrete ones may be thought of as reflecting this asymmetry: if a change in the use of a term affects the range of physical objects with which the use of the term is associated, the greater number of possibilities that would be described as "abstract" than "concrete" makes increasing concreteness unlikely. One way in which the use of an abstract expression may change is by another (concrete) expression replacing a limited range of the use of the abstract expression (as a hypothetical example, *cash* could replace *money* during transactions involving notes and coins, rendering the sentence "you have to give the cashier your money" incorrect); in such a case, the concrete expression may replace the abstract in only a limited range of physically specifiable contexts. While this could, in principle, lead to an association between the (previously) abstract term and specific physical circumstances, the coincidence required (i.e. that the difference in the range of applicability for the abstract and concrete expressions correspond to some recognisable physical circumstance) may be so unlikely that this kind of change be effectively unidirectional.

This argument is accentuated when tied to the communicative affordances offered and exploited by a group. While it is conceivable that in one individual's experience, an abstract expression is used only in connection with an identifiable physical circumstance, the chances that this coincidence would be regularly repeated over individuals such that eventually the experience of every member of the group was commensurate with that expression having concrete meaning may be relatively small.

Similarly, the general trend for deontic uses to give rise to epistemic uses but not vice versa can be explained by noting that frequently a speaker's use of a deontic expression affords conclusions about their epistemic attitude whereas the use of an epistemic expression does not generally indicate whether the speaker believes some obligation obtains. In this sense, a large proportion of deontic uses may form a subset of epistemic uses, but not vice versa. The development of a deontic from an epistemic need not be thought impossible (if in an individual's experience an epistemic has only been used in an apparent connection with the speaker's belief about obligations, that individual may learn to use the expression deontically). But the coincidence required for this change to occur makes it statistically a virtual impossibility.

5.3.3.2 General tendencies in semantic change

The broad statistical perspective addresses the question of why certain changes are unlikely to happen, but cannot deal with the first question identified above, why certain kinds of semantic change happen in parallel in unrelated languages. Traugott (1989) condensed these regularities into three broad tendencies:

Tendency 1 "Meanings based in the external described situation > meanings based in the internal (evaluative/perceptual/cognitive) described situation," (Traugott 1989, p. 34). This tendency subsumes pejoration and melioration, the extension of words from the "sociophysical" domain to the emotional and psychological (for example, modern English *feel* derives from Old English *felan* which initially meant only 'touch'), and the tendency for expressions of perception to be used to describe so-called cognitive processes (e.g. the extension of expressions relating to vision to uses relating to understanding as in "I see what you mean" Ibarretxe-Antuñano 1999). The development of *mot-* from relating to an imposed obligation to a more general deontic sense ("They look delicious, I must try one") which may be interpreted as a "participant internal obligation" also exemplifies this tendency.

Tendency 2 "Meanings based in the external or internal described situation > meanings based in the textual and metalinguistic situation," (Traugott 1989, p. 35). That is, the degree to which expressions are used to manage or refer to the flow of discourse (their "metatextuality") increases. Traugott's paradigm example is the development from Old English *þa hwile þe* (in which *hwile* was a noun meaning 'time') to a temporal connective and later a concessive (*while*). Other examples offered by Traugott and Dasher (2002) are the development of performative function of expressions which previously only had descriptive functions (e.g. *I recognise I was wrong to do that*) and of adverbials used as discourse markers (e.g. *anyway* to mark the speaker's return to a topic).

Tendency 3 "Meanings tend to become increasingly based in the speaker's subjective belief state/attitude toward the proposition," (Traugott 1989, p. 35). Pejoration and melioration exemplify this tendency (in addition

> to tendency 1). Other examples include the development of epistemic uses of *mot-* (and other modals), concessive meaning developing from temporal (e.g. *while*), and the development of honorifics in Japanese from non-honorifics (i.e. the meaning becomes a reflection of the speaker's attitude to their interlocutor). Traugott and Dasher (2002) see this tendency as dominant in semantic change.

A generalisation over these three tendencies can be made: in each of them, the grammar of an expression becomes less related to the non-human domain and more related to increasingly sophisticated patterns of human behaviour.[8] Why do the communicative affordances offered and exploited by groups change in these ways? In several of the examples above, changes in communicative practices have often begun as regular associations (e.g. the association between boys and unprincipled behaviour was hypothesised to be the ground for the development of English *knave*; the associations between permission and obligation, and between acknowledging obligation and believing something would happen were factors in the development of *mot-* outlined above). How could associations like this come to be part of the grammar of a word?

Consider a simple language game in which a word is used in a manner analogous to "slab", "pillar" and "beam" in Wittgenstein's builders' activities (*PI* §2, section 4.2.3 above). However, in this language game A is a pie maker and orders B to fetch him certain fruits. Suppose that some of the fruits contain seeds which C discovers make pleasing decorations. In this situation, the communicative affordances available to C allow him to use A's expressions to get B to bring him seeds: the fact that there is a way of asking B to fetch a fruit and the fact that fruits contain seeds mean that C has a way of getting B to bring some seeds (and this is something he can learn about his environment). That A and C intend to do different things with what is brought on their command means that they present different affordances to B. For example, B may learn that it doesn't matter if he eats the outside of what he brings to C, but doing this when bringing things to A provokes an angry response. That is, A and C, through their expectations of the result of using an expression (which are in turn a product of their experience with the expression) impose different normative standards on the use of the expression. Either A's or C's standards could become generally accepted in the community (which standard will depend on a myriad of factors). But the

[8] While cognitive and perceptual terms are not used as descriptions of behaviour, their grammar is nonetheless grounded in behaviour (section 4.3.1).

important point is that regular association can become part of the grammar of an expression by leading individuals to expect certain consequences of the use of that expression.

The development of *mot-* outlined above exemplifies this process. When the word was used in relation to obligation the regular connection between someone's citing another's (or their own) obligations and their belief that what was obliged would be carried through was something that could be learned. For example, when a speaker declares that a third party is obliged to do something, then generally the speaker will, e.g., be willing to make plans that depend on the third party's actually doing what is obliged. As these various connections are learned by hearers, speakers may then be able to exploit the hearers' expectations and, for example, use the expression to indicate that they believe that someone will carry out some action (and, e.g., that they will be willing to make plans on the basis that that action will be carried out). Likewise, the relationship between seeing something and knowing something about it may be a fact learned by individuals leading to a change in communicative affordances and the possibility of using *see* in a knowledge sense without any explicit relation to vision (Andrews 1995). These changes need not be seen as the result of individuals' *innovative* use of language. Instead, they are the result of individuals learning about what generally happens in the world (including how others respond) in relation to expressions, forming expectations on this basis and thereby changing what others can learn.

That semantic changes tend to increase the degree to which the grammar of an expression relates to human behaviour (action, attitudes, etc.) can be seen as a reflection of the fact that human beings are ubiquitous elements of human communication. This fact may be seen as channelling the development of the use of expressions in various ways, which can be broadly grouped into Traugott's (1989) tendencies:

Tendency 1 The tendency for expressions to develop from meanings tied to external circumstances to meanings tied to patterns of human behaviour (that is, to "internal" circumstances: perceptual, cognitive, emotional) may be seen as a reflection of the fact that the uses of "externally based" expressions are regularly associated with such "internal" factors. For example, the development of an expression meaning 'touch physically' to meaning 'perceive through the sense of touch' (a development exhibited by, but not restricted to English *feel*, Sweetser 1990) may reflect the fallibility of first person descriptions of what one is physically touching: thus, it can be learned that when a person says they are touching X, they may not *actually* be touching X but can be relied on to act in a similar manner to the way they would act when actually touching X. This would produce the affordances for using the word with a 'sensory perception' rather than 'physical contact' meaning. Or the regular development from a verb meaning 'to hear' to one relating to heedfulness and later to obedience may be seen as a reflection of the fact that when A cannot hear B, A cannot be heedful to what B says (and so if B wants A to be heedful to what he says, making sure A hears will in some circumstances be important). The connection between hearing speech and being heedful to it may thus be exploited in various ways (e.g. the statement that A did not hear what B said may function as a statement that B did not do as A commanded).

Tendency 2 The tendency for expressions to assume metatextual uses may be seen as deriving from the regular metatextual effect non-metatextual expressions may have. Consider, for example, the development of English *indeed*. According to Traugott and Dasher (2002), prior to 1600 *indeed* (or, *in deede*) was used in the sense of "in the act" and "in truth" (the former giving rise to the latter). The development of discourse marking uses (as a signal that what follows is not only in agreement with what preceded but also counts as additional evidence) can be seen as the exploitation of a discourse effect of mentioning one's commitment to the truth of something.

Tendency 3 The tendency for expressions to come to reflect the speaker's attitude may be thought of as deriving from the fact that speakers have attitudes toward whatever it is they are saying. Consider the development of evaluative uses (pejoration and melioration): suppose A has a certain attitude toward something (X) which he refers to with a non-evaluative expression; B takes a different attitude toward X and uses a different non-evaluative expression; if C has experience of interacting with A and B, then in his experience one expression is used to refer to X by people of a certain attitude while the other expression is associated with a different attitude; C may now contribute to the development of evaluative uses, perhaps by using the expression which (in his experience) is connected to his own attitude, but perhaps more significantly by altering the affordances for other individuals to use the expression evaluatively (through his responding to people who use A's expression on the assumption that they share A's attitude). This pattern of development relies on the coincidence of attitude with expression, and it might be objected that taken in the round the use of a non-evaluative expression would be evenly distributed over speakers with positive and negative attitudes. Note, however, that the described process above involves a positive feedback: the asymmetry in C's experience is then reflected in other's experience with C. Thus local coincidences between expression and attitude may be amplified and spread through a group.

The tendency for expressions to develop subjective uses from uses grounded in objective circumstances reflects in part the fact that speakers can get no closer to expressing the objective truth than by expressing their subjective beliefs. In part, subjectification may also be a reflection of the fact that individuals may be able to learn how to successfully engage in a given language game in spite of not knowing the "objective" rationale behind it.

5.3.4 *Theoretical apparatus*

The communicational affordances account of language change is distinctive in not attributing change to psychological tendencies of individuals to reform their language in particular ways. Instead, explanations for change are located in individuals' learning[9] about the sorts of things that happen in relation to their own and others' activities (including verbal activity), and in the consequences this learning has on their expectations and thereby what is available in the community for others to learn. In this account, the theoretical construct of "representations of meaning" plays no role.[10]

In contrast, many contemporary approaches to semantic change begin with an externalisation based model. For example, Traugott and Dasher (2002) diagram the diachronic relationship between form–meaning pairs (lexemes) as (7) (L is a lexeme, M a meaning):

(7) $$\mathrm{L} \rightarrow \begin{bmatrix}\text{Form}\\ \mathrm{M}_1\end{bmatrix} > \mathrm{L}\rightarrow \begin{bmatrix}\text{Form}\\ \mathrm{M}_1{+}\mathrm{M}_2\end{bmatrix}$$

[9] "Learning" here is not restricted to children learning to speak, but could characterise changes in an individual throughout their life.

[10] There is scope for metalinguistic language games such as "explaining meaning" to play a role, or to be the subject of an account, but they do not assume a privileged position.

This decontextualised representation of form and meaning as entities arbitrarily conjoined gives the impression that semantic change is a process whereby a new meaning is simply added to an extant lexeme. This addition may come about by a speaker innovatively attaching a new meaning to an old form, or by a learner mistakenly attaching the form to a different meaning. These processes (especially speaker innovation) appear to be unconstrained, so psychological principles are postulated to constrain changes to be just those found in the histories of languages. For example:

> In the on-line production of language, [speakers and writers] use mechanisms such as metaphorization, metonymization (including invited inferencing, subjectification, intersubjectification), and objectification in the context of spoken and written discourses. (Traugott and Dasher 2002, p. 34)

> Words do not randomly acquire new senses . . . , new senses are acquired by cognitive structuring. (Sweetser 1990, p. 9)

Metaphorical innovation is perhaps the most widely accepted cognitive processes underpinning change (Hopper and Traugott 2003). Interpreted as simply non-literal use, creative metaphor doesn't seem to impose any regularity, yet various conventional metaphorical uses do conform to regular patterns (e.g. spatial terms often develop temporal uses). In order to account for these regularities, theorists postulate special relationships between meaning domains which often show a conventional metaphorical relationship: meaning representations from one domain (e.g. space) are thought to "map onto" representations from another (e.g. time), and this mapping is thought to underpin the selection of new meanings for an old lexeme. Thus conventional metaphors are seen as a reflection of the structure of the human "conceptual system" (Lakoff and Johnson 1980; Talmy 1988; Sweetser 1990; Narayanan 1997; Feldman and Narayanan 2004; Evans and Green 2006). The particular mappings that characterise our cognitive structures are referred to as "cognitive metaphors".

Theories of cognitive metaphor tend to focus on the abstract description of representation of meaning rather than the plausibility of the individual-based origins of conventionalised metaphors. However, as argued in section 4.4, such theories do not meaningfully describe structures in the brain. If theories of cognitive metaphor say anything about semantic change they say something about the behaviour of humans as individuals; for example that it is "natural" for individuals to (spontaneously) use language in certain metaphorical ways. However, this kind of individual-based thesis is difficult to accept for a number of reasons. First, there is no independent evidence that it is true: conceptual metaphor theorists do not cite instances of spontaneous innovative metaphorical use (employing the conceptual metaphors in question) as support for their theories, but relationships found in languages (i.e. the phenomena their theories are supposed to explain). As argued in section 5.2.2, the inference from a change in language to a tendency of human beings as individuals to reconstruct their languages according to such a change is invalid. Secondly, there is variation across languages in the ways different conceptual domains are related to each other. For example, in Aymara, terms used with both spatial and temporal senses show an uncommon pattern: when referring to the future, Aymara speakers use terms that are also used with a spatial sense of "behind", and the past is linked to "in front" (Núñez and Sweetser 2006). Similarly, in my own experience, in Scotland and the north of England, the phrase "the back of seven" is used in different places to refer to roughly the time 6:50–6:55, to roughly 7:05–7:20, or to roughly 7:45–7:59. Thus, it seems that if individuals *do* spontaneously employ a space–is–time cognitive metaphor, there are various ways in which a particular spatial expression can "map" on to a temporal meaning. This causes a problem for the idea that the conventional metaphors generally have their roots in individuals' spontaneous innovation. Were an individual, for example, to spontaneously use the phrase "the back of seven" (e.g. in making plans to meet someone) he would quickly learn that others didn't understand, either because they would ask him what he meant or because only some people, having guessed correctly what he meant, would turn up at the right time. (From the perspective of communicative affordances, these communicative breakdowns would indicate to the innovator that the spatial expression cannot be used in a temporal sense).

A third reason to doubt the individual-based thesis is that the relations between senses of an expression are somewhat idiosyncratic. For example, epistemic uses of *can* are relatively limited (8):

(8)
- That may not have happened.
- That can't have happened.
- That may have happened.
- * That can have happened.

One way a conceptual metaphor theory of semantic change can accommodate this kind of idiosyncrasy, and the cross linguistic differences in how expressions from one domain maps onto another, is by describing these fact as conventions. But if the theory accommodates convention, then the supposed "naturalness" to individuals of certain metaphorical mappings becomes superfluous as it is conceivable that *un*natural metaphorical relationships and the exclusion of "natural" domain mappings from a language could both be maintained by convention.

Another way facts about conventional metaphors can be accommodated by conceptual metaphor theories is by adding extra properties to cognitive domains on an *ad hoc* basis. For example, a number of languages (including English) have extended the use of expressions originally meaning 'go' and 'come' (in spatial senses) to function as futures. Bybee et al. (1994, p. 267) found a cross-linguistic asymmetry in the uses of 'go' and 'come' verbs as future-markers: seven out of eleven 'come' futures in their database were immediate futures while none of the ten 'go' futures were. Bybee et al. account for this asymmetry by adopting Emanatian's (1992) conceptual metaphor account of future-sense uses of 'go' and 'come' (Fig. 5.2). The model relies on the idea of a (conceptually represented) time-line onto which the speaker projects himself and then describes the path of the subject along the line in spatial terms. Future uses of 'come' involve the speaker projecting himself to a point in the future close to the future event mentioned. The asymmetry is explained thus: "it is only reasonable to suppose that this dislocation of perspective would not usually involve a projection into the distant future but would more often be a point in time near at hand, yielding an immediate future," (Bybee et al. 1994, p. 269). However, it is not clear (and Bybee et al. give no indication) why this is a reasonable supposition. Indeed, Emanatian (1992, p. 9) did not seem to consider this a reasonable supposition on the basis of this metaphor model, as it was constructed for a language (Chagga) in which "there is no overtone of imminence evoked by the metaphorical use" of either the come- or go- futures. With no *reason* or *reasoning* to support their assertion it is wrong for Bybee et al. to say it is *reasonable*; instead it is an *ad hoc* modification to a metaphorical theory, designed to fit the facts it is intended to explain.

From the perspective of communicational affordances, no such ad hoc theoretical paraphernalia is required to understand the difference between 'come' and 'go' futures. The difference may originate in the different distributions of the temporal relationship between an utterance of 'come' and 'go' (spatial) expressions and the action or event that constitutes the reason for spatial movement. The

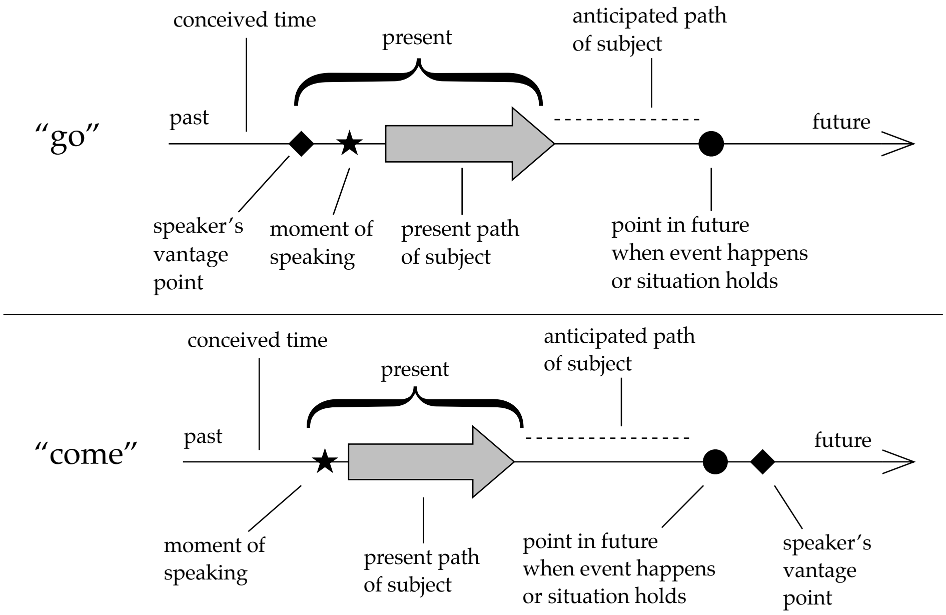


Figure 5.2: Metaphorical accounts of future uses of spatial movement expressions (from Emanatian 1992).

regular use of 'come' and 'go' expressions as futures may be seen as deriving from the fact that a justification for someone's journey is often their intention to do something at the end of the journey (i.e. in the future). If I tell someone that I am going somewhere to do something, this act of telling can occur at any point along my journey except at its end. Conversely if I tell someone that I have come somewhere (or come from somewhere) to do something, we will both be at the location where I intend do whatever it is I intend to do.[11] Thus in instances when I tell someone I am going somewhere and give a reason for my going ("Why are you going to market?" "I am going to sell my pig."), the gap between the moment of utterance and the time at which I will do what I intend will be far more variable than when I tell them I have come ("Why are you here?" "I've come to fix the boiler."). Similarly, when I tell someone that a third party is going somewhere, their arrival may be at any time in the future whereas when I tell someone that a third party is coming here, *more often* the time at which they arrive will be the near future (the relevance of someone's future arrival often being more relevant closer to the time). Thus as individuals learn that reasons given for coming and going are actions that individuals are likely to carry out (e.g. because they intend to), and others then exploit this understanding with non-spatial uses of the 'go'

[11] If this is *not* how the expressions translated as 'go' and 'come' are used, this would be reason for rejecting the translations.

and ‘come’ expressions, what is learned as a future may be different in just the way Bybee et al. (1994) found. The association between an expression meaning ‘come’ (spatial) and immediacy is a tendency rather than a hard rule, and this is reflected in the fact that it is a tendency of future uses of ‘come’ expressions to relate to the immediate future.

The origins of conventional metaphors may be obscure. For example, the expression “time flows” may have its origins in the long and geographically diffuse history of using water as a time-measuring device (e.g. the Greeks used a small earthenware vessel with a hole in its side near the base; speakers in court were allocated the time it took for the vessel to empty, Kenny 2004). Such an origin is now obscured by the replacement of water with mechanical and electronic time keeping devices, and this may lead to the supposition that the origin of the conventional metaphor was a creative metaphor. Indeed, it is conceivable that the origins of our conceptual distinction between literal and metaphorical meaning lie in conventional metaphors which arise by means other than creative metaphor and the obscurity of the source of these inherited patterns. Explanations of meaning that *on the whole* succeed (e.g. explaining the meaning of “flow” by reference to fluids) would be confronted with cases (such as “the flow of time”) which appear aberrant yet are communicatively efficacious; this mismatch could form the basis of a literal/metaphorical distinction which individuals would then be able to exploit in creative metaphor. That is, it is conceivable (though this is only a tentative suggestion) that the relationships between expressions dealt with by proponents of conceptual-metaphor theories are actually the inspiration for rather than the result of creative metaphorical use.

5.4 Projecting communicational affordances onto early stages in the evolution of languages

The theories of protolanguage discussed in chapter 2 attempted to bridge the gap between animal behaviour and human language. If we reject those accounts as based on the unfounded assumptions of externalisation, can an approach based on communicational affordance dynamics offer a better alternative?

5.4.1 Biological evolution

> [O]nly of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious. (*PI* §281)

The description of language development as the shifting of communicational affordances employs an idiom which is purposefully neutral with respect to "cognitive processes" and the differences between humans and other animals: what animals do (and learn to do) with what they find in their environments can be described as the exploitation of affordances. Indeed, some animals can learn to exploit some communicative affordances presented by humans: for example, captive chimpanzees learn (without training) that they can make humans do things by pointing (Leavens, Hopkins, and Bard 2005). What, then, are the biological differences between humans and animals that affect our abilities to use language?

The human brain has undergone pronounced changes in its evolution from the brain of our most recent common ancestor with chimpanzees (Wilkins and Wakefield 1995), and it is likely that those changes have affected our abilities to exploit the affordances offered by others, and the ways we act and react. Chapter 3 argued that we currently have a poor understanding of what the brain does in relation to linguistic behaviour, and chapter 4 urged that our ignorance of such empirical questions cannot be removed by currently popular abstract descriptions of cognition.

However, the brain is not the only aspect of human biology that has changed since the human and chimpanzee lines diverged. Other aspects of our make up with relevance to our abilities to exploit the communicative affordances we make available to each other have also developed. A may learn about the kind of thing B will do in relation to an expression by observing what B does, but this is not the only way. If A can successfully predict something about B's actions on the basis of observing B, this may help A discover details about the roles B plays in communicative practices. For example, if A can tell the difference between cases in which B does and does not understand this will help him learn how to interact successfully with B. Prior to the development of expressions such as "I don't understand" or "what do you mean?" individuals may still have presented to each other cues to their subsequent actions. The human face is interesting in this regard, giving myriad cues to individuals' degree of comprehension, disappointment, anger, satisfaction, etc. etc. The muscular anatomy of chimpanzee and human faces are broadly similar and both are thought to be more complex than more distantly related primates. While many facial expressions appear to be preserved across many primate species (Parr, Waller, and Vick 2007), a number of differences between human and chimpanzee faces leads to a greater range of visually perceptible contrasts: eyebrows, fatty cheeks, everted lips and a bony chin boss serve to make even subtle movements of human facial muscles far more visibly discernible than corresponding chimpanzee muscular movements (Vick et al. 2007). The connective tissue of the human face lies loosely on top of the facial muscles in comparison with the more firm connection found in chimpanzees, and this may make the human face more mobile than the chimpanzee's (Burrows et al. 2006). Another important factor in the visual effect of movement of the facial muscles is the unusual shape and coloration (i.e. visible white sclera) of the human eye (Vick et al. 2007; Kobayashi and Kohshima 2001). Thus the human face is able to display a greater range of visual contrasts than the faces of our closest relatives, allowing greater reliability in the presentation of more fine-grained visual cues to an individual's emotional state and likely behaviour.

In addition to offering cues from which others' behaviour may be inferred and hence learned about, the human face may play a role in opening up communicative affordances to an individual. Recall the experiments described by Masataka (2003) (section 3.3.2 above) which showed that adults differentially respond to children according to the degree to which the child auditorially and visually resembles adults using language (in a culturally variable way). Such cues appear to influence the degree to which adults interpret infant behaviour as communicational and thus the degree to which they offer the opportunity for infants to learn about the kinds of thing adults do in communicative interaction. That animal faces are unable to respond in the same manner as human faces may impact on the degree to which animals can affect (and learn to affect) human behaviour. That is, animals physically cannot give the visual signals that humans are used to and this means that in certain circumstances they cannot appear to humans to (e.g.) understand a human action in the way human infants can. In this sense, the communicative possibilities afforded by mature humans to animals and to other humans may be significantly different as animal faces fail to elicit the range of human behaviours provoked by human infant faces. Consider in this regard differences between human infant and captive chimpanzee pointing: captive chimpanzees only seem to point to request a human pass some object (usually food) while human infants (around one year of age) point "declaratively", simply to draw an adult's attention to something without wanting the adult to pass whatever is pointed to (Tomasello in press). It is possible that human infants but not chimpanzees learn to point declaratively in part because their faces are better suited than chimpanzees' to expressing interest and enjoyment while pointing and when adults respond as if to a declarative point, further prompting the adult to behave in certain ways. The differences between chimpanzee and human faces may act as a barrier to humans playing the recipient role of a chimpanzee's declarative point. Thus whether or not a chimpanzee has the neurological capacity[12] to learn declarative pointing it may not be something their *environment* allows them to learn.

Biological differences between us and our ancestors compound the difficulties attendant on describing the kind of interaction they would engage in. Differences in the degree to which individuals can (in general) "read" each other (which we may attribute to both neurological and morphological differences) would likely impact on their abilities to learn about regularities in each others' behaviour. The effects of this may be difficult to predict: the capacity individuals have to engage with each other in various different kinds of interaction may vary across the evolution of language.

Fitness and exploitation of affordances In the above discussions, the question *why* an individual would exploit the communicative affordances made available to him was not addressed. One way of addressing why organisms do what they do is to ask what the evolutionary advantage of this behaviour is, i.e. what is the function (Tinbergen 1963) or "final cause" (Aristotle) of the behaviour. However, when thinking about the exploitation of communicative affordances, we need not suppose that every move in every language game has a positive impact on individuals' fitness.

Natural selection operates by the differential fitness of individuals with different heritable traits. Thus the relationship between language games and natural selection depends on the range of heritable traits historically present in a population and the impact these traits may have on the way individuals engage with each other. The grammar governing our descriptions of language have nothing

[12] This question is complicated by the fact that the grammar of the expression "neurological capacity" is indeterminate in this context: chimpanzees do not have the capacity to learn declarative pointing, but we have no rules as to how to share the attribution of this lack of capacity between the chimpanzee's face and brain.

to do with either differences between individuals' abilities or the fitness of individuals. It would be an extraordinary coincidence if our arbitrary classification of linguistic practices corresponded on a one-to-one basis to genetic differences between our ancestors and their contemporaries who fared less well.

Consider, for example, the hypothetical development of role reversal imitation outlined above. In that scenario, it is conceivable that individuals who engaged in role reversal imitation at the end of the process were biologically no different to those who did not at the beginning (that is, role reversal imitation *could* develop without a corresponding biological development, though this is not to say that it *did*). Particular differences between humans' and other animals' behaviours cannot simply be assumed to correspond to differences in heritable biological traits.

However, it may well be the case that natural selection has in the past operated on the abilities of individuals to exploit affordances (generally or specifically communicative) or on their display of cues or signals to which others may respond (both visual and auditory). Equally, aspects of our biology which impact on our ability to engage with each other communicatively may have been selected for on bases *other* than their current communicative utility (Fitch 1994). Engaging in meaningful speculation about these possibilities for the various relevant aspects of our biology is beyond the scope of this thesis. What is important to note is that there is no reason to read off genetic differences or selective pressures presumed to have existed in the past from descriptions of our activities as language games, and no reason to demand that each language game we describe should show a positive fitness impact. Heritable differences are selected for because they have a positive fitness impact *in the round*. The impact of biological changes that influence our abilities to interact with each other should be considered in terms of their total effect on fitness, rather than the effect of a limited range of the consequence of change. For example, heritable biological traits unique to humans may form part of an explanation for why we are able to (and do) engage in communicative practices (e.g. giving directions, giving to charity) which can be described as altruistic in a biological sense (i.e. the action reduces the fitness of the actor relative to the recipient). Were these behaviours the *only* effects of these heritable traits they would pose an evolutionary puzzle (as individuals with the trait would be at a disadvantage compared to those without it). However, if such altruistic human interactions result from traits with more general consequences for human behaviour, such altruistic consequences need not imply that individuals with these traits have lower fitness than individuals without them. Thus we should resist the temptations to identify each aspect of our descriptions of language with a regime of natural selection (as Pinker and Bloom 1990 do) and to generate evolutionary paradoxes by presenting language use as primarily the altruistic donation of information (as Dessalles 2007 does).

5.4.2 The emergence of human languages

There are a number of reasons for doubting our ability to theoretically reconstruct a replacement for the accounts of protolanguage reviewed in chapter 2: we currently lack a firm basis on which to predict the likely behaviours of our ancestors given the differences both in their experiences and in their biologies; in addition, if modern languages are seen as the outcome of numerous generations repeatedly exploiting, and thereby changing, the communicative affordances in their environments, then this process likely obscures the interactive practices of the past by continually changing them. However, we may be able to use evolutionary considerations to shed light on the development of the language games we have inherited in spite of these limitations.

The dynamics of communicative affordance exploitation make it likely that widespread conventional communicative practices would have their roots in rather simple interactions between individuals as simpler practices are likely to have spread through a group faster than more complex ones. In addition, a relatively complex practice engaged in by just two members of a group may be imperfectly learned by a third from whom a fourth learns it, etc., diluting the degree to which what develops as a widespread practice (that is, what is common to the way different individuals engage in this practice) is a reflection of the original relatively complex interaction.

A simple interactional practice may involve a certain kind of object, or be dependent on a regularly occurring event, and these relationships may lead to further structuring of practices as individuals learn the affordances offered by these objects or events and by other individuals. Connection to physical objects may have been one way in which more complex communicative practices could have developed and spread through a group. Objects of a given kind would regularly offer different individuals the same affordances, acting as a stabilising force as an interactional practice spreads through a group, increasing the likelihood that two individuals could learn an interactive practice separately from each other (i.e. from different members of the community) yet still be able to engage with each other successfully. That is, the regularities offered by concrete objects may have increased the scope for individuals to interact with others with whom they may not be familiar. In this sense, early communicational practices may commonly have been, to use Traugott and Dasher's (2002) terminology, "based in" the external world (though this is not to say that they could be analysed as having "meanings" analogous to our concrete nouns). Regularities in the physical world may thus have formed a scaffolding on which various early communicative practices could have been based. However, as individuals learn to engage in these practices, the possibility opens up for regularities resulting from structures in human relationships with the physical world to become conventional aspects of language games, relatively divorced from the original physical structuring.

Labelling concrete objects The connection of a communicative practice with a certain kind of object doesn't mean that that object is the meaning of the practice (or an expression in the practice) as urged in chapter 2. For an object to be "the meaning" of an expression may be taken to mean that the use of the expression can be explained to someone (who is already proficient in using words as labels for objects) by pointing to the object. Such an explanation is successful if the recipient then goes on to use the word appropriately in a range of language games in which it plays the role of "referring" to the object (e.g. requesting, fetching, recognising, etc.). At earlier stages in the evolution of language such a range of practices presumably didn't exist. The development of this variety of practices would likely have depended on the role of objects in human life. Various kinds of objects have families of properties relative to human life: they can variously be picked up, carried, passed, fetched, looked for, cooked and eaten, lost, etc., and these roles in human affairs are variously connected. Thus interactive practices relating to one of these properties of a kind of object would exhibit regular connections with other activities, and these connections could form the basis of further developments of language games centred around a kind of object. The commonalities among different kinds of objects mean that different communicative practices related to different objects would have the potential to develop along similar lines. This is a rough and simplistic sketch of the evolution of a nascent repertoire of expressions whose commonalities of use mirrors our use of the expressions whose meaning we can explain by pointing to concrete objects.

As the interactive activities of our ancestors are unknown to us, it is not possible to describe trajectories of change (such as the one outlined for *must* above) for the development of label-like usage (and if these developments happened more than once they may have followed different trajectories). However, as recurrent structuring factors in human life, concrete objects are likely candidates for shaping interactive practices not only across individuals within a community, but also across communities that have no contact with each other. A fuller account of the development of nouns would deal not only with expressions whose use relates to concrete objects, but also the place of these expressions in a broader repertoire of expressions and language games as it is characteristic of a noun that it plays its role in the context of a sentence. This compounds the difficulties attendant on describing pathways by which expressions come to play the modern role of naming concrete objects. However, these considerations suggests ways in which the variety of uses of concrete nouns could develop without attributing to individuals the inborn motivation to use expressions in these ways.

From this perspective, the place of objects in human life is the (historical) scaffolding on which the grouping of objects under linguistic expressions is built. Thus, linguistically coded categories of concrete objects and relationships of category inclusion are a reflection of objects' affordances and the ways these have historically been exploited by humans interacting with each other. This view contrasts with the view that categories are the product of categorisations made "in the mind" and then expressed linguistically (c.f. the theoretical positions discussed in chapter 2). Rosch (1978) was a proponent of a view that the categorisation of objects by a language was product of the way individuals perceived objects' attributes: objects with many attributes in common would fall under the same category. A grave problem with this position is the lack of clarity as to how attributes are to be enumerated and how this enumeration would lead to the use of an expression in relation to various objects. As various schemes can arbitrarily be adopted for the enumeration of attributes, Rosch's proposal risks falling into vacuity: the correct scheme for enumerating attributes could be taken to be the one most consistent with the categorisation of objects in our languages. It is notable that attempts to account for linguistic categorisation by attribute similarity (experiments both by Rosch (1978) and more recently by Malt et al. (1999)) have failed to match linguistic categories to patterns of attributes individuals list in response to questionnaires. From the perspective of communicational affordance dynamics, that various objects fall under a particular label is a reflection of their similar roles in certain communicative situations (that is, their similar roles in human life). For example, objects that fall under the term *furniture* are not perceptually alike, but are alike in that they may often be made of similar materials, made by the same person or group of people, bought in the same place and new items required at a similar time (e.g. when moving house), and these similarities underpin the communicational dynamics from which the particular structure of the category *furniture* developed.

From the outer to the inner The grammars of our concepts which may be characterised as relating to our "inner" lives (i.e. thoughts, beliefs, emotions, feelings, intentions, etc.) are particularly complex, relating to each other, to more concrete concepts and to characteristic human behaviours. As noted in section 4.3.2, the rules for the application of these concepts to an individual may relate to stretches of behaviour, to generalisations about the kind of thing a person might do and to their experiences and actions in other contexts. How could such convoluted and involved practices have developed. From the perspective of communicational affordances, there is little reason to presume that an individual could somehow spontaneously imbue an expression with such a meaning, as this assumption amounts to attributing to the individual the ability to create the rules for the complex use of such an expression.[13]

Learning to exploit the communicational affordances of one's environment involves learning about the ways other individuals behave. The greater their communicative repertoire, the more scope there is for learning cues to the kinds of things they are likely to do and experiences they have had. That is, as engagement by individuals in practices "based in the external world" (as described above though not necessarily of the complexity associated with e.g. referring to objects) becomes widespread, more regularities across individuals' behaviours develop and can be learned (thereby enabling further regularities to emerge). For example, a practice widely adopted which is connected to the presence of some object can form the basis of a signal that an individual believes that object to be present (i.e. an individual can learn that when others engage in this communicative behaviour, whether or not the object is present, this is a cue that they will act as if the object were present); without the communicative practice, only the presence of the object would serve as an (unreliable) cue that another's object-relevant behaviours can be expected. While this development would not in itself be the development of a concept relating to the "inner" (i.e. a concept of 'belief'), it does illustrate a way in which language games could develop in this

[13] As things currently stand, it is possible for an individual to imbue an expression with such a meaning (e.g. "when I say 'abracadabra' I mean 'toothache'," c.f. *PI* §665). However, this depends on the prior existence of a word or expression with the requisite grammar.

direction once an initial structure built on the scaffolding of the "external" was in place.

The development of language games proposed here is a process whereby individuals develop expectations on the basis of regularities and thereby create new regularities for others to learn from. Common characteristics of human behaviour (that we cry out when injured, smile when amused, avoid and seek situations that resemble situations which have harmed or pleased us in the past, that certain situations can make us freeze with fear after which we may run away or proceed with caution, etc. etc.) provide opportunities for language games to develop on these bases. For example, modern Japanese *kowai* (whose uses can be glossed as "be (physically) stiff, tough" and "fearsome, be fearful") derives from from the Old Japanese *koFasi*, "(physically) stiff" (Traugott and Dasher 2002, p. 95).[14] The development can be seen as reflecting the physical relationship between stiffness and fear (this relationship need not be direct: physically stiff things are relatively immobile and a characteristic of certain instances of fear is remaining motionless; indeed it seems *kowai* developed a number of uses relating to physical immobility including "exhausted", "painful" and "embarrassed" most of which have since been lost). This regularity would have allowed, e.g., the mention of someone's being motionless to be used (in some contexts) as an expression of how afraid they are.

By setting the development (and maintenance) of our concepts in the context of general aspects of human interaction (and other behaviours) we can get a handle on why the grammar of these concepts is as it is.[15] Consider, for example, the visually based use of the English verb "to see". The grammar of this word relates to various activities humans can engage in: someone who can see can, *inter alia*, reach out and grasp objects infront of them, navigate around obstacles without touching them, describe what something put before them looks like, recognise things they have seen before, point to things in the distance, make judgements based on the movement of objects infront of them (e.g. judge when one moving object will hit another), etc. Various of these abilities that fall under the concept of "seeing" can be selectively impaired by neurological damage (see Greene 2005 for a review). Why then do these recognisably different abilities fall under the same concept? (The answer cannot be that our *experience* is similar when

[14] Japanese words are represented here using Traugott and Dasher's (2002) phonemic representation.

[15] This is not to say that the grammar is not arbitrary in the Wittgensteinian sense that the grammar of a concept cannot be right or wrong.

performing each of these feats, as this answer relies on the grammar of "experience" and more particularly of "visual experience", and so simply defers the question.) For most individuals (and throughout the history of the verb), these abilities have been highly correlated: damage to the eyes generally removes these abilities from individuals on a permanent basis, and individuals whose eyes are closed or covered or who are in a generally dark environment have temporarily reduced powers to perform these acts. Thus in spite of the variety of abilities that grammatically make up the concept of *seeing*, their relationship to light and the undamaged eye groups them together communicationally (so if A is told "he is unable to see" by B as an explanation of C's failure to grasp an object in front of him, A can safely assume that C will be unable to navigate around obstacles without touching them). The relationship between seeing and undamaged/uncovered eyes or good light creates a pathway for expressions used in relation to the eye or to light to develop into vision verbs, and these pathways have been followed a number of times in Indo-European languages (Sweetser 1990).

Thus, the relationships between these human abilities can be appealed to to account for why our concept of *seeing* (or *vision*) has developed as it has. Were humans endowed with echo-location as bats are, the communicative affordances approach would predict that a concept matching *to see* would be unlikely to develop (and in its place would be a different concept or concepts). (This is not to say that seeing means having undamaged eyes and being in the light, as the statements "he can see in the dark", "he can see even though his eyes are damaged", and "he can see even though he has no eyes!" are not contradictions.)

This sketch of the development of the concept of *vision* appeals to structuring factors imposed by human biology and by physics. However, structures that shape communicative practices and create pathways for the development of grammar need not all relate to such basic and universal aspects of human life. The structure of conventional interactive practices may also be a ground for the development of new ways of using expressions, as in the case of *mot-* where the social structure of obligation (supported by conventional language games) and anticipated consequences of obligation opened a pathway for epistemic uses of *must* to develop. In this sense, language is a constantly emerging phenomenon in Hopper's (1987) sense: as individuals learn to exploit the affordances and structure in their environment, their action produces new environmental regularities which may then impact on the development of other related language games.

The communicational affordances view of the evolution of language allows us to replace the idea that it is natural for humans as individuals to spontaneously use expressions relating to our "inner lives" with historically contingent processes on which various structuring factors impact. We may expect different unrelated languages to develop similar concepts not only because of similarities in the ways humans respond to certain kinds of experience, but also because of similar structuring factors impacting on different groups' lives. That concepts with similar grammars may emerge in parallel in different communities, therefore, does not imply that these concepts are inevitable parts of language.

> [I]f anyone believes that certain concepts are absolutely the correct ones, and that having different ones would mean not realizing something that we realize — then let him imagine certain very general facts of nature to be different from what we are used to, and the formation of concepts different from the usual ones will become intelligible to him. (*PI* p. 230)[16]

If the development of language games relies on regularities in the world (including in human behaviour), then their institutionalisation can be seen as offering a way in which individuals can learn about those regularities. Take the concept of belief: from the communicational affordances perspective, there are likely various ways in which an expression of this concept could develop. One route relies on the existence of communicational practices tied to observable events, such that accidental transgressions of the ordinary rules for using an expression (namely the presence or existence of the observable events), based on individuals' beliefs, can be observed and learned about by other individuals. Once a concept of belief has developed and reflects in its conventions the possibility for differences between individuals' beliefs and reality, the possibility of such differences can be learned by learning the rules for the expression rather than having to discover the possibility (and circumstances) of such disconnect. This contention is supported by experiments that show adults who have learned a language which does not have expressions corresponding to English *believe* perform poorly on (non-linguistically presented) false belief tasks in comparison

[16] The context of this quotation is Wittgenstein distinguishing his attempts to clarify the grammar of our concepts (in order to clear up philosophical problems) from the kind of attempt undertaken in this chapter to account for the formation of our concepts.

with younger individuals who have learned a similar language which does have such expressions (Pyers 2006).[17]

5.5 Summary: communicative affordances

This chapter has developed a way of thinking about the evolution of language, based on evidence from semantic change. Rather than seeing language as the external manifestation of internal structures, it is seen as a series of techniques by which individuals can interact with each other. As individuals learn how to engage using these techniques and exploit the affordances available to them, they alter the possibilities available to others for interaction.

Much of the discussion about the early development of language has been suggestive of trends and possibilities rather than the presentation of a clear history of our language games. The picture of language evolution presented here is one of constant emergence, contemporary languages being the result of long historical processes. According to this perspective, communicative practices are shaped by various factors impacting on human life, not least other existing practices. This layering of language games on top of old language games makes reconstruction of practices deep in our history on the basis of trajectories of language change likely to be practically impossible. This, coupled with the difficulties associated with using our current understanding of the human–as–machine to reconstruct human and pre-human behaviours in very different communicational contexts, puts concrete speculation about earlier communicative practices beyond our current reach. This is especially true for practices which did not leave (physically) fossilised consequences.

While such reconstruction may not be possible, the communicative affordances approach suggests that it may not be entirely necessary when developing an evolutionary based account of the development of certain basic aspects of our language games. While we may not be able to present the rules for early interactions, I have suggested that they would have been based in human interactions with concrete objects since these would have been a source of regularities in (pre-) human life. The structuring of human affairs around concrete objects was appealed to as opening up the possibilities for myriad interrelated language

[17] Of course, learning the rules for using the word "belief" is not the *only* way an individual could come to appreciate differences in people's beliefs. However, as an *additional* route by which an individual can learn such facts, language may be seen as facilitating the development of such an appreciation.

games to develop, thereby producing expressions whose use we can characterise as "referring" to objects. I also suggested that the development and spread of language games on the basis of regularities in the roles of physical objects in human life may have provided the scaffolding on which expressions relating to our internal lives were built. Neither of these accounts amounts to a clear history of these language games (and indeed, these may have various different histories), and the language games dealt with do not exhaust the ways we use language. What I hope to have achieved in this chapter is to demonstrate that it is possible to approach questions of the origins and evolution of our various language games without having to rely on the notion that these come spontaneously to humans as individuals (through the externalisation of internal representations or by any other means).

The focus of this chapter has been on the development of rules for the use of an expression, rather than on changes in the forms of expressions. It may be possible to approach changes in form in an analogous manner, by considering the affordances associated with various forms in the kinds of communicative context the generally occur. For example, variation in form is more communicationally tolerable when interlocutors can tell what is likely to be said, and this may apply to the trends in the development of spoken form that are part of grammaticalization (Hopper and Traugott 2003): in contexts where an expression is highly predictable, individuals may be able to guess what the speaker intends even if the form of the speaker's utterance shows a greater degree of divergence from others' utterances of the "same" expression than the normal range of variation communicatively acceptable in other contexts. That is, the affordances for speakers in these contexts allow relatively large variation in spoken form. This tolerance of variation may lead to phonological reduction (rather than any other kind of change) for a number of possible reasons: there are fewer ways to change a form by reduction than there are by addition or substitution, so if convention changes by what is most common changing, it is likely to be in the direction of reduction; additionally, a state in which phonological material has been reduced may be a sink (both because individuals rarely add phonological material and because if they did individuals would likely add different sounds). Thinking about development in terms of shifting communicative affordances may also be applied to morphosyntactic aspects of language, as regularities of expression become conventionalised (see Hawkey 2008). Again, these brief notes are not exhaustive of phonological and morphosyntactic change, but they are intended to indicate how an approach that doesn’t rely on hidden mechanisms can be brought to bear on these issues.

In accounting for the development of the grammar of our expressions, great emphasis has been put on the ways usage can be structured and these structures adopted as conventions. The next chapter takes up these ideas and applies them as a quantitative account of commonalities in the ways people in unrelated linguistic communities use colour in their communicative activities.

Chapter 6

The evolution of colour terms

Relationships between different languages' colour terms and the evolution of colour term systems has a long history in debates about linguistic relativity, the question being something like "does language determine colour perception, or vice versa?" (e.g. Kay and Maffi 1999). In a particularly influential study, Berlin and Kay (1969) collected native speakers' judgements as to which colours (presented on a fine grained colour chart) fell under which of their languages' colour terms. That study concluded that there are universal properties of languages' colour term systems, a conclusion that has commonly led to the inference that colour terms are (somehow) a reflection of internal representations of colour (e.g. Mervis, Catlin, and Rosch 1975; Kay and McDaniel 1978; Dowman 2003; Belpaeme and Bleys 2005a). This chapter examines some evidence for universal structures in languages' relationships to colour and tries to account for these from the perspective of communicational affordances. Colour categories are seen not as the reflection of individuals' internal representations but as the product of universal aspects of the communicationally relevant colour environment of human beings.

6.1 Universal structure in colour-chip naming

Berlin and Kay's (1969) study was criticised on several methodological grounds: small sample sizes (both in terms of languages and numbers of speakers), informants were English bilingual, judgements were not elicited in native contexts, simultaneous presentation of colours on a grid may have biased speakers' judgements, and results were based on experimenters' subjective interpretations of speakers' responses (Kay and Maffi 1999; Lucy 1997). In 1976, the World Colour

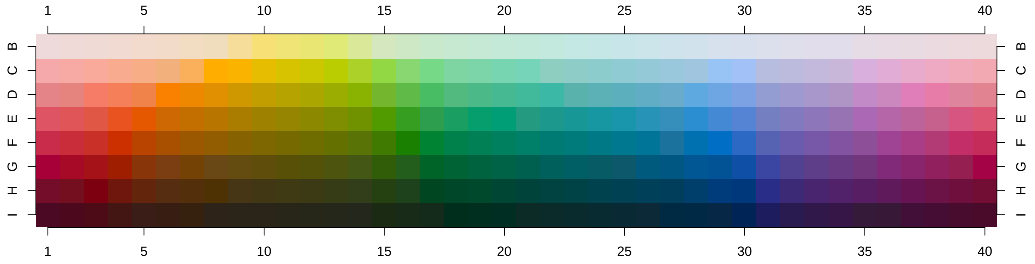


Figure 6.1: The Munsell array used in the WCS. (Reproduction of colours throughout this chapter serves only as an approximate guide to Munsell colours.)

Survey (WCS) was initiated to address these methodological concerns. The WCS gathered judgements of monolingual (as far as possible) speakers *in situ* for 110 unwritten language spoken in small scale, non-industrialised societies. An average of 24 speakers per language were consulted.

The WCS uses 10 "neutral" (black/grey/white) and 320 chromatic chips produced by the Munsell Color Company, designed to represent 8 gradations of lightness and 40 gradations of hue at the maximum available saturation. Fig. 6.1 is an approximate representation of the array of coloured chips. These Munsell chips were presented to speakers one by one in a fixed random order. The instructions given to field linguists collecting these data asked them to try to elicit "basic colour terms" from their informants (see section 6.3.1 below).

6.1.1 Using colour space to analyse the WCS

The WCS collected a total of 839,243 judgements[1] from 2,616 speakers. Analysing this data presents a formidable challenge. Berlin and Kay's (1969) original analysis had relied on visual inspection of the arrangement of chips falling under a given term; such analysis for the WCS data would be impractical and suffer the charge that results were based on subjective judgement. To deal with these problems, Kay and Regier (2003) used the positions of the Munsell chips in CIE L*a*b* space[2]: for each term used by each speaker, a single point (a "speaker centroid") in CIE L*a*b* space was identified as the average position of the chips

[1] Plus 24,037 instances where no judgement is recorded.

[2] CIE L*a*b* space is a system for representing colours in three dimensions such that distances correspond to certain psychophysical measurements. See section 6.1.1.1.

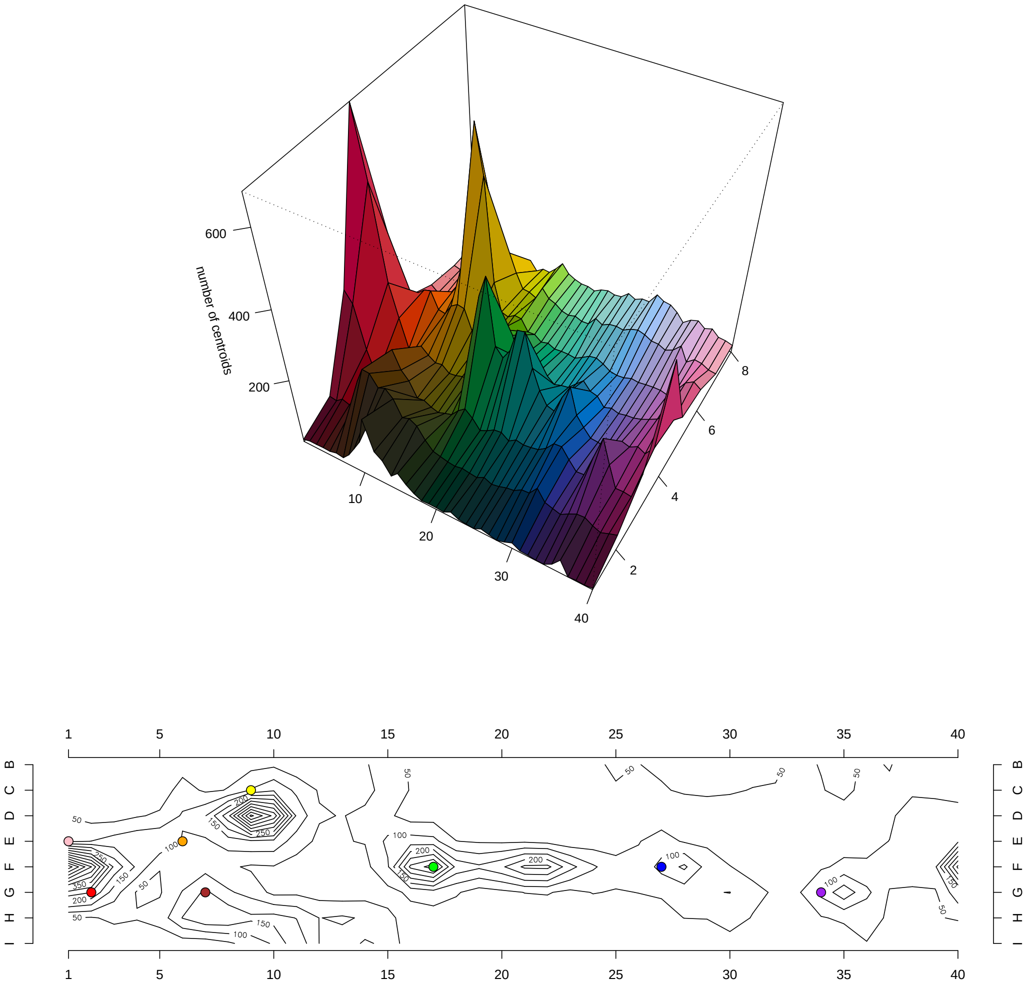


Figure 6.2: Histogram and contour plot of coerced speaker centroids using the WCS data. Histogram floor plane represents the positions of chromatic Munsell chips in the array (Fig. 6.1) and height represents the number of speaker centroids coerced back to that chip. (Histogram colours serve only as a rough guide to chip colours.) Coloured circles on contour plot indicate (from left to right) English terms “pink”, “red”, “orange”, “brown”, “yellow”, “green”, “blue” and “purple”.

that speaker labelled with that term; this position was then “coerced” back to one of the WCS Munsell chips, and the numbers of speaker centroids coerced to each chip were counted. This method of analysis produces particularly striking results, shown in Fig. 6.2.

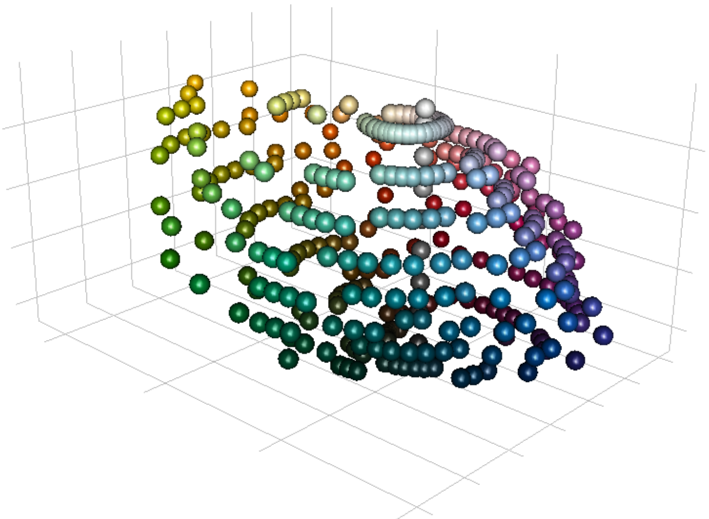


Figure 6.3: Three dimensional representation of the uneven distribution of WCS Munsell chips in CIE L*a*b* space

Kay and Regier (2003) compared the peaks of this distribution to the coerced speaker centroids for English colour terms: English "blue", "green", "purple" and "brown" fall at or very near peaks in the WCS distribution while "red", "pink", "orange" and "yellow" fall "in the neighbourhood of WCS peaks," (p. 9089). The number of coerced centroids falling on the same chips as English term centroids was significantly greater than the numbers falling elsewhere on the Munsell array.

While Kay and Regier interpret these results as revealing common structure in the WCS languages' colour term systems, there are in fact two possible sources of structure in the distribution of centroids: speakers' judgements and the distribution of Munsell chips in CIE L*a*b* space. Figs. 6.3 and 6.4 show the uneven distribution of Munsell chips in CIE L*a*b* space: yellow chips form a bulge at higher L* values, and at each L* value chips have variable radii (saturation) and angular (hue) separation, forming discernible clumps. The source of these irregularities lies in the irregular relationship between the Munsell and CIE L*a*b* systems.

The procedure for choosing a chip to represent a speakers' term is rather complex involving several different notions of "distance".[3] Thus the effect of the uneven

[3] Speaker judgements are converted into a point in CIE L*a*b* space on the basis of Euclidean distances (the point is the position at which the mean squared Euclidean distance to all the chips

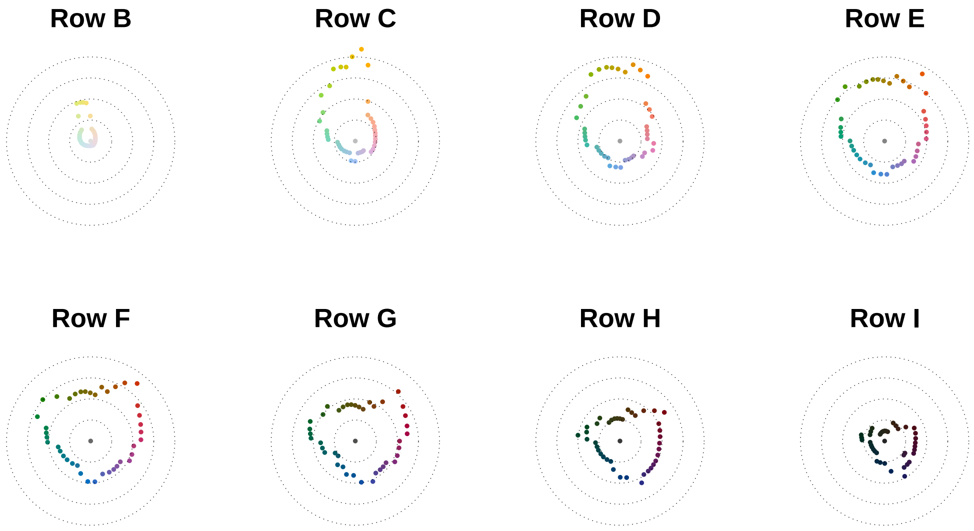


Figure 6.4: Two dimensional representation of uneven distribution of WCS Munsell chips in CIE L*a*b* space. Planes of constant L* are shown with rings of equal CIE L*a*b* radius indicated by black dotted lines.

distribution of Munsell chips is also rather complex. In order to examine these possible effects without going through them analytically, I asked what results would be expected if the same centroid/coercion procedure was performed on a randomised data set which did not contain the structure of the WCS speakers' responses. I generated random data sets by rearranging the WCS judgement data according to a procedure outlined by Regier et al. (2007). Each speaker was assigned a random number (a "speaker shift") between zero and 39 and their responses were translated across the Munsell array by this number of chips. If, for example, a speaker was assigned the number 23, each of their judgements for chips in column 1 of the Munsell array would be used as judgements for column 24 in the randomised data set, etc. Judgements shifted beyond the end of the

covered by the term is minimised). However, chips are coerced back to Munsell chips not on the basis of which chip is nearest in Euclidean terms, but according to the following rules: (1) the coerced chip is from the row on the Munsell array that has the nearest L* value to the centroid, (2) the coerced chip is either the neutral chip from that row or the chromatic chip with the nearest hue angle, (3) of these two chips, the coerced chip is one whose radius in the a* b* plane is closer to the average radius of the chips represented by the centroid than the radius of the neutral chip (Kay and Regier 2003).

Munsell array were “wrapped round” so, for example, the judgements in column 33 of speakers assigned a shift of 23 would appear in column 16 (= 22 + 33 - 40) of the random data set. This method of randomisation preserved some features of the original WCS data such as distributions of the numbers of terms per speaker, number of chips per speaker’s term, relationships between terms and lightness, and the spatial relationships between a speaker’s terms on the Munsell array (namely roughly contiguous patches being named with the same term, contiguous patches wrapping around the array horizontally but not vertically, etc.). These features would be lost if randomisation was done by assigning to each speaker’s term a random subset of the Munsell chips. Randomly shifting the data in this manner affects only the relationship between Munsell hue (i.e. column on the Munsell grid) and speaker responses, so if the peaks in Fig. 6.2 reflect only the structures of the WCS, randomly shifted data should not produce within-row peaks when analysed into centroids.

Fig. 6.5 shows the result of performing the centroid/coercion operations on a randomly shifted data set. A number of peaks (e.g. around the red and green regions) are clearly identifiable, indicating that the centroid/coercion method in conjunction with the CIE L*a*b* space favours certain chips over others. Thus randomly generated data produces similar evidence to that which prompted Kay and Regier (2003, p. 9089) to claim that “certain privileged points in color space appear to anchor the color naming systems of the world’s languages.”

In order to test whether the WCS results in Fig. 6.2 could plausibly be a reflection of these biases, I performed Spearman’s rank correlation tests on the coerced centroid counts produced by the original WCS data and 20 randomly shifted data sets. Shifting the data does not affect the relationship between individual speakers’ terms and the rows of the Munsell grid, so the number of centroids falling in each row is the same for all data sets. As some rows have more centroids coerced to them than others, significant correlations between chip counts across data sets would be unsurprising (as, for example, chips in row F would generally have higher numbers than chips in row C for all data sets). For this reason, correlation tests were performed on a row by row basis. In addition to the twenty tests per row comparing WCS data with the twenty random data sets, random data sets were compared with each other to assess whether the peaks in Fig. 6.5 were consistently produced (this produced an additional 190 tests per row). As the focus here is on a possible bias towards certain chromatic chips, neutral chips

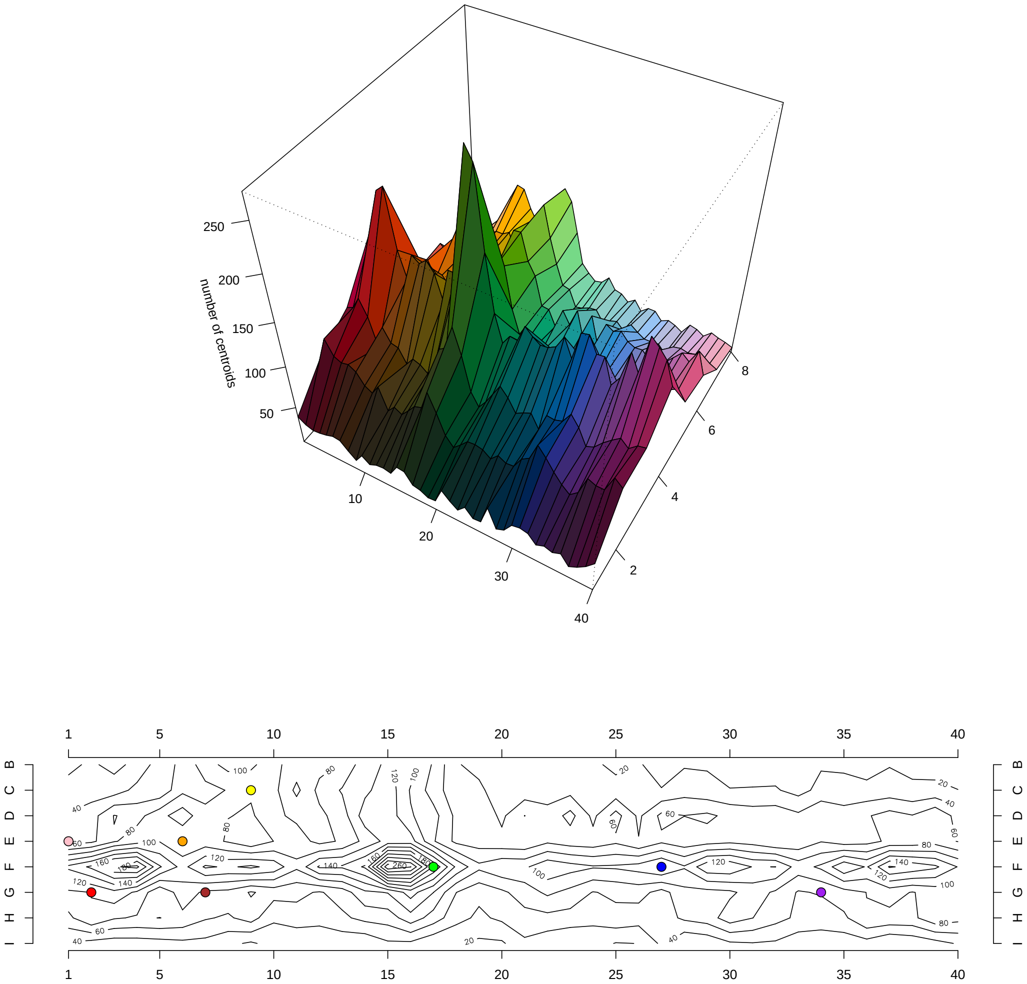


Figure 6.5: Histogram and contour plot of coerced speaker centroids using randomly shifted speaker responses. Coloured circles on contour plot represent English terms as in Fig. 6.2.

were omitted from the analysis. A summary of the numbers of tests reaching significance ($\alpha = 0.01$) is shown in table 6.1.

In every row of the Munsell array other than the darkest, *all* the pairs of randomly shifted data sets produced significant correlations, indicating that the peaks in Fig. 6.5 are the result of a biased procedure rather than a random distribution. Rows D and E show significant correlations between WCS data and every random data set indicating that in these rows the distribution of coerced centroids mirrors that produced by the biases in the analysis. It is possible that the single

Munsell row	*WCS vs shifted*	*Shifted vs shifted*
B	0	190 (100%)
C	0	190 (100%)
D	20 (100%)	190 (100%)
E	20 (100%)	190 (100%)
F	0	190 (100%)
G	1 (5%)	190 (100%)
H	0	190 (100%)
I	0	144 (76%)
Total	41 (26%)	1474 (97%)

Table 6.1: Significant correlations between coerced centroid counts produced by the WCS data and 20 randomly shifted data sets (Spearman's rank correlation, one tailed, $\alpha = 0.01$).

correlation in row G is the result of chance (excluding rows D and E there were 120 tests at $\alpha = 0.01$ so the expected number of type 1 errors is 1.2).

In summary, the analysis performed by Kay and Regier (2003) employs a biased method, but the results are not attributable solely to this measurement bias, suggesting there is non-random structure in the WCS data. The chip counts in Fig. 6.2 are the result of looking at this structure through the lens of a biased procedure.

6.1.1.1 What is colour space?

The biases involved in the centroid/coercion method raise the question, Why did Kay and Regier (2003) use the CIE L*a*b* colour space to generate surrogates for speaker's naming patterns? In their paper they describe the CIE L*a*b* colour space as being something with which "psychologically meaningful distances can be calculated" (p. 9086), suggesting that they view the centroid of a number of chips labelled with the same term as being psychologically meaningful. But what psychological meaning could centroids have?

The CIE L*a*b* colour space is based on two kinds of psychophysical experiment and an attempt to produce a useful tool for the production of colour (e.g. to help engineers make colour televisions which maximise the range and quality of colour). The first kind of psychophysical experiment is colour matching: subjects with normal vision in laboratory conditions are shown a test light and have to vary the intensity of three standard lights until their combination appears to match the test light. Because human colour vision depends on three kinds of photoreceptor in the eye (section 6.4.1 below), it is possible to match any spectral distribution with three lights. A set of standard lights is the CIE XYZ lights: any spectral distribution of light can be converted into XYZ values such that, under laboratory conditions, that spectral distribution will appear the same colour as the corresponding mixture of X, Y and Z lights.[4] These three values can be thought of as a space: different spectral distributions at the same "point" in XYZ space will be seen as the same colour by a standard observer, and a straight line between two points identifies the XYZ values that can be created by mixing lights represented by those points. Distances in this space are not, however, psychologically meaningful (Wyszecki and Stiles 1982).

The second kind of experiment used to create CIE L*a*b* space involve discrimination. Subjects are shown colour patches (whose XYZ values are known) and asked to perform some kind of task requiring them to discriminate on the basis of colour (e.g., select the odd one out). When differences between test and target colours are relatively large, subjects are able to discriminate on every trial. As colour differences become smaller, performance degrades gradually (as opposed to falling directly to chance levels below some threshold). Thus, for a given set of XYZ values, experiments can find which XYZ values subjects will be able to discriminate on (e.g.) 85% of trials, which values will be correctly discriminated on 70% of trials, etc. CIE L*a*b* coordinates are derived from CIE XYZ by a mathematical manipulation designed to produce a space in which pairs of lights which are equally discriminable will be separated by equal Euclidean distances.[5] CIE L*a*b* achieves this goal approximately (though better than CIE XYZ space, Wyszecki and Stiles 1982).

The CIE L*a*b* space is a tool for estimating whether two colours will be discriminable from each other, and Euclidean distances in CIE L*a*b* are better described as psychophysical than psychological.[6] CIE L*a*b* distances are expressed in terms of "ΔE" units, one ΔE being roughly the threshold level of colour difference. At distances significantly greater that one ΔE, the notion that distance corresponds to discriminability becomes meaningless (as a pair of lights

[4] CIE X, Y and Z are actually mathematical constructs rather than real lights: there are wavelength regions over which the energy they emit is negative.

[5] XYZ values do not determine L*a*b* values uniquely as conversion is done relative to a variable "white point".

[6] Certainly, Belpaeme and Bleys's description of it as "a representation of the psychological color experience of the average human," (2005b, p. 297) is, to the degree it is comprehensible, an overstatement.

cannot be *more* discriminable than a pair which can be correctly discriminated on every trial). For the WCS Munsell chips, the average distance from each chip to its nearest neighbour is about six ΔE units. What this means is that between one chip and its nearest neighbour six evenly spaced colours (in CIE L*a*b* space) would be just discriminable from each other. On average, a speaker's term covers about 40 Munsell chips, so generally the colours of the chips labelled by a given term and the colour represented by those chips' centroid are easily discriminable. What, then, should the interpretation of a centroid in CIE L*a*b* space be? The grammar of the CIE L*a*b* space is analogous to a table or an almanac for looking up certain facts about colour discriminability. As such, averaging over points in this space has very little meaning. At best the centroid is something like a representation of the colour for which the average number of just discriminable colour differences between that colour and the Munsell chips labelled by a given term is minimised.[7] But it is unclear what relevance this colour (which is in part dependent on the arbitrary colours of the Munsell chips) has for the use of the colour term.

However, if CIE L*a*b* space is misinterpreted as a representation of the inner workings of humans as they see colours, the centroid may be confused with something to do with how a human judges that a colour should fall under a given term. This misinterpretation of colour space as something in the mind/brain, whereby "*the reality* of the colour model comes to be confused with *the model* of reality" is not uncommon (Saunders and van Brakel 2002). Indeed, Regier, Kay, and Khetarpal (2007) suggest that colour naming is determined by an "optimal partitioning" of this colour space and the irregular distribution of Munsell chips in it. These unwarranted extensions of a mathematical model (of the results of specific psychophysical investigations) on the grounds that it is (somehow) a mechanistic description of colour vision parallels the use of the models of theoretical linguistics to infer linguistic behaviour of our ancestors on the ground that the theoretical model is (somehow) instantiated in the brain (c.f. section 4.4.3).

Colour spaces are tools for organising and reproducing colours (Saunders and van Brakel 2002). Grammatically they have nothing to do with the processes by which a language develops and maintains terms which fall under the family resemblance category of "colour terms" or the processes by which a human–as–machine describes the colour of an object. As such they offer a poor basis on which to construct an analysis of the WCS data or a theory accounting for any

[7] A centroid will only get this point roughly as it minimises mean *squared* distances.

cross-linguistic structure. Because the centroid of a collection of Munsell chips in CIE L*a*b* space has no meaning relevant to colour terms, there is no sense in which the bias detected above could be corrected for as there is no relevant meaning to the notion of the *correct* or unbiased centroid of a collection of chips.

6.1.2 *Analysing the WCS without a perceptual colour space*

An alternative approach to analysing the WCS data is to ask, for each chip, to what extent do speakers of a language agree on the appropriate colour term, and then to compare chips to see if some are generally more consistently named by speakers of a given language than others (c.f. Lindsey and Brown 2006). To measure the agreement between speakers, I calculated the Shannon entropy of the responses from a given language for each chip as follows: for chip c, the proportion of speakers of language l who used term t was calculated (and is represented here as $p_{l,c,t}$). The "naming entropy" of a language for each chip ($H_{l,c}$) was then calculated according to equation (1).

$$H_{l,c} = -\sum_{t} p_{l,c,t} \log_2(p_{l,c,t}) \qquad (1)$$

Shannon entropy is a measure of information, and in this case $H_{l,c}$ can be interpreted as a measure of how informative the information that a chip was labelled with a certain term is to identifying which speaker produced that judgement. If all the speakers of a language use the same term for a chip, then being told that an unknown speaker produced that term for that chip does not help identify the speaker, and the entropy is low. In contrast, if all the WCS speakers of a language use different terms, then being told what term an unknown WCS speaker produced for that chip will uniquely identify them and the entropy will be high. Judgements are informative in identifying the speaker (i.e. entropy is high) when agreement is low, and vice versa.[8]

Entropy values, averaged over languages, are shown in Fig. 6.6. This figure bears some resemblance to Fig. 6.2 (Kay and Regier's 2003 analysis): roughly put, red

[8] Entropy is a better measure of agreement among speakers than the proportion of speakers using the most popular term for a given chip (the measure used by Lindsey and Brown 2006) as it takes account of the use of all terms. For example, if the proportions of speakers using terms a, b and c for a chip are 0.51, 0.49 and 0.0 respectively, the entropy will be lower than would be the case were the proportions 0.51, 0.25 and 0.24 respectively, reflecting the greater consistency in the first case than the second.

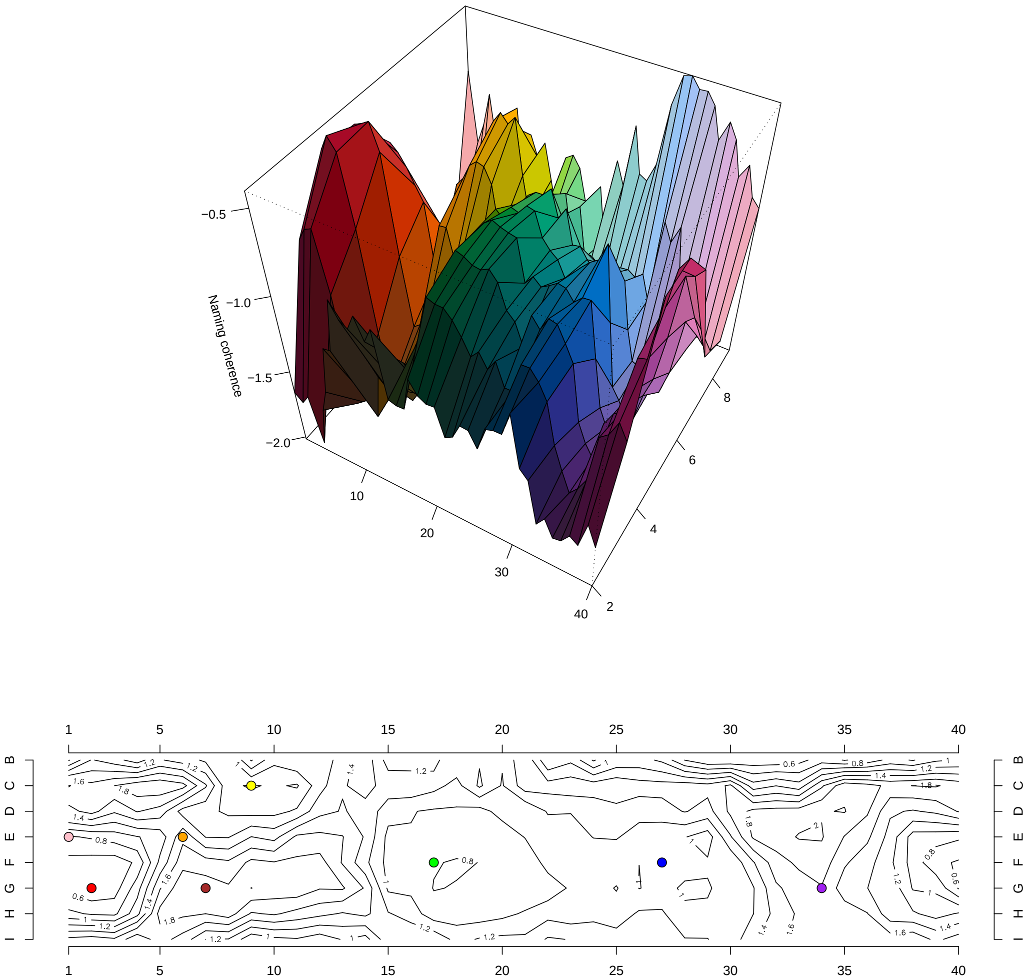


Figure 6.6: Coherence of WCS responses (measured by chip entropy) averaged over languages. z-axis on surface plot represents $0 - H_c$ so higher values represent greater agreement among speakers.

and yellow areas stand out, as does the green/blue area. However, a striking difference between the two can be seen around the chip identified as representing English "purple" (G34). While Fig. 6.2 shows this area to be a peak in terms of number of centroids (leading Kay and Regier 2003 to claim this as one of eleven "privileged points in colour space [which] appear to anchor the color naming systems of the world's languages", p. 9089) Fig. 6.6 shows this is an area for which speakers generally do not agree on an appropriate colour term.

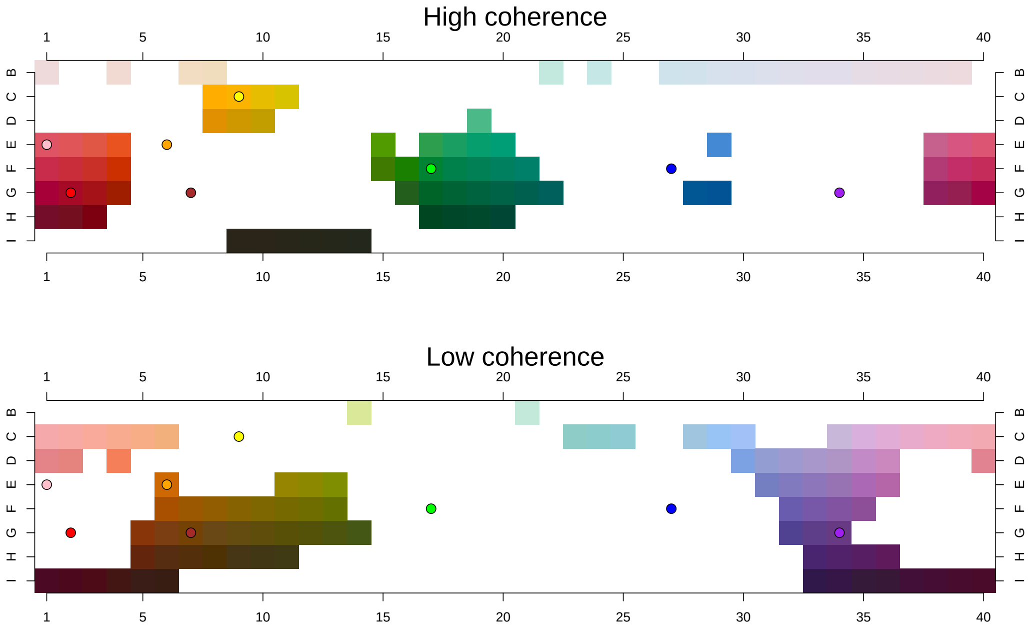


Figure 6.7: Munsell chips whose distribution of naming entropy values was significantly different to the distribution over other chips. Chips with lower entropy values (higher coherence) are shown separately from those with high entropy (low coherence). Coloured circles represent English terms from Kay and Regier (2003).

Naming entropy measures were subjected to *t*-tests to determine which chips' distributions of coherence measures were significantly different to the global distribution (i.e. the distribution of values over all chips *except* the chip being tested). As the global distribution ignores language and chip information, it is equivalent to the distribution that would be obtained if responses were randomly shifted across the Munsell array on a language-by-language basis. Thus chips with significantly different coherence measures to the global distribution indicate common structure across the WCS languages (both in terms of chips whose naming is coherent and incoherent across languages). A Bonferroni corrected significance level (correcting for 330 tests) of $\alpha = 0.00016$ was used to give an overall significance level of 0.05.[9] Two-tailed tests were conducted (mean Welch-corrected df=109.7), and figure 6.7 shows the chips for which tests showed significance.

[9] What this means is that if the null hypothesis that no chip has a different distribution of values to the global distribution is true, the probability that one type 1 error (false positive) will be made in the 330 tests is 0.05

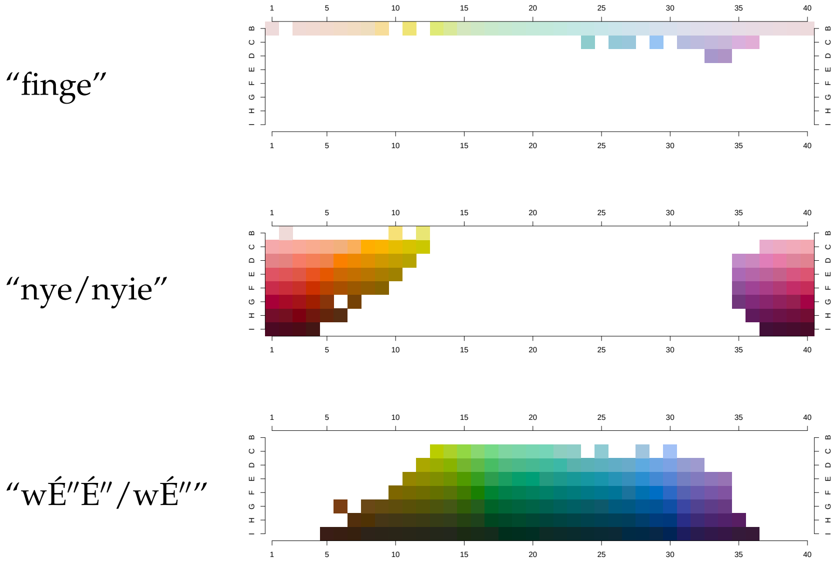
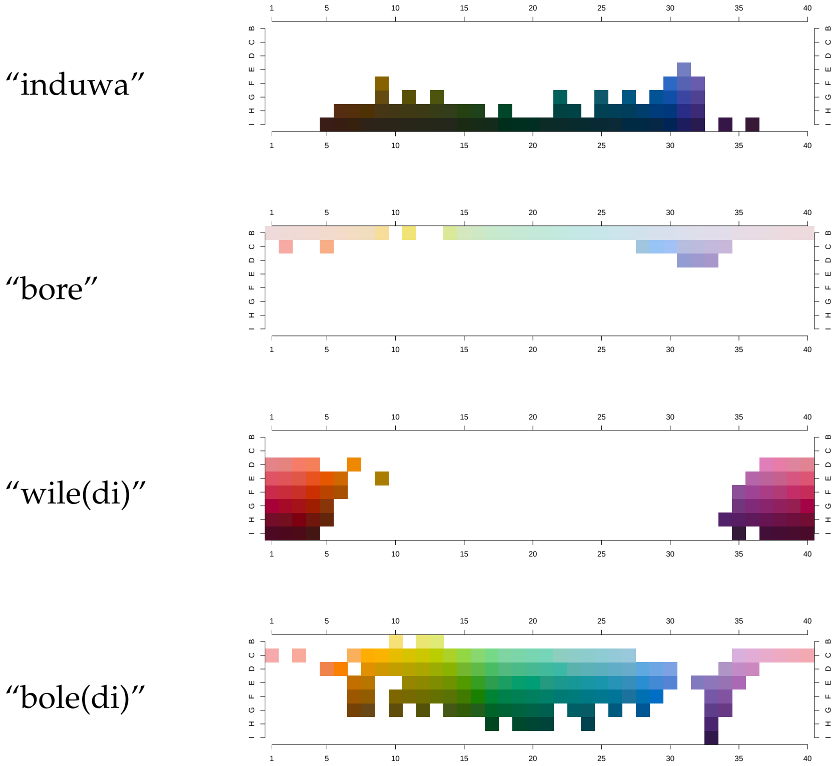


Figure 6.8: Munsell chips organised by their most popular term in Nafaanra

As can be seen from these displays, Munsell chips in the regions labelled by English as “red”, “green” and “yellow” show significantly greater agreement among speakers of the WCS languages than the distribution over all chips. While a small number of “blue” chips show high naming coherence, these form less of a well defined patch. The representation of English “purple” used by Kay and Regier (2003) falls on a chip for which naming coherence is significantly lower than the global distribution.

6.1.3 *Same/different colour*

Fig. 6.7 shows that some Munsell chips are coherently named across the languages of the WCS. The measure of coherence ignores whether different chips are labelled with the same or different terms. The different patches of coherent naming do not necessarily correspond to different colour terms. Different languages in the WCS show different naming patterns over these patches. For example, in Nafaanra (Ghana), the patch of yellow chips in Fig. 6.7 are labelled with the same term as the patch of red chips, while in Lele (Chad) the yellow chips fall under the same term as the generally coherently named green chips (see Figs. 6.8 and 6.9).



"induwa"

"bore"

"wile(di)"

"bole(di)"

Figure 6.9: Munsell chips organised by their most popular term in Lele

To investigate how WCS languages generally treated Munsell chips as the same or different, I produced a "similarity space" in which distances between pairs of chips reflect the degree to which WCS languages use different terms for those chips. For every pair of Munsell chips (i and j), I calculated within each language (l) a distance measure ($d_{l,i,j}$) as the proportion of speakers who used different terms for both chips.[10] These values were then averaged over languages to give a mean dissimilarity measure for each pair of chips ($d_{i,j}$). I then used Torgerson's (1952) multidimensional scaling method to project the Munsell chips into a 329 dimensional space such that Euclidean distances between chips in this space were equal to the mean dissimilarity measures.[11] Torgerson's method conveniently delivers the principle axes of this space: the first axis (or dimension) lies closest to all chips and accounts for as much of the scatter of chips around their centre as a single axis can; the second dimension is orthogonal to the first and

[10] These values obey the triangle inequality $d_{l,i,j} \leq d_{l,i,k} + d_{l,k,j}$ and so are appropriate for mathematical projection into a "space".

[11] N.B. while the analysis presented here describes a "space" this is not intended to be interpreted as any kind of psychologically meaningful colour space.

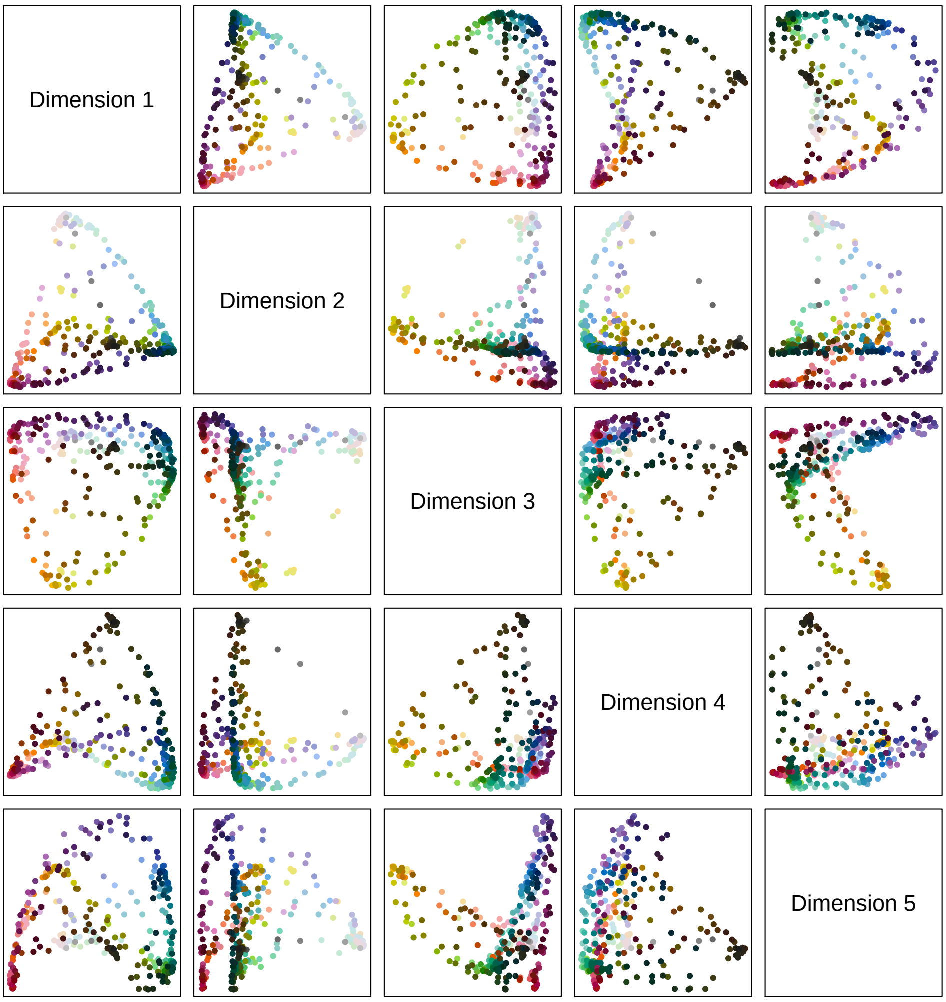


Figure 6.10: Similarity space derived from speaker judgements averaged across all languages. Vertical axis for each plot in a row represents the dimension (principle axis) indicated by text in that row. Horizontal axes represent the dimension indicated by text in each column. Munsell chips are identified by (approximate) colour.

accounts for most of the chips' scatter not accounted for by the first; the third is orthogonal to the first two and again accounts for most scatter not already accounted for etc. (Borg and Groenen 1997). Thus successive dimensions provide an increasingly good approximation to the space in which Euclidean distance corresponds to chip dissimilarity.

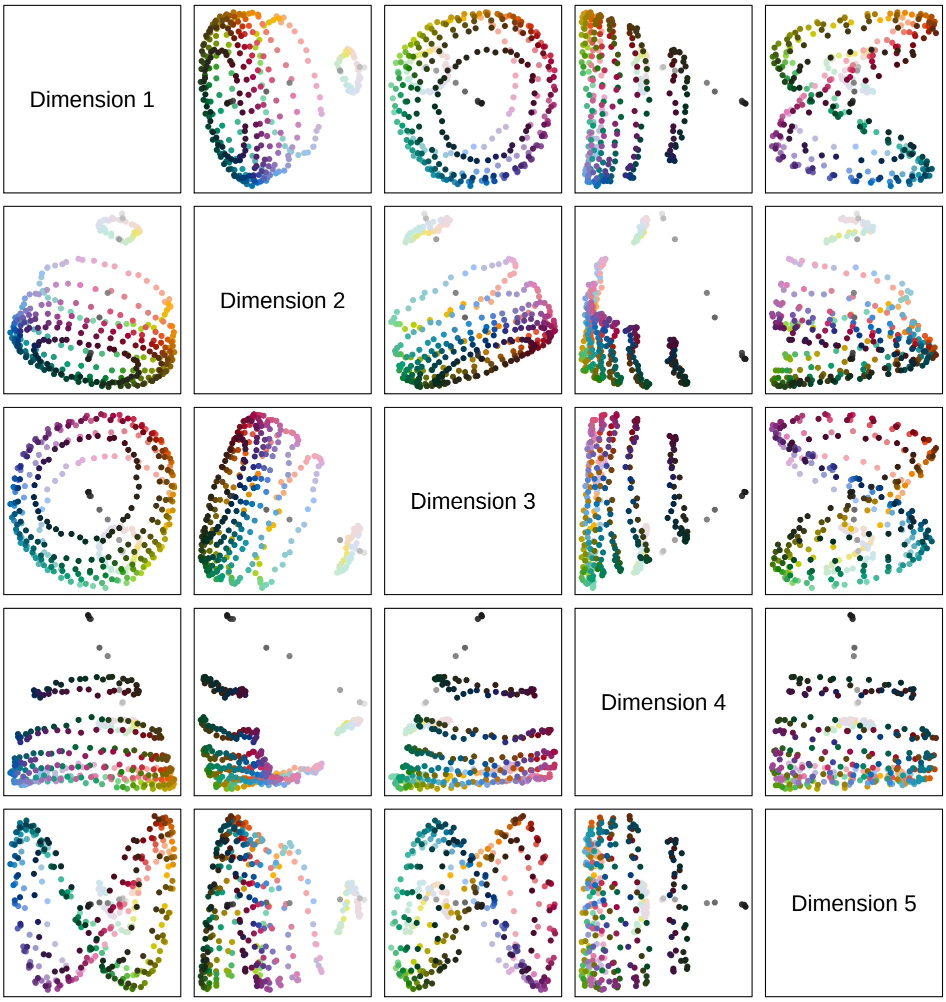


Figure 6.11: Similarity space derived from speaker judgements randomly shifted across the Munsell array on a per-language basis.

Fig. 6.10 shows the Munsell chips' positions in the first five dimensions of the similarity space. At the extremes of the first dimension are the red and green/blue chips, indicating that the strongest signal in the WCS data is a difference between these colours. The second dimension appears to correspond to lightness. Adding the third separates off yellow chips, the fourth darker chips and the fifth blue (as can be seen particularly in the plots of the third dimension against the fifth). In the first three dimensions, the chips roughly form a tetrahedron with red, green, yellow and pale chips forming the vertices.

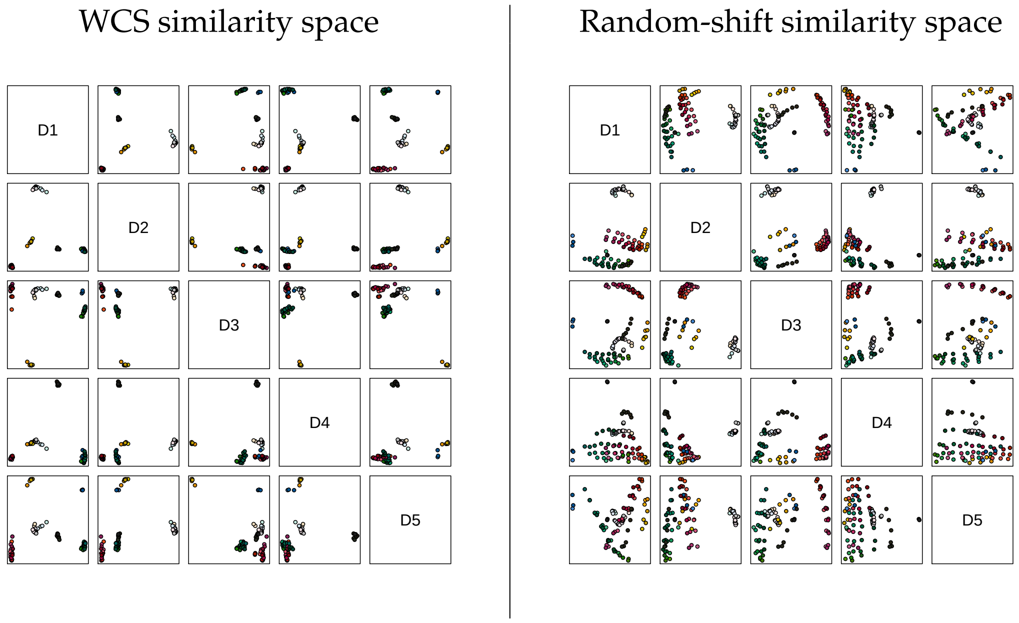


Figure 6.12: Chips with significantly greater coherence measures (see Fig. 6.7) plotted in similarity spaces.

These are rough impressionistic observations about the similarity space; more quantitative analyses follow below.

While Fig. 6.10 shows Munsell chips' positions are related to their colour (i.e. colours do not appear to be randomly distributed through the space) care should be taken not to over-interpret this structure. Fig. 6.11 shows the result of performing the same procedure on data randomly shifted (as above) on a language-by-language basis. Again, structure is apparent: this time the chips form a series of concentric rings ordered by lightness in the first three dimensions. This is to be expected as the procedure for producing randomly shifted data (a) preserves the relationship between naming and lightness, and (b) makes the probability that two chips from the same row appear in the same category in a given language depend only on the number of chips between them on the Munsell array (wrapped round so column 1 is next to column 40). Thus the mean dissimilarity over shifted languages between two chips in the same Munsell row will also depend simply on the distance across the Munsell array between the two chips.

Chips appear more clustered in the WCS similarity space than in the random-shift similarity space. The more chips are clustered, the lower the dissimilarity between each chip and its nearest (most similar) neighbour will be. A one tailed paired samples t-test on nearest-neighbour-dissimilarity measures for the two spaces, revealed significantly more clustering in the WCS similarity space than in the random-shift space ($t(329) = 7.780, p < 10^{-13}$).

Fig. 6.12 shows the positions of only those chips found to be consistently coherently named across languages. In the WCS space the coherently named chips clearly form clusters conforming to the apparent patches on the Munsell array (Fig. 6.7). In contrast, these chips are spread out through the random-shift space.

6.2 Accounting for colour term structure

The idea that across the planet different groups end up with colour term systems bearing similarities to each other has been taken as indicating that colour term systems reflect (are externalisations of) innate internal categories or similar hypothetical internal structures. Kay and McDaniel (1978) suggested that the perceptual sensation of the primary colours red, yellow, green and blue correspond to the firing of four classes of "opponent cell" in the lateral geniculate nucleus (discovered in Macaques and assumed to exist in humans). The firing of these opponent cells depends on the relative stimulation of the three cone classes on the retina. This colour term hypothesis has been discredited both in terms of its own predictions (e.g. the red and green opponent cells referred to by Kay and McDaniel show neutral firing rates at a greenish-yellow spectral hue rather than one described as "neither red nor green", Dowman 2007) and the erroneous assumption that activity in the cone cells is transformed only into a limited range of opponent channels (Webster and Mollon 1994).

Less physiologically grounded proposals for an externalisation–based account of colour terms commonly appeal to colour spaces, focal colours and perceptual distances (e.g. Heider 1971; Bornstein, Kessen, and Weiskopf 1976; Kay and Maffi 1999; Franklin and Davies 2004; Belpaeme and Bleys 2005b; Dowman 2007). A tacit assumption shared by these approaches is that there is a way of representing colour that is somehow "correct", as it is within this representational format that structure is discerned. Consider Bornstein et al.'s (1976) experiments which used a habituation/dishabituation paradigm to probe infants' perception of spectral hues. Stimuli were sets of three spectral hues, evenly spaced in terms of wavelength. The central (habituating) stimulus and one test stimulus were from the same English colour term category (e.g. "blue") while the third was from a different category (e.g. "green"). For most (but not all) sets of stimuli infant looking to the different-category test was longer than to the same-category test, which was the criterion for saying that infants perceive the former test as "different to" and the latter as "the same as" the habituating stimulus. Bornstein et al. considered this an important result as the wavelength differences between the habituating light and the two tests were the same. Presumably, had infants looked for equal time to the two test stimuli, and only looked longer to stimuli of a greater wavelength separation from the habituating stimulus, Bornstein et al. would not have concluded that infant perception matched linguistic categories. However, the choice of wavelength as the measure with which to characterise the stimuli was entirely arbitrary. Had the same stimuli been characterised in terms of their frequency (which is proportional to 1/wavelength) physical differences between stimuli would not have been equal. This is not to say that frequency characterisation would be better (or would show the dishabituation patterns to rely on uneven frequency differences) but to point out that different, arbitrarily chosen ways of characterising stimuli will give a different baseline against which to present the results of the dishabituation experiments. Indeed, one way of characterising the difference between spectral lights is in terms of their effect on infants in dishabituation experiments (i.e. differences could be expressed as the length of time for which an infant looks at one light having been habituated to the other).[12] This characterisation would (by definition) show that "equally spaced" stimuli are seen as "equally different" independently of their relationship to linguistic categories. Similarly, it would be possible to characterise differences between stimuli in such a way that within linguistic categories infants perceive "equal" differences in spectral lights as greater than differences across categories.

Aware that characterising stimuli in terms of wavelength (and restricting stimuli to spectral lights) is problematic, Franklin and Davies (2004) repeated Bornstein et al.'s experiment, but used Munsell chips rather than spectral lights, and measured colour difference in Munsell hue units. Franklin and Davies describe the Munsell units as forming a "perceptually uniform metric". However, this notion rides on the indeterminacy of what a perceptual colour distance is. Suppose for the sake of argument that adjacent Munsell chips are separated by equal perceptual distances in the sense that for any pair of adjacent chips, human adults will on average rate[13] the similarity of the colours as being the same as the similarity of any other pair of adjacent chips. We may now say that any pair of chips

[12] N.B. the CIE XYZ and CIE L*a*b* colour spaces represent attempts to produce colour spaces that reflect the results of psychophysical experiments.

[13] This may depend on *how* subjects are asked to rate similarity.

separated by one chip are also separated by equal perceptual distances if by that we mean that the number of intervening adjacent Munsell chips is the same (and these adjacent Munsell chips all receive the same pairwise similarity judgement). However, if by “perceptual distance” between non-adjacent chips we mean average similarity judgements for those chips, it is a *hypothesis* that any given pair of Munsell chips on a hue circle separated by a single chip is separated by the same perceptual distance as all other such pairs. Franklin and Davies (2004, p. 360) actually tested this hypothesis and found that it did not hold, though they did not take this as disconfirming their assumption that the Munsell system gives a “perceptually uniform metric” for non-adjacent chips. That such a hypothesis does not hold is only surprising on the erroneous assumption that the Munsell system should map in a particularly simple way onto the colour-relevant mechanics of human biology. On this assumption what seems reasonable in Munsell space should be reflected in human behaviour. But as stressed in section 4.4.3 such neat mapping of structures would be an incredible coincidence. Without the assumption that the Munsell representation is privileged by being a straightforward representation of colour vision mechanisms, adult similarity judgements (and hence Munsell space) are just as arbitrary a measurement system as wavelength for Franklin and Davies’s experiment. Again, an alternative distance metric could be constructed on the basis of infant behaviour for which it would be definitional that infants’ colour discrimination is independent of adult colour categories. By relying on arbitrary systems of representation in order to discern “structure” (that would disappear in different representational formats), infant discrimination experiments are impotent to explain the origins of languages’ colour term systems.[14]

6.2.1 *Agent based models*

Two different simulations have been offered to account for languages’ colour term systems[15]: Belpaeme and Bleys’s (2005b) model relies on the CIE L*a*b*

[14] Bornstein et al. (1976) and Franklin and Davies (2004) were both attempting to demonstrate infant “categorical perception”. However, if the notion of categorical perception is not to be beholden to an arbitrary choice of measurement system it needs to be demonstrated that individuals’ discrimination of stimuli from a continuum can be predicted solely from their categorisation of those stimuli (Studdert-Kennedy et al. 1970). In this sense, adult colour vision is not categorical as two colours may be identified as “red” with 100% consistency *and* be 100% discriminable. (In addition whether infant colour vision is categorical is an indeterminate question as it is indeterminate what it means to say that a pre-linguistic human categorises colours.)

[15] Steels and Belpaeme (2005) present a very similar model to Belpaeme and Bleys’s, though they couch it in terms of creating communicating robots rather than an account of natural languages’ colour terms.

space and Dowman's (2007) uses a bespoke representation of colour whose parameters were tuned "by a process of trial and error" (p. 125) to produce results that approximated certain features of the WCS data. These models are explicit about the assumed relationship between language and colour space, positing that colour terms are learned by association with patches of colour space. In these models, agents use their internal associations to produce a signal to label an object of a given colour. Agents learn from each others' communications and arrive at a stable shared system of colour communication. The geometries of the objects' colours' representations in the colour space determine the resulting relationship between language and colour.

The use of colour spaces in these models represents unfounded assumptions about how humans learn and produce colour terms. Interpreted simply as mathematical constructs from which to predict human colour-language behaviour they represent simplistic and untested hypotheses. In Dowman's (2007) case, the hypothesis is that colour term learning can be described by a rather arbitrary model whose *design* is based on the WCS data. In Belpaeme and Bleys's (2005b) case, the hypothesis is that the use of colour terms can be predicted on the basis of a mathematical object designed to roughly capture results of colour discrimination experiments.[16] The power of these models to explain structures in WCS data is thus limited by these unfounded assumptions.[17] It is worth noting that studies of children's colour term learning often find that colour terms are learned relatively slowly and are frequently misused by children for the wrong colour (Soja 1994). Children learn that "red" is an appropriate response to a question of the form "what colour is it?" before they master the relationship between this response and red objects (Sandhofer and Smith 2001). Thus there is little reason to suppose that learning colour terms is a simple matter: children gradually learn to exploit the colour system offered to them by other people's knowledge.

There are many different ways of constructing colour spaces which could be inserted into agent-based models to produce different results. Interestingly, several measures of colour discriminability (which could be used as the basis of

[16] The results of Belpaeme and Bleys (2005b) simulation seem simply to be that agents' categories form at the corners of an RGB cube translated into CIE L*a*b* space (a result that is a rather transparent consequence of the way agents identify regions of colour space). As such, Belpaeme and Bleys' results depend on the choice of coloured light in computer screen technology (which cannot present every visible colour).

[17] It is worth noting, Dowman's (2007) model was tweaked until it fit the WCS data in a number of respects, so the fact that it does fit the WCS data in those respects is unsurprising. Belpaeme and Bleys (2005b) produce results which have only superficial similarity to the WCS data (and identifies purple as a region in which languages develop consistent colour terms).

a colour space) appear to be malleable, depending to a degree on experience. Özgen (2004) reports that discriminability of colours within a colour category can be improved by training, and that speakers of certain African languages which have no green/blue distinction are outperformed by English speakers on a visual search task when distractors and target straddle the green/blue distinction (when target and distractors are from the same English category, Africans were faster than English speakers). Similarly, Roberson et al. (2000) found that various measures of perceived colour difference were dependent on language, results showing greater distinction between colour pairs for speakers of a language in which the colours are given different labels than for speakers of a language in which the colours are given the same label.

Dowman (2007, p. 126) claims that the results of his model depend crucially on cultural transmission. This is true in the sense that in the model an agent's colour categories could only be interpreted as its linguistic categories, and these were acquired through cultural transmission. However, while it may not be obvious what colour systems will emerge from simplistic inspection of the modelled agent, the emergent systems depend only on the properties of individuals (in the sense that nothing other than the learning model *can* be changed to affect the outcome). In Belpaeme and Bleys' model, the distribution of colours in the environment is a variable on which the resulting colour categories depend, but when the distribution of coloured pixels in digital photographs of natural scenes was used, the success of the model in producing WCS-like categories was diminished.[18] Colour categories in Belpaeme and Bleys' model can be identified in two ways, as linguistic categories or as non-linguistic projections onto the CIE L*a*b* colour space. Simulations in which agents did not communicate with each other produced similar colour categories to simulations of communication (while similarity to the WCS data is quantified, no statistics are available to determine whether the difference between the simulations is significant). Thus in both simulations the resulting language is largely determined by the make-up of individuals.

A more fundamental way in which both of these models rely on the properties of individuals is in the assumption that the language games played by agents are solely related to colour. In this sense, colour is built directly into these simulations (van Brakel 2005).

[18] This adds support to the view that the success of Belpaeme and Bleys's model is due to the model selecting the (arbitrary) corners of the RGB cube projected into CIE L*a*b* space, (c.f. fn 16).

6.3 Communicative affordances of colour

6.3.1 Colour grammar

> When we're asked "What do the words 'red', 'blue', 'black', 'white' mean?" we can, of course, immediately point to things which have these colours, — but our ability to explain the meanings of these words goes no further! For the rest, we have either no idea at all of their use, or a very rough and to some extent false one. (*RoC* I/§68)

Throughout this chapter I have been using the expression "colour term" quite loosely to refer to expressions in various languages whose use resembles the use of English colour terms. In order to develop a communicative affordances account of the evolution of colour terms, it is necessary to say something about the grammar of those terms, and the respects in which they may be thought of as the same kind of term across languages.

What kind of thing is colour? This question suggests an Augustinian response, the search for something for which a colour term can be the name. Such a response (e.g. colours are patterns of neural firing, Zeki 1983, or classes of reflectance spectra Byrne and Hilbert 2003, or simply that colour terms are used to refer to colours) is generally unhelpful when asking whether an expression in a foreign language corresponds to a colour term until we determine a criterion for stating that the foreign word is a name for such a thing. Rather than try to assimilate our use of colour terms to other uses (e.g. terms labelling chemical substances, as in "water is H_2O") we should characterise what role a term has to play in the lives of its users for it to count as a colour term. Not only will this clarify what I have been loosely describing as "colour terms", but it will also (I suggest) offer the basis of the solution to the problem of commonalities found in the WCS data. What follows is a brief and incomplete sketch of the use of English colour terms, followed by the respects in which it may be inferred that the WCS colour terms have similar uses.

A principal use of English colour terms is to identify objects. I can, for example, ask someone to pass me the blue (or the brown) book and there are cases in which this will prompt that person to pass me a particular book. In such cases, the difference between the correct book and an incorrect choice is determined by sight (as opposed to, say the correct response to the request "pass me the heavy book"). For some colours (e.g. "Munsell colour 10.00R 16/5") whether an object is that colour may be determined by various technical measures, but for the most commonly used English colour terms (“red”, “white”, etc.) whether an object is that colour is determined by human judgement. That is, someone who knows what “red” means (and has normal vision) is generally able to tell whether something is red simply by looking at it. This is not to say that whether an object is “red” or not is a subjective matter (different for different people, as is whether a given object is “pretty”). People, who have normal vision and understand colour terms, generally (though not always) agree whether the colour of an object is this or that. This agreement without regulation by a measurement procedure is an important aspect of colour terms’ grammar, and the circumstances in which it can be achieved will be used below to characterise the emergence of colour terms.

A colour term can be explained to someone familiar with the practices associated with colour naming by pointing to a sample. Using samples to determine whether something is a given colour (e.g. “chartreuse”) is a particular technique or set of techniques: e.g. the sample may be held against the object to see if they are the same *in respect of their colour*. This technique is difficult to characterise as teaching it may involve demonstration, encouragement and correction, but not a formulation of principles (telling someone who doesn’t know how to compare colours that the technique relies on two things being the same colour informs that person that what they are learning to do is make judgements of “the same colour”; such information cannot tell them what counts as being the same colour). While visual indistinguishability of patches of coloured objects is important to sameness of colour, sameness of colour does not mean indistinguishable to sight: objects may have the same colour, shape and size yet still be generally visually distinguishable (e.g. matte vs glossy). This is a respect in which WCS terms may differ from English “red”, “blue”, etc. as it *may* be the case that objects which we would say were the same colour but differed in their visual appearance in some non-colour respect would be appropriately labelled with different terms in a foreign language, even though those terms are the closest approximation to colour terms. For example, Saunders and van Brakel (2002) state that in Karam (WCS language number 51), leaves of plants, herbs and trees with the same Munsell colour code can be identified as *mosb*, *waln* and *lban*.

Not only objects, but also light may be “coloured”. One way of determining what the colour of a given light is is to see what a white object looks like when it is only illuminated by that light; that is, under red light white objects appear

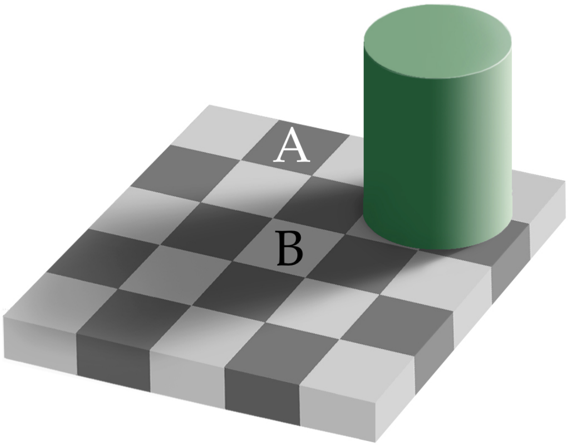


Figure 6.13: Grey square optical illusion. Squares A and B (excluding labels) are "the same colour" in terms of the light leaving them and entering the eye. (From Wikimedia Commons.)

to be the same colour as red objects in ordinary light. This "appearing" to be a given colour generally does not rely on techniques additional to telling whether two objects are the same colour in ordinary lighting conditions: people can be fooled by illumination into thinking (e.g.) a white object is red. Objects are seen by their reflecting light into the eye, though the colour of an object is not necessarily identical to the colour of the light it reflects. White objects illuminated by coloured light are a case in point as we do not say that the white object becomes a red object when illuminated with (and reflecting) red light but that it is a white object that looks red. Similarly, an object may have unvarying colour across its surface yet the colour of the light leaving different regions of the surface and entering the eye might vary. Fig. 6.13 illustrates this difference: squares A and B are different colours in the sense that Fig. 6.13 is a picture of such a difference, yet the light leaving the two squares and entering the eye is the same colour (because in the picture the two squares are produced either with the same ink or with the same coloured pixels). In this respect, "same colour" is somewhat indeterminate, sometimes being used in relation to "same coloured light" and at others in relation to "same coloured surface". This too may be a difference between the grammar of English colour concepts and those of the WCS languages: if a community does not manipulate lighting, and does not produce naturalistic representations (in which the colour of paint may be distinguished from the colour of the object of the painting) equivalently distinct ways of using an expression translated as "same colour" may not exist.

Various additional aspects of the grammar of English colour terms may or may not be shared with the WCS languages: that "orange" is "both yellow and red"[19] but that no colour is "both red and green" in the same sense (c.f. section 4.2.4.2); or that transparent objects may be "red", "green", "yellow", etc. but not "white" or "black" (c.f. *RoC* I/§§19–20)[20]. These grammatical relationships may serve as clues to the history and development of colour terms in English. However the methodology of the WCS makes it impossible to infer from that data-set whether different languages display these or similar grammatical relationships.

6.3.1.1 *WCS stipulation of colour terms*

Field workers collecting data for the WCS were instructed to elicit "basic" colour terms, and were told that "Essentially, the basic color terms for a given speaker are the smallest set of simple words with which the speaker can name any color."[21] The idea that these terms "name colour" was built into the methodology for collecting data as the only differences between Munsell chips were differences in colour. While the grammar of the terms collected in this fashion may vary considerably, we may infer that these terms share with English colour terms the fact that they can be used in relation to a variety of objects — including, of course, Munsell chips (at least, in cases where there is general agreement among speakers), and that they are the terms speakers think of when confronted with a series of chips varying only in colour. It seems to be a safe assumption on these grounds that these terms can be used in certain circumstances to indicate subsets of objects which we (English speakers) would pick out by their "colour".

The instructions given to field workers can be used to make additional inferences about the uses of the WCS terms. While "basic colour term" was not given a unique operational definition, field workers were told that elicited terms should ideally exhibit the following four characteristics:

1. Monolexemic (i.e. meaning not predictable from the meaning of the expression's parts).
2. Signification not included in that of any other colour term (so e.g. "scarlet", being a kind of "red" would be excluded).

[19] Perhaps this way of speaking about orange has its roots in mixing paint, as mixing yellow and red paint does produce orange.

[20] A term may, for example, be used to refer to white opaque things and to what we would call colourless transparent things. This *may* be taken as a ground for denying that the term is rightly called a colour term.

[21] Instructions to field workers can be found at http://www.icsi.berkeley.edu/wcs/data.html

3. Applicable to more than just a narrow range of objects.
4. Psychologically salient (indices of which include occurrence at the beginning of elicited lists of "colour terms", stability of reference across informants and occasions of use, and occurrence in the idiolects of all informants.)

The applicability of these characteristics is a matter of individual judgement[22] (and for individual languages questions have been raised as to whether these characteristics apply to the terms collected by the WCS, e.g. Everett 2005). Compounding this difficulty are inconsistencies in the field workers' instructions (e.g. they were told "Some informants may lack entirely terms which are basic for others" which contradicts one of the indices of characteristic 4).

Thus while the grouping together of terms in the WCS as "colour terms" may mask significant differences in the grammars of these terms (including aspects that may be called on to question whether the terms are appropriately described as "colour terms") I will assume that they share with English colour terms the conventional role of identifying objects or groups of objects in ways that can be assessed by visual inspection alone; that the same terms can be used in this role for a variety of different objects; that objects that are the same colour as each other often (though not in every case) fall under the same term. I also assume that, rather than existing in a communicational vacuum, these "colour terms" are used because they allow speakers to achieve certain desirable effects in their environments, not simply because they are a reflex of an instinct to "name colours".

6.3.2 Development of "colour terms"

How could a language develop terms that operate like colour terms? Here is one fictitious and idealised possibility: imagine again a simple language game in which A instructs B to fetch a particular object. For the purposes of this example, the objects which A instructs B to fetch are differently coloured: suppose B is to fetch A a certain kind of fruit which changes colour with ripeness (being green when unripe and red when ripe), and A indicates to B whether he is to bring a ripe or an unripe fruit. The colours of the fruits A requests may make B's job relatively easy: if he knows where to find the appropriate fruit he doesn't need to inspect them much beyond a quick glance. If B collects several fruits at a time, he

[22] Four additional criteria were listed for "doubtful" cases, and these too are open to interpretation.

may on occasion choose something wrong that, to a quick glance, resembles the fruit he is to fetch (i.e., as we would describe it, is a similar colour). What effect this might have on the use of the term will depend on A's reaction: for example, if A requests a ripe fruit, but what B brings is a different kind of fruit which happens to be green when ripe and red when unripe, A will respond negatively and B will learn to be more careful in future. However, if when B makes a mistake he brings something that pleases A, B may learn that picking objects on the basis of a quick glance is OK. This will depend in part on what A wants the fruit for, and in part on what kinds of thing B is likely to accidentally pick up on the basis of visual similarity.

Others may also learn to engage with B in order to get him to fetch objects. If C requires different kinds of object which can be differentiated on the basis of the same colour distinction as the ripe/unripe fruit A requests, C may be able to exploit B's behaviour, thereby extending the range of objects for which the expressions are used.

This simplistic scenario is not intended as a realistic history for any actual colour term, but illustrates the basis on which I propose to account for the common structure in the WCS. In the scenario above, whether the original expression used by A and B would develop a more abstract relation to colour (i.e. become an expression used for various different kinds of object of the same colour) depended on the attitude of individuals to the objects fetched on the basis of visual similarity to the original sets of objects. If the colours (and the colour contrast) in the original language game failed to correspond to a meaningful contrast for other objects (e.g. if C's interest in the new objects B brought was independent of their colour differences as defined by the original game) then there are two reasons for supposing that the practice wouldn't develop into something like a colour term. One possibility is that B may learn two techniques for collecting objects on the basis of what kind of object they are (or who asked for them), the two techniques being based on different colour contrasts would mean that analysis of the expression as a colour term would be inappropriate. Alternatively, C may find his interests better served by using some other conventional expression and so not adopt the same expression as that used by A and B (we need not assume that C spontaneously invents such an expression; it may develop in a slow and complicated manner, but once it becomes available the affordances of this expression would offer C a better way or dealing with objects than the original A-B interaction).

Human interactions with each other and coloured objects are far more complex than the simple practices of object-fetching in the imaginary scenario above. However, the scope for an interactive practice tied to particular objects (for which a colour contrast is important) to be extended to other objects *on the basis of colour* will depend on the degree to which meaningful colours and colour contrasts match across different kinds of object.

The suggestion here is that (relatively) object-general colour terms develop from other language games, and that this route means that colour terms are structured by the meaningful colour contrasts in human environments. However, this proposal relies on the existence of some other communicative practice whose origins I haven't described. It might be objected that in the fictitious scenario above the language game of fetching objects was simply assumed, and if that is valid then why not simply assume the language game of naming colours? However, the proposal here does not rely on any specific communicative practice, merely on the possibility that a practice concerns specific objects in such a way that correlates with the colour contrast of those objects. In addition, a colour based system which relies on contrasts that are relevant across various objects, will show the kind of stability characteristic of colour terms (i.e. that individuals with different experience can agree in their judgement as to what colour to call certain objects without recourse to some technical analysis): once an individual has learned to deal with colours in a limited set of the possible domains of colour term use, they will be able to engage with others (who have learned similar practices for other kinds of object) on the basis of colour.

The communicative–affordances based proposal presented here, then, is that similarities in the WCS data are reflections of similarities in the object colours and colour contrasts that are generally relevant to human life in the various communities studied. The next section asks why this would be so.

6.4 Evolution of colour

The WCS took over 20 years to collect data for 110 languages, so collecting data in the field about the role of colour in the lives of the WCS languages' speakers would be far beyond the scope of this thesis. However, even if such field work found a close correlation between colour and colour terms, the direction of causality would remain an open question: colour terms may derive from the role of colour in human life, or colour terms may shape the colour of the human-relevant environment if individuals use these terms in some of their activities which result in organisation of the environment. Saunders and van Brakel (2002) argue that the colour environments of Western societies are largely constructed on the basis of extant colour-based practices (e.g. a manufacturer may decide that his products should be sold in "red" and "green" boxes but not "yellow" ones). Therefore, this section lays out some reasons for thinking that across human cultures, in different environments, correlated relevant colours and colour contrasts which form the scaffolding on which colour term systems can be based will be similar prior to the emergence of colour terms. In outline, the reasoning takes the following steps:

- The evolution of trichromatic vision among primates was a driven by the selective advantage of being able to make certain colour discriminations, indicating that these colour discriminations were significant to primate life.
- The presence of trichromatic primates in an environment changes the selection pressure on the colours of organisms in that environment. This leads to some structure in the colour environment.
- This increasingly structured colour environment heightened the selection pressure for selective responses to colours which often functioned as a human-relevant signal. These responses became "hardwired" to a degree, but may also have a learned component.
- This low-level differential response is relatively small on an individual basis, but can cumulatively build up into more distinct environmental colour structure. This, in addition to the colour response to the selection pressures imposed by the presence of humans, led to various colour distinctions frequently relevant to human life (in various ways) being correlated across cultures.

6.4.1 *The evolution of colour vision*

Fig. 6.14 shows the structure of the human eye. Light passing through the cornea and lens is focused as an image on the retina. Photoreceptor cells on the retina change their polarisation (and hence firing rate) in reaction to the amount of light falling on them. The processes by which light is absorbed and converted into neural impulses are rather complex, but depend on the photopigments (light absorbing chemicals) in the photoreceptor cells. The human eye contains four kinds of photoreceptor cell each of with contains a single kind of photopigment.

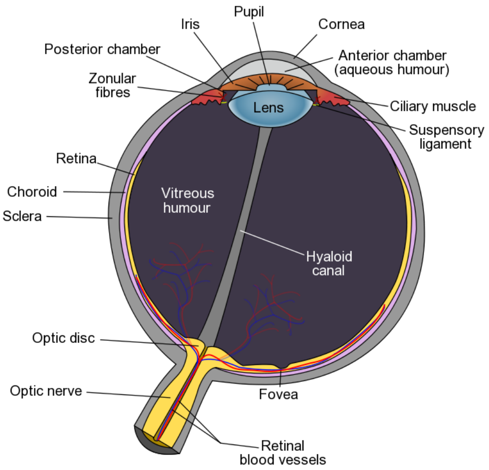


Figure 6.14: Schematic diagram of the human eye (from Wikimedia commons).

The four photopigments differ in their absorption spectra, so the photoreceptor cells differ in their response to light of different wavelength composition. Three classes of photoreceptor cell, the cones, mediate photopic vision (that is, vision at relatively bright levels of illumination such as daylight) while the fourth, the rods, are important for scotopic vision (vision under low illumination levels, Nathans 1999).

The three cones may be classified according to the wavelengths of light they are most responsive to as long, medium and short wavelength sensitive cones (L, M and S cones respectively).[23] Fig. 6.15 shows the sensitivities of the three cone classes in terms of the proportion of incident energy that is absorbed by the cone, resulting in changes to the cone's firing rate (these curves are normalised to peak at the same level). Light reflected from objects with different reflectance spectra (under the same illumination) will generally stimulate the three cone classes to different degrees, and these differences are the basis of colour vision. The responsivity curves of the three cone types overlap, so any light entering the eye (even light perceived as a "pure" colour) generally stimulates more than one type of cone (thus it is a mistake to identify the three cone types as individually responsive to red, green and blue).

[23] It is something of a simplification to suggest that there are just two L/M cone types as there are different versions of the L cone photopigment, roughly equally distributed in the human population with slightly different absorption spectra (Nathans 1999).

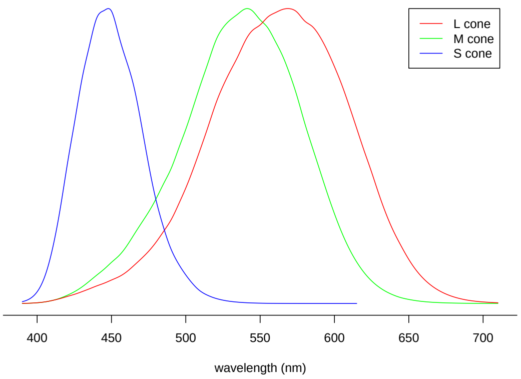


Figure 6.15: Normalised cone energy responsivities (Stockman, Sharpe, and Fach 1999; Stockman and Sharpe 2000).

6.4.2 *Visual ecology and the evolution of colour vision*

Fig. 6.15 shows the cone sensitivities for humans. Different animal species differ in both number and spectral responsiveness of photoreceptors. For example, honey bees have three photopigments, one of which has peak responsivity in the ultraviolet range; birds generally have four cone classes (including an ultraviolet sensitive cone) and most mammals have just two (Goldsmith 1990). The cone photopigments of jawed vertebrates can be arranged into four classes on the basis of molecular analysis (small differences within each class fine tune spectral sensitivities). The same four classes of photopigment have been found in a jawless vertebrate, indicating their presence in vertebrate's eyes before the divergence of jawed and jawless forms more than 540 million years ago (Vorobyev 2004).

Paleontological evidence suggests that, prior to the extinction of the dinosaurs, mammals were nocturnal. As such they would have relied more on rod mediated (low light) vision than cones, and this may be part of the reason the mammals lost two of the four molecular classes of cone (Goldsmith 1990). The three human cone photopigments evolved from the mammalian pair, the M and L cones being differently tuned versions of photopigments from the same (molecularly specified) class. Why the mammals retained the photoreceptors they did is a matter of some debate (nocturnal mammal retinas are dominated by rods, and some mammals have retained only one type of cone, Peichl et al. 2000). One possibility is that the two retained photopic photopigments were *not* used to discriminate objects on the basis of colour (for which there would be little opportunity in low-light environments), but to control circadian rhythms on the basis of the varying composition of ambient light as the sun moves across the sky (Peichl et al. 2000).

The development of trichromacy (having three spectrally distinct photopic photopigments) in primates indicates a selection pressure for better colour vision. This is made more apparent by the *detrimental* effects of increasing the number of cone classes. Patterns of cone activation in the retina signal not only differences in wavelength composition but also differences in illumination due to the spatial detail of the observed scene. As differences in firing rates between adjacent cones could be due to either chromatic or luminance differences, the presence of differently tuned photopigments reduces the spatial acuity of vision (Nathans 1999). This negative impact on vision may be part of the explanation for why primate photoreceptor sensitivities do not span the visible spectrum more evenly, and why platyrrhine colour vision is polymorphic rather than universal (Regan et al. 2001).

What visual advantage does trichromacy confer on primates over their dichromatic (two photopic photopigment) conspecifics? Chlorophyll is a ubiquitous aspect of primate environments, largely determining the reflectance properties of foliage, and one possible advantage of primate trichromacy may have been the ability to discriminate between foliage and other objects. Fig. 6.16 shows responsivities of human cones and the radiant spectra of foliage when viewed from above or below on a sunny day. The peak of these spectra fall between the M and L cone responsivities, activating them to roughly the same degree. Objects presenting different spectra to the human eye would (in most cases) produce unequal activation of the M and L cones which would serve as the basis for discrimination from foliage. Sumner and Mollon (2000a) measured the reflectance spectra of a large number of leaves in the environments of certain catrrhine species (whose cone responsivity curve positions are similar to our own), and computed the activations of these animals' M and L cones when viewing these leaves under typical daytime illumination. While cone activation levels varied for different leaves, the ratio of M to L cone activation remained roughly the same. Sumner and Mollon (2000a) compared these results with hypothetical cone activation levels that would be found if the M and L cones' peak spectral sensitivities were at wavelengths significantly different to those of catarrhines. Much greater variation was found in L to M cone activation ratios, indicating

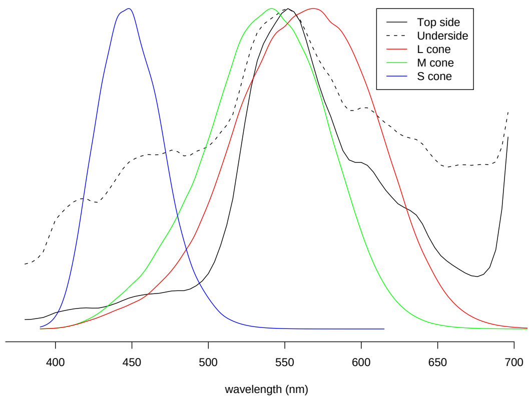


Figure 6.16: Normalised human cone responsivities and normalised mean radiance spectra for foliage (top side and underside) recorded *in situ* in French Guiana under sunny conditions (data from http://vision.psychol.cam.ac.uk/spectra/spectra.html).

that the positions of L and M cones are tuned to provide a reliable low-level cue to whether what is being looked at is (just) a leaf.

6.4.2.1 *Evolution of colour environment*

Thus the advantage originally conferred by primate trichromacy appears to have been the ability to detect non-foliage objects against a foliage background. Whether this advantage was related to a specific kind of object (e.g. ripe fruits, edible yellow or red leaves, conspecifics) or was more general is unknown (Regan et al. 2001). A difficulty in determining which objects would have become more visible against the foliage to trichromats is that the reflective properties of objects appear to have changed in response to primate trichromacy.

Sumner and Mollon (2000b) and Regan et al. (2001) studied the cone responses of trichromatic catarrhines and platyrrhines (respectively) when fruits at various stages of ripeness are viewed under a range of daylight conditions. In both cases, differences in L and M cone activation offered a cue to the ripeness of fruits eaten by these primates. This was more reliable than the differences in cone activation for dichromatic platyrrhines and the inferred dichromatic catarrhine ancestor. Regan et al. (2001) found that fruits which rely on primates for seed dispersal (either in dung or by being spat out some distance from their original location) were more conspicuous against a foliage background to trichromatic individuals than dichromats, and produced similar patterns of cone activation. Fruits whose seeds would be destroyed when eaten by primates showed much greater variability in cone activation patterns. Regan et al. suggest that plants from various botanical families which rely on primates for seed dispersal have convergently evolved surface reflectance that exploit trichromatic primate vision to provide a consistent cue to the location of ripe fruits against a foliage background. Other fruits were found either to be cryptic to trichromatic primate vision, or to display colour signals apparently tuned to the visual systems of other animals (e.g. birds). Ripe, primate-seeking fruit were generally found to produce larger L/M cone ratios that the background foliage. In colour terms, this corresponds roughly to being yellow, orange or red against a background of green.

Using similar techniques, Sumner and Mollon (2003) demonstrated that differences in L and M cone activation could be used to easily discriminate the orange fur of various polymorphic and uniformly trichromatic primates (including the orangutan) from a foliage background. The cone activations of dichromats generally did not differ between foliage and fur. Fur colouration may have evolved in some species, exploiting trichromatic vision, though the advantage of orange colouration among platyrrhine species whose colour vision is polymorphic (and so should lead to behavioural differences between trichromat and dichromat individuals in relation to seeing conspecifics) is unclear (Sumner and Mollon 2003).

However, identifiability against a foliage background is not the only way primate bodies may exploit trichromacy. L and M cone sensitivities are well placed to detect changes in the oxygenation level of blood beneath the skin surface, so exposed skin can serve as a visual cue to various factors (emotional or sexual state, state of health[24]) and species with uniform or polymorphic trichromat vision tend to have bare faces while monochromats and dichromats tend to have furry faces (Changizi et al. 2006). For many catarrhine species, female sexual skin swells and reddens cyclically in coordination with ovulation as a cue to males that copulation may result in conception; males also often have sexual skin which reddens with testosterone levels and may signal dominance or fitness (Dixson 1998).

The presence of trichromatic primates can influence the colours in their environments through natural selection. While the colour signals fruits use to attract

[24] Dichromat humans find judging the state of someone's health from their appearance difficult (Vorobyev 2004).

primates may have begun by simply enhancing the visibility of fruits to trichromats, it seems likely that primate audiences could learn the colour signals, using them as an indication of ripeness irrespective of the background against which fruit is viewed. Tests performed on captive chacma baboons (*Papio ursinus*) have demonstrated that the increased redness of female sexual skin is visually arousing to males indicating that the function of reddened skin is more than simply a visibility enhancement (Dixson 1998).

6.4.2.2 *Attention to red*

There is some evidence that different colours provoke different responses in humans with some degree of consistency. Bornstein (1975) measured the length of time 4–5 month old infants would look at a uniform spectrally illuminated patch, and found looking times varied with wavelength. His data are represented in Fig. 6.17 along with a rough representation of the difference in L and M cone activation.[25] Bornstein interpreted his results in terms of infant possession of colour categories. However, it seems a simpler interpretation is that looking times may be related to the difference in L and M cone activation: a Spearman's rank correlation test on looking times and the rough measure of difference in cone activation showed a significant correlation ($r = 0.970, N = 8, p < 0.001$).[26]

These results suggest that the exploitation of L and M cone differences by aspects of the environment signalling to human vision may have produced innate heightened attention to objects whose reflectance spectra produce this imbalance. This need not be the sole cause of individuals' responses to colour (Dixson 1998

[25] The measure of difference in L and M cone activation is the absolute value of the logarithm of the ratio of cone responsivities. This measure was chosen because (a) Bornstein (1975) used stimuli equated for luminance (and so at different power intensity levels) so taking the ratio of L and M cone responsivities corresponds (roughly) to the ratio of cone activation levels whereas the difference between L and M responsivities would not correspond to the difference between L and M activation; (b) taking the absolute value of the logarithm produces a measure which is neutral with respect to whether L or M cone activation is greater and indicates only the extent to which one cone is more activated than the other. However, the responsivities of the L and M cones, taken from Stockman and Sharpe (2000) are arbitrarily scaled to have a maximum of one. Differences in the appropriate scale of L and M cone responsivities will therefore make the measure used here different to the absolute value of the logarithm of the ratio of cone activation levels. However, the similarities of L and M cones and their photopigments suggest a correction to the measure used here would not be particularly large.

[26] These results should be handled with care and taken as suggestive rather than conclusive: Bornstein's experiment was not designed to test the relationship between looking time and cone activations, and further investigation into the possible relationship introduced here would employ more carefully controlled stimuli specifically targeting regions in the spectrum predicted to show certain effects. Additionally the rough measure of difference in cone activation should be supplemented with a more accurate measure.

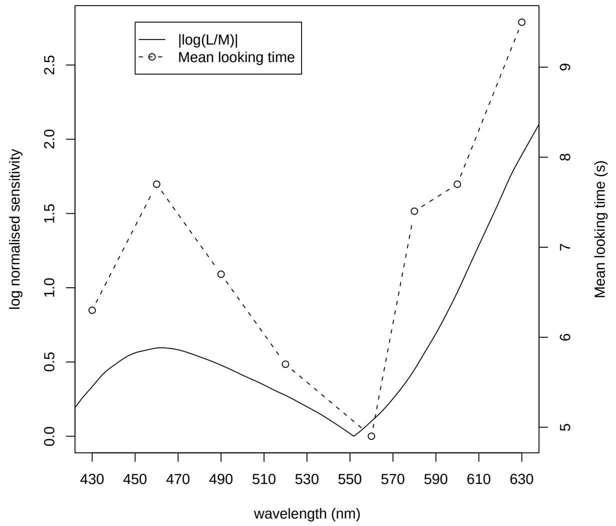


Figure 6.17: Difference between L and M cone sensitivities plotted with mean infant looking times at spectral lights.

suggests that the arousal response of male baboons to female reddened sexual skin may in part be learned through experience of, e.g., female proceptive behaviour), but it could guide an individual's experience of the world, influencing what they learn about colour signals.

A number of studies have investigated physiological responses to colour. While the results are not always conclusive and admit various interpretations, when colour is found to have a distinctive effect on adults it is usually red that stands out: red light is more conducive to producing epileptic seizures than blue light, induces greater galvanic skin responses and increases eye blink frequency in comparison with other colours (Kaiser 1984).[27]

6.4.2.3 *Cultural colouration*

If human attitudes towards objects are influenced by their colour, this may in turn lead to structure in the colours of objects organised by various human relevant factors. Hill and Barton (2005) compared the performance of competitors in boxing, taekwondo, freestyle wrestling, and Greco-Roman wrestling at the 2004

[27] Unfortunately, research on physiological responses to colour seems to have dwindled some time before Kaiser's (1984) review, and doesn't appear to have picked up since.

Olympics. In these events competitors were randomly assigned red or blue uniforms. Hill and Barton found those wearing red were more successful in each sport than would be expected by chance. Similarly, Rowe et al. (2005) found that at the same Olympics, judo was won more often by the competitor wearing blue (as opposed to white). Attrill et al. (2008) found that among English football teams, wearing red has been associated with long-term success since 1947. While the mechanisms by which these asymmetries are manifested is a matter of debate (Rowe et al. 2005; Barton and Hill 2005) these results suggest that certain colours may become associated with sporting success.

The association between colour and success is suggestive, but aspects of the arrangement of professional sports diminishes the degree to which colours could become ubiquitously associated with success (e.g. in the Olympics, colours are reassigned in every round as their purpose is to differentiate the two competitors). However, if we construct a hypothetical scenario in which competitors keep the colour they are assigned, the proportion of competitors wearing the "winning colour" would increase at every round. Fig. 6.18 shows what would happen in a knockout competition involving initially equal numbers of players wearing red and blue if players kept their colour from round to round, red beat blue 55% of the time (the average advantage found by Hill and Barton 2005), and at each round players were randomly assigned opponents from the remaining pool (so players with the same colour sometimes play each other). As with natural selection, small advantages can over time add up to produce large cumulative differences.

We need not assume that sporting contests are the only (or even a significant) way small but consistent differences in the effects of colour on humans could develop into structuring of the colour environment. Another route could be through trade: merchants may find that for a certain kind of item (e.g. fabric) demand is higher for some colours than other (e.g. he may find that he more frequently has to replenish his stocks of one colour than another). This could, perhaps influence his own stock purchasing behaviour or the prices he charges for coloured items. These colour based alterations could have knock on effects, for example on the prestige attached to items of the colour for which a higher price is charged. Once a colour association has emerged, individuals may entrench it by exploiting it. What is important for a small colour based differential response to produce a more categorical colour distinction is that individual preferences accumulate over time, and this may happen in a variety of ways.

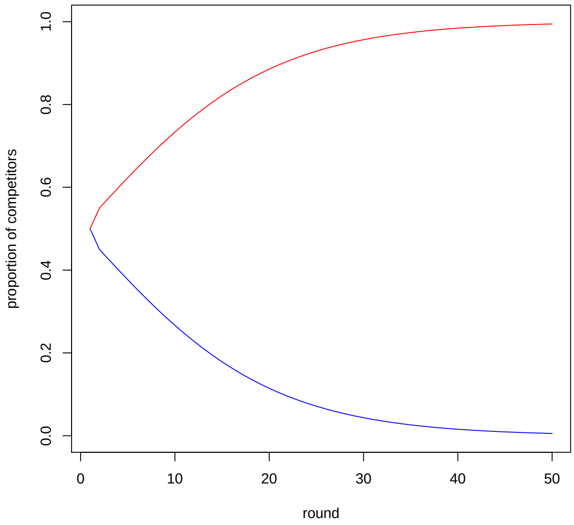


Figure 6.18: Proportion of "red" and "blue" players in a knockout competition. Probability that red beats blue $p = 0.55$

6.4.3 *WCS data and cone responses*

The model of colour terms adopted here rests on human-relevant colours and colour distinctions being correlated across various objects. However, this does not mean that it unreasonably assumes that *all* objects fit into such a colour structure. The patterns of colour terms across languages are here assumed to reflect the patterns of colour of those objects whose colours are relevant to human life and whose colour structure more or less matches the colour structures of other kinds of object. The argument presented here is that when various objects' colour distinctions do correlate, those distinctions will follow certain patterns. The presence of trichromatic organisms creates colour based selection pressures which appear to have influenced the colour of ripe fruits. Human trichromacy may also have created a selection pressure for bare skin giving visual access to information about blood flow beneath the skin. In addition to producing structure in the colour environment, these factors seem to have led to low level differential attention in humans which could contribute to the cultural structuring of the colour environment either directly or by guiding individual's exploration of the world. Combined with the reflectance and radiation spectra of universal aspects of human lives (e.g. chlorophyll, blood, fire, uncooked meat) these factors may underpin physiological responses to colour which in turn could influence individuals' behaviour in relation to different colours. Such influences need not be particularly pronounced on an individual level for them to produce larger cultural effects.

The evidence reviewed above points to red as being particularly important to human life. This is corroborated by ethnographic surveys of the use of colour in ritual. Turner (1966), for example, found that colour in Ndembu ritual culture was structured around the triad of red, white and black, each colour having various significance in different ritual contexts. Sagona (1994)[28] reviewed a number of case studies and found that the Ndembu's colour triad was common across various cultures, with red having particular symbolic significance. Hovers et al. (2003) present archaeological evidence (from the Qafzeh Cave terrace in Israel) that red ochre was used for colouring (either objects or the body) as early as 92,000 years ago. It seems that the red colour was particularly valued as ochre mines yield both yellow and red ochre, both of which have similar non-colour properties, yet in the Qafzeh cave only red ochre was found.

Returning to the WCS data, the approach adopted here suggests that, through the mechanisms discussed above, the colour environment of humans will frequently become organised in a manner that reflects discriminations made possible by the evolutionarily recent L / M cone differences, and that colour term systems develop by the exploitation of this colour organisation. To test this, I compared the positions of Munsell chips in the WCS similarity space (section 6.1.3) with measures of cone activation. In order to do this, I converted the CIE L*a*b* values for each chip (published with the WCS) into CIE XYZ values (using the D65 white point) by standard equations and then converted these into representations of L, M and S cone activation using the formula in equation (2) derived from CIE XYZ functions and normalised cone responsivities (Stockman et al. 1999; Stockman and Sharpe 2000). (Hereafter, L, M and S refer to these numerical representations of cone activation.)

$$\begin{pmatrix} L \\ M \\ S \end{pmatrix} = \begin{pmatrix} 0.2683 & 0.8466 & -0.0349 \\ -0.3859 & 1.1649 & 0.1029 \\ 0.0212 & -0.0245 & 0.5342 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{2}$$

[28] Cited in Hovers et al. (2003, p. 493).

Because values calculated this way are arbitrarily scaled (differently for each cone) direct comparison is inappropriate. To deal with this, I also calculated the luminance (V) corresponding to each chip's CIE L*a*b* value using the formula given by Sharpe et al. (2005). Roughly stated, luminance is a measure of the brightness signal (measured in specific experimental situations) produced by the summation of L and M cone activation (S cone activation is thought to contribute little to luminance). Differences in L and M cone signals were represented by $\log(L/V)$. As L and V are both arbitrarily scaled, $\log(L/V)$ values have an arbitrary additive constant: absolute values are not particularly meaningful, but relative values are, higher values indicating greater L than M cone activation. L/V corresponds to the $L/(L+M)$ measure used standardly in MacLeod-Boynton diagrams to represent cone chromatic responses (Wyszecki and Stiles 1982). The second MacLeod-Boynton axis is $S/(L+M)$ which represents colour distinctions available to dichromats, so I also used $\log(S/V)$ as an additional chromaticity representation. As a third variable I adopted $\log(V)$ to represent luminance. Logarithms were used as under this transformation the distribution of values was more even, aiding graphic representation (no tests reported here were affected by this transformation). The combination of these three values determine the relative activations of L, M and S cones.

Figs. 6.19 and 6.20 show scatter plots comparing the three cone response measures with the dimensions of the similarity space derived from the WCS (all chips and significantly coherently named chips respectively). As predicted, the measure of the difference between L and M cone activation shows a strong relation to the principal axis of the similarity space. These variables show a rough "threshold" relationship, chips above a certain $\log(L/V)$ value being placed at one end of the similarity axis, chips below this level appearing at the other end, confirming that the primary signal in WCS chip naming patterns can be organised by the evolutionarily recent L/M cone distinction which, I have argued, carries significance in various aspects of human life. Notably, the principal axis shows little relationship with $\log(S/V)$, a representation of chromaticity available to dichromats (both modern and ancestral).

Other axes of the similarity space also appear to be related to measures of cone activation patterns. The second dimension appears to be related to luminosity (separating the chips with greatest luminance from the rest) while the third distinguishes chips with low $\log(S/V)$ values (which are also chips which have mid-range $\log(L/V)$ values). These distinctions would be visible to dichromats,

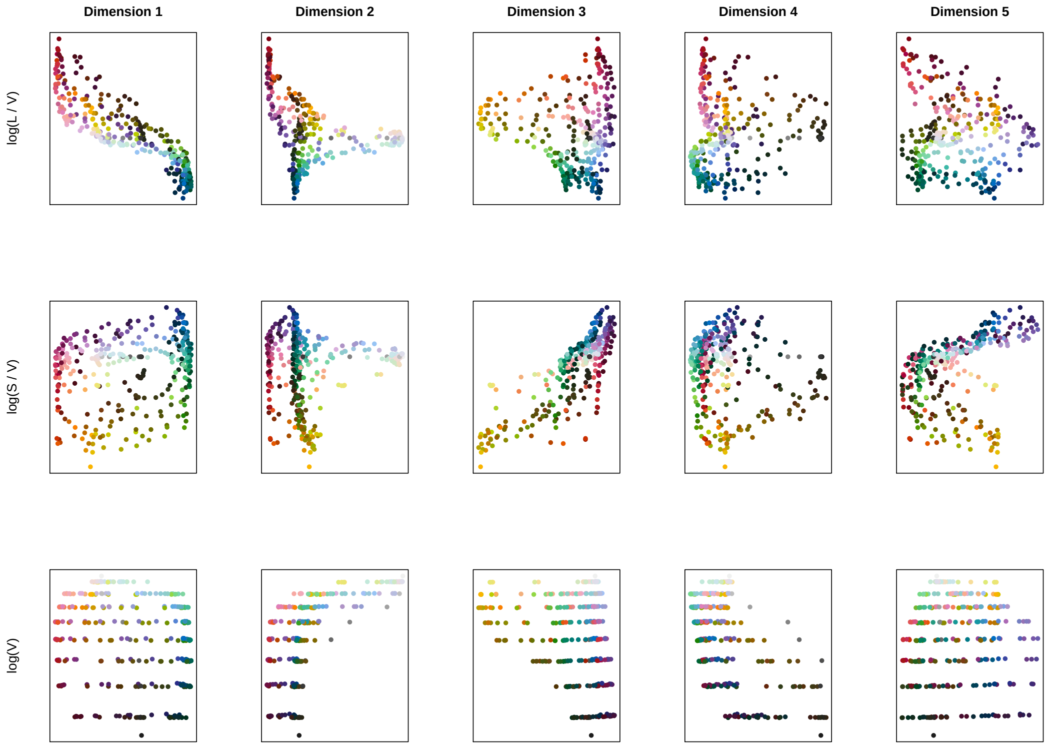


Figure 6.19: Chip positions in similarity space plotted against measures of cone activation patterns.

though the chips picked out by the third dimension noticeably do not seem randomly related to $\log(L/V)$ values (as appears to be the case for dimension 1 and $\log(S/V)$): the chips separated by dimension 3 lie between the extreme $\log(L/V)$ values of red and green/blue.

I performed Spearman's rank correlations comparing each dimension of similarity space with the three measures of cone chromaticity response. Because cone responses vary systematically across the Munsell array, and the axes of discrimination space preserve the array structure even for randomised data (Fig. 6.11) significant correlations (found, for example, for all correlations with the principal similarity axis) are unsurprising. However, instead of significance testing, I asked whether the axes of the similarity space would show stronger relationships with patterns of cone activation if the WCS data was systematically shifted across the Munsell array (i.e. all responses shifted by the same amount rather than different shifts for each language or speaker as above). Fig. 6.21 shows the

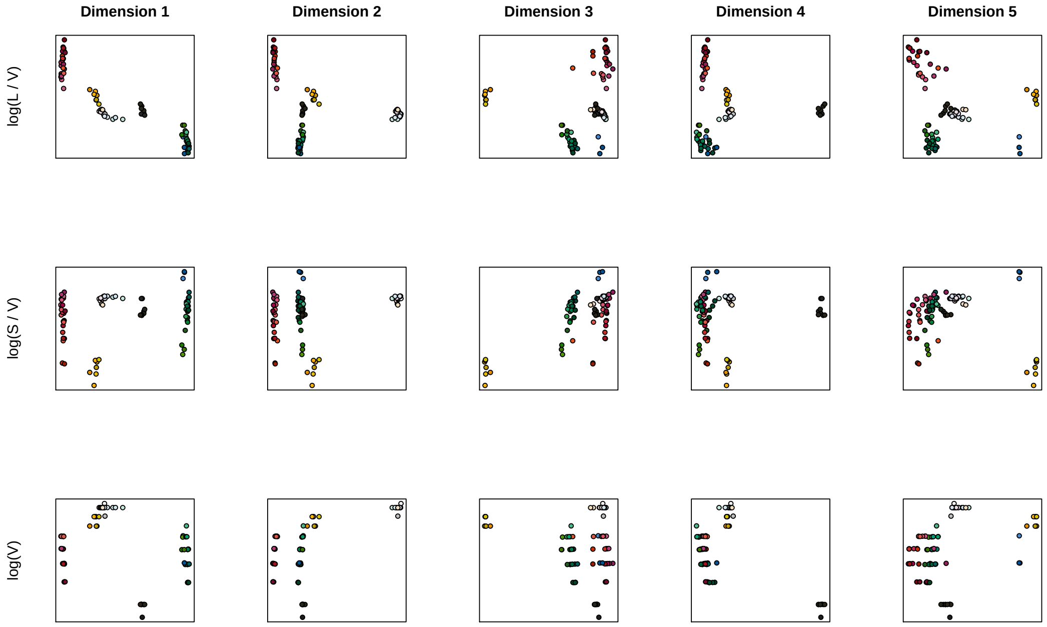


Figure 6.20: Positions of chips whose naming coherence measure was significant plotted in similarity space against measures of cone activation patterns.

result of these calculations. The top left panel shows the correlation between similarity space dimension 1 and the L/M cone difference measure for non-shifted data is almost maximum (only two small shifts, of one and two chips across the Munsell array, produce marginally stronger correlations). Dimension 1 shows a relatively weak correlation with $\log(S/V)$ and $\log(V)$ confirming that the distinctions picked out by the principal axis of the discrimination space would be particularly difficult for dichromats to handle.

The bottom row of Fig. 6.21 indicates that luminance shows the strongest correlation with dimension 2 (and shifting the WCS responses does not affect this correlation particularly). Dimension 3 shows minimal correlation with $\log(L/V)$ (indeed, this was one of the few correlations that failed to reach significance at $\alpha = 0.05$). The correlation between dimension 3 and $\log(S/V)$, while large, was relatively small compared with the maximum possible produced by shifting the WCS. This suggests that dimension 3, while making chip distinctions visible to dichromats, is nonetheless related to trichromatic vision: identification by comparison of S cone activation and luminosity is important for this dimension, but L and M cone differences are also relevant.

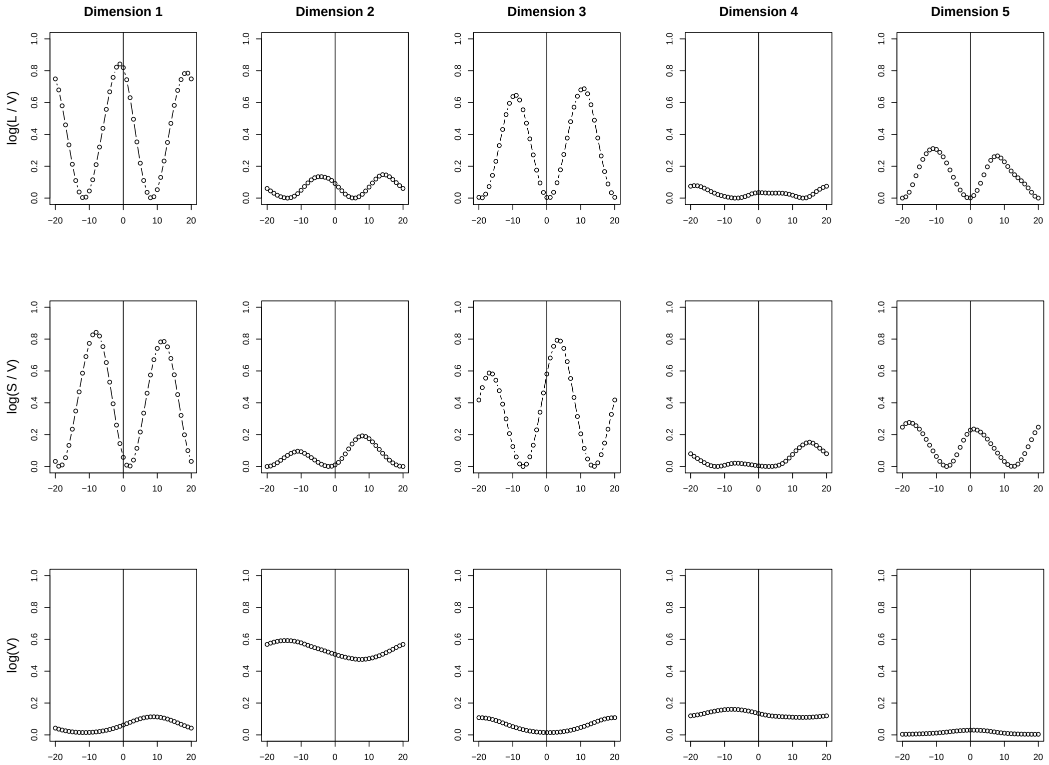


Figure 6.21: Spearman's correlation between Munsell chips' positions on similarity space's principal dimensions and their effect on the human eye. (Labels indicate the correlated variables, not axis scales. On all plots, x-axis represents the shift of chips across the Munsell chart, y-axis r^2. Vertical lines indicate zero shift, i.e. the original WCS data.)

Dimension 4 appears to be related to luminosity, picking out darker chips. Dimension 5 shows a locally maximum correlation to $\log(S/V)$, and minimum correlation with $\log(L/V)$, apparently separating the high $\log(S/V)$ chips from the others (particularly, from the green chips).

6.4.4 Discussion

The importance of the evolutionarily recent L/M cone distinctions for the naming distinctions of the WCS data support the proposal that this aspect of human biology plays an important role in structuring human colour environments. Across languages, the chromatic Munsell chips consistently named within languages are coloured red, green and yellow, the distinction between red and green being most prominent. This latter distinction corresponds to the extremes of the

L/M cone activation ratio and also to the colours of blood and foliage, which appear to have played roles in the evolution of trichromacy, the colour environment and human physiological colour responses.

Yellow If the colour environment (and hence colour terms) are principally organised in terms of L/M cone activations, why do yellow chips (intermediate between red and green in L/M terms) appear to form a category across the WCS languages? As there are various ways in which small differential colour responses may culturally become associations between colour and some other aspect of human life, the importance of the L/M cone distinctions may be manifest in different ways in different cultures. If red and green objects lie at the extremities of such differential responses, then different cultural processes may isolate one or the other of these colour regions from other colour regions. That is, some processes may produce a green vs red/yellow distinction while others lead to red vs yellow/green distinctions. These possibilities are reflected in the two languages depicted in Figs. 6.8 and 6.9. Consider a culture in which both distinctions were present: this culture may develop a repertoire of colour terms structured as A="red", B="yellow/green", C="red/yellow", D="green". In such circumstances, the increasingly object general use of colour terms A (or D) may replace some uses of C (or B) leaving only yellow-related C (or B) uses. Indeed, English "yellow" appears to have followed something like this route, deriving from Indo-European **ghelwo* which seems to have denoted both green and yellow (OED).

Blue The distinction between green and blue Munsell chips in the WCS similarity space is relatively small and a minor aspect of the positioning of chips in that space (Fig. 6.10). According to Kay and Maffi's (1999) analyses, the majority of WCS languages do not distinguish between green and blue Munsell chips. This may be taken as a reflection of the relative unimportance of green/blue distinctions to human life. The blue/green distinction is visible to dichromats (Fig. 6.20, middle row) but it seems that the evolutionary history of the human capacity for this distinction does not depend on distinguishing objects on the basis of their colour, but has more to do with the composition of ambient light in the environment of small nocturnal mammals (Vorobyev 2004; Goldsmith 1990; Peichl et al. 2000). We might therefore expect that differential responses associated with L/M differences would not be mirrored by differential responses associate with S/(L+M) differences, and so blue/green distinctions would become part of a culture's colour environment less frequently.

Light/dark The focus on cone activation patterns presented above was aimed at accounting for chromatic rather than lightness distinctions in the WCS. However, both dimensions 2 and 4 of the similarity space appear to order Munsell chips by lightness (Figs. 6.10 and 6.21). Fig. 6.8 represents a language (Nafaanra) that appears to have a term denoting very light colours, while Lele (Fig. 6.9) has distinct terms identifying the lightest and darkest chromatic Munsell chips. The former pattern is more common among the WCS languages than the latter though this may result in part from the fact that the lightest chromatic chips are closer together in CIE $L^*a^*b^*$ space than the darkest: six nearest neighbour distances fall below the $1\Delta E$ discrimination threshold for the lightest chips while all of the darkest chip nearest neighbour distances exceed this level.

Lightness distinctions are visible to monochromats and to low light vision which relies only on rods. As such, the lightness of an object may generally have relevance to human life, correlating with how easily the object may be found under poor illumination. Other respects in which the darkest and lightest Munsell chips may correspond to relevant colour distinctions include the effects of burning and sun bleaching (producing relatively dark and light colours respectively). Thus lightness may also be a structuring factor in human-relevant object colours.

The evolution of colour terms This chapter has attempted to account for universal patterns in languages' colour terms by adopting a communicational affordances perspective. That perspective suggested that similarities in colour term structure is a reflection of similarities in the colours and colour distinctions relevant to human life. Language independent reasons for common patterns of colour were found in the evolution of colour vision and the interaction between trichromacy and colour organisation of the environment. The relationships between cone activation and linguistic colour distinctions were interpreted in terms of natural colour signals exploiting trichromacy, trichromatic individuals responding by differentially responding on the basis of colour and this response feeding into cultural processes which amplify small colour-differences into more categorical associations. However, the results in Figs. 6.19, 6.20 and 6.21 could also be interpreted in other ways (e.g. as evidence that colour systems result from the externalisation of language-independent internal representations of the meanings of colour terms). The discussions in this chapter have shown that if the idea of externalisation is rejected (on the grounds set out earlier in this thesis), structures in the colour term systems of the worlds' languages can still be accounted for.

In this chapter I have made little use of the idea that languages may change from one colour term system to another. The evidence that languages follow trajectories through colour term systems is weak, being based not on longitudinal observation of language changes but on the assumption that attested systems develop from other attested systems with fewer colour terms (Berlin and Kay 1969; Kay and Maffi 1999). If all colour term systems fit onto a single trajectory, the evolutionary hypothesis would be plausible (but would still benefit from longitudinal evidence). However, several apparently *ad hock* trajectories are required to force the attested colour term systems into evolutionary sequences, making suspect the assumption that languages all develop towards a colour term system like English's (Saunders and van Brakel 2002). The approach to colour term systems taken in this chapter neither relies on progression from one system to another, nor does it rule this possibility out.

What is distinctive about the approach to colour terms presented here is that the use of words to talk about object colours is not assumed to be in place prior to the development of particular colour term systems (as is the case with Dowman's and Belpaeme and Bleys' models). Instead, a generalised account of the development of such language games was presented and this formed the basis of the explanation for structure within those games.

CHAPTER 7

Conclusions

This thesis is about language games. Not only does it concern the evolution of the linguistic practices in which humans engage, but it also addresses how this language game, the description of language evolution, should be played. It argues for a reorientation of the field of language evolution in such a way as to make it more consistent with other language games in which the distant past (beyond the reach of memory and written records) is described. To do this, certain inadequacies in existing research approaches have been highlighted, and alternative research directions have been proposed which have (in the case of colour terms) been shown to be fruitful.

More specifically, this thesis has argued that the study of language evolution should go *beyond the individual*. There are two senses in which this is intended. Firstly, the use of models of internal processes underpinning language should be eschewed as there is little reason to suppose that existing or foreseeable models can tell us anything about hominid behaviour at earlier stages in language evolution. Secondly, a fruitful line of investigation into the evolution of language is to consider the community of others in which an individual is embedded and how learning and exploiting the communicational affordances offered by that community in turn impacts on the affordances offered to others. This direction of thought goes beyond the individual in the sense that properties of emergent communicational systems and tendencies for these systems to change in certain ways need not be thought of as reflecting the tendencies or preferences of individuals to use or to change their language in corresponding ways.

A recurring theme in this thesis is that the study of the evolution of language is often hampered by the adoption of the theoretical apparatus of contemporary linguistics. Much of the criticism of language evolution theories can also be applied to (and in some cases derives from) criticism of linguistic theories, and finds parallels in other critiques of these theories (e.g. Harris 1998a). A central aspect of this criticism is the inadequate treatment of what it is for words to have meaning. Jackendoff (2002) laments the "syntactocentrism" of contemporary linguistics, but the oversight of meaning in linguistic theorising is not simply a failure to incorporate a notion of meaning into a mechanistic model of language production. It is rather, as Harris (1998a) emphasises, a failure to reflect on the linguist's own use of language. Such a failure leads, for example, to the erroneous assumption that because a word (such as "phoneme") has a meaning (that is, a use in a language game) it must also have a referent. When no *thing* is forthcoming to fulfil that role, "we say that a *spiritual* [mental, intellectual] activity corresponds to these words," (*PI* §36). One aspect of the reorientation of the study of language evolution called for by this thesis involves a closer attentiveness to the theorist's own language games in order to avoid falling into conceptual traps.

This final chapter draws the themes of the thesis together to provide an overview of how the thesis hopes to influence the field of language evolution.

7.1 Describing the distant past

As noted above, theorising about the evolution of language involves engaging in the language games of describing the distant past. What are the rules by which this set of games is played for events and processes other than the evolution of language? In *On Certainty*, Wittgenstein draws attention to some of the rules governing our descriptions of the past (e.g. what counts as evidence that the world existed before my birth). An important Wittgensteinian point is that, like other language games, describing the distant past is an activity governed by *arbitrary* rules (in the sense described in section 4.2.4). For example, justification of something as good evidence concerning the past bottoms out at some point: clearly justifying something as evidence by appealing to some other evidence (that shows the first piece to really be evidence) simply postpones the question (for on what grounds is the latter piece of evidence justified). Noting this arbitrariness in the language game should not be taken to imply that descriptions of the distant past are "mere" constructs (as if there were some other "distant past" of which we could talk, that is, some other language game with which the distant past could *really* be accessed). To question whether what we conclude about the distant past on the basis of what we (arbitrarily) take as evidence *really* corresponds to what happened in the past is to go round in a logical circle (*OC* §191).

The importance of noting the arbitrary rules constituting the language game of correctly describing the distant past lies in the fact that in constructing theories of the distant past of language evolution, we should abide by these rules (otherwise we will be playing a different language game). Botha's (2006a) "windows approach" to language evolution, described in section 2.1, can be seen as outlining some of the features required of such language games. That a window on language evolution should be grounded, warranted and pertinent are arbitrary requirements of the language game, but they are requirements nonetheless. Chapters 2 and 3 of this thesis address the degree to which theories of language evolution which rely on the notion of the externalisation of internal meaning representations satisfy these requirements. While the notion of externalisation was found to play crucial roles in various competing accounts of language evolution, the lack of empirical evidence showing the warrantedness of these assumptions undermines the position of these accounts in broader descriptions of the past. Without these features, purported accounts of language evolution hang disconnected from the distant–past–language–game as it is ordinarily played.

7.1.1 The protolanguage language game

A common feature of the accounts of protolanguage discussed in section 2.2 is that little attention is paid to what it amounts to to suggest that a protolanguage utterance had a particular meaning. In this, they parallel an idea of "the awakening of consciousness" which Wittgenstein mentions:

> The evolution of the higher animals and of man, and the awakening of consciousness at a particular level. The picture is something like this: Though the ether is filled with vibrations the world is dark. But one day man opens his seeing eye, and there is light.
>
> What this language primarily describes is a picture. What is to be done with the picture, how it is to be used, is still obscure. Quite clearly, however, it must be explored if we want to understand the sense of what we are saying. But the picture seems to spare us this work: it already points to a particular use. This is how it takes us in. (*PI* p. 184)

Attributions of meaning to protolanguage present a picture, as does the statement that the meaning so attributed is "in the mind" of the protolanguage user. This kind of picture is where discussion of meaning in protolanguage theories generally stops, and this makes it difficult to assess the claims of these theories. The attempt to explore this picture, to understand the sense of these attributions, is a novel contribution of this thesis. Meaning attributions were cashed out in chapter 2 in terms of the observations a field linguist would require in order to be satisfied that a particular meaning attribution was a good translation. This needn't be the *only* way in which we could give sense to these attributions, though it has the distinct advantage of paralleling what is actually done when a field linguist attributes meanings to a language previously unknown to him. Any other approach, by relying on some novel criterion (such as, for example, the protolanguage user's brain being in a particular configuration) would be disconnected from our ordinary practice of meaning attributions, in which case we might as well use a word other than "meaning" to describe the attribution.

The discussion of meaning attributions to protolanguage concluded that if these attributions amount to anything, they amount to the assumption that protolanguage users would use expressions in various, flexible ways. At this point in the discussion there was a certain degree of vagueness as to what behavioural observations the field linguist would make. There are two reasons for this vagueness: different sets of criteria would be appropriate for different meaning attributions, and so attempting to give a concrete description good for all protolanguage meaning attributions would distort this variety; and by hypothesis the forms of life of protolanguage speakers would have been different to those of speakers of modern languages, leading to various indeterminacies (as in this unusual context it may not be clear how to apply the rules of our language) which the field linguist may *decide* to clear up one way or another. However, in spite of this inherent vagueness, flexibility of use emerged as an important criterion and was used in later chapters to judge the role of externalisation: evidence was sought that usage generalisations corresponding to protolanguage–theory meaning attributions could be safely assumed on the basis of externalisation of internal (neural) structures.

7.2 Empirical judgement of the role of externalisation

Chapter 3 sought empirical evidence which could be called on to justify the externalisation assumptions made by protolanguage theories outlined in chapter 2, and concluded that such evidence is sorely lacking, both in terms of neural structures and ordinary human linguistic and language-learning behaviour.

A point worth stressing in terms of neuroscience is that while the literature makes much use of the idea of semantic representations as the internal counterparts to external meaningful words, this is not generally due to the *findings* of neuroscience but is a feature, rather, of the way in which research is framed and organised. That is to say, the idea of semantic representations appears in the empirical research literature as a preconception rather than a result. The chaotic, contradictory and confusing empirical results that emerge within this framing do little to inspire confidence that this is a fruitful way to approach the mechanistic workings of the brain involved in language use. More specifically, nothing emerges from this research literature to suggest that spontaneous usage generalisations corresponding to meaning attributions in theories of protolanguage may safely be assumed on the basis of neural structures.

Does this mean that there are no semantic representations in the human brain? This question is misleading, suggesting that there is a clear sense to the expression "semantic representations in the brain". Much of the discussion of language games concerning "meaning" and concerning "neural structures" in chapter 4 sought to show how unclear (how indeterminate) this particular expression it. "Semantics" and more broadly "meaning" have roles in language games totally unconnected with neuroscience as can be seen from the fact that these games can be played (e.g. judgements can in many cases be made about whether a particular meaning attribution is correct or not) in complete ignorance of the physical workings of the brain. Section 4.4.3 dealt explicitly with the assumption that while we do not know *how* the semantic representations dreamed up by linguists relate to the brain, nevertheless they must relate somehow. This assumption, I suggested, traded on an analogy with computers: while a computer user may be unaware of the relationship between the abstract notion of information stored in the computer and the physical structure of the computer, nonetheless such a relationship exists. However, the analogy breaks down when it is realised that what this computer user is ignorant of is something someone else knows (that is, there exists a language game, a set of practices, against which a proposed relationship may be judged right or wrong). In contrast, there is no such standard, of which we may be ignorant or knowledgeable, for the relationship between the physical brain and the abstract descriptions of the mind/brain produced by linguists.

In the absence of any such standard, the question shifts from "are these semantic representations present in the brain?" to "would it be possible to produce an abstract description of the brain, couched in terms of these semantic representations?" I argued in section 4.4.3.1 that the answer to this latter question is, as a matter of logic rather than empirical discovery, "yes". In the absence of established bridge laws connecting physical to abstract descriptions, it only takes a degree of ruthlessness to insist that these laws be gerrymandered to satisfy the requirement that a rule governed relationship obtain. Wittgenstein offers an illuminating parallel case in *TLP* § 6.341:

> ... Let us imagine a white surface with irregular black spots on it. We then say that whatever kind of picture these make, I can always approximate as closely as I wish to the description of it by covering the surface with a sufficiently fine square mesh, and then saying of every square whether it is black or white. In this way I shall have imposed a unified form on the description of the surface. The form is optional, since I could have achieved the same result by using a net with a triangular or hexagonal mesh. ... *TLP* § 6.341

The important point is that the possibility of describing ink spots with a certain mesh (or the brain according to a particular scheme of semantic representation), is that it tells us nothing about the ink spots (or the physical brain). This is because any configuration of ink spots could be described using a sufficiently fine square mesh, so the possibility of using such a mesh does not alter our state of ignorance concerning the actual configuration of ink spots.

By adding epicycles within epicycles, it would in principle be possible to translate physical descriptions of the brain into descriptions couched in terms of semantic representations. However, given the likelihood (based on the empirical lack of a simple fit between neuroscientific results and linguistic-theory descriptions) that such a scheme would be thoroughly complicated, the question remains whether behavioural predictions made on the basis of the externalisation-framework are borne out (and so can be assumed for earlier stages of the evolution of language). Section 3.3 addressed this question, looking for evidence of spontaneous usage generalisation early in language acquisition and setting this against evidence that usage patterns are highly experience dependent. This section highlighted findings that early usage generalisations can be experimentally manipulated, and that as early as around three months, adults give contingent stimulation to infants on the basis of their own expectations of the behaviour of a communicative partner. These and other findings discussed in section 3.3 suggest that immersion in a linguistic culture shapes the usage generalisations made by language learners, and undermines the protolanguage assumption that analogous generalisations would have been spontaneously made by our ancestors in very different communicative environments.

7.3 Learning to exploit communicative affordances

This thesis's assessment of language evolution theories' reliance on models of the "internal structure" of individuals resulted in a generally negative conclusion. These models are not supported by empirical neuroscientific results (which could be used as justification for projecting models back onto our ancestors) but by various conceptual confusions: for example, grammatical rules are misinterpreted as empirical laws due to superficial similarities (section 4.4.2), and behavioural observations are re-packaged as mechanisms which are the source of those behaviours (section 4.4.1, c.f. Harris 1981, p. 27). However, this thesis aims to have a positive impact on the field of language evolution, and so sets about describing how language evolution can be studied without recourse to models of individuals' internal structures.

The approach advocated rests on the notion of communicative affordances, detailed in chapter 5, but already touched on in the discussion of caregiver construction of infant/child meaning in section 3.3.2. Not only are adult responses to infants influenced by (culturally variable) adult expectations about the behaviour of communicative partners, but infants in turn respond to adult responses and, as Masataka (2003, p. 88) notes, learn how to elicit certain responses in adults. That is, there is evidence that as early as three months, infants are learning how to exploit the communicative affordances offered by adults in their environments.

Chapter 5 applied the idea of affordance exploitation and the shifts in affordances brought about by others' learning, to address questions of the form "why do languages have (or why does this language have) this particular feature?" This form is characteristic of some (but not all) questions that language evolution theorists have attempted to answer using approaches based on models of individuals' internal structure (e.g. Kirby 1999). In this thesis, affordance dynamics were applied to various attested language changes to show how the approach could be operated, and to demonstrate the positive possibilities of going beyond the individual when thinking about language change and language evolution. Certain generalisations about language change were shown to match general properties of affordance dynamics. For example, changes from concrete to abstract uses, when considered in terms of the different characteristic affordances for use of concrete and abstract expressions, could be understood as reflecting two simple features of affordance dynamics. First, in its early life a novel expression will likely have a relatively limited range of uses (a characteristic of concrete expressions), and second, there are many more usage patterns which would be characterised as abstract than concrete (so changes are more likely to produce abstract expressions than concrete ones).

Other tendencies for language to change in particular ways could also be accounted for by the affordances approach at varying levels of generality. For example, the development of futures from spatial expressions can be seen as a reflection of the fact that very often people travel from one place to another in order to do something at their destination. This basic structure of human life acts as a scaffolding on which the development may occur (e.g. that describing someone's "going" to a particular place can be linked to their intention to do something there). The characteristic difference between expressions translated as 'go' and 'come' (namely that the latter are used in connection with the place where the speaker is while the former relates to some other place) can then feed through simply (though not without exception) to differences in uses of those expressions as futures (section 5.3.4).

7.3.1 Communicative affordances in the distant past

Having shown how an affordance–dynamics approach can be applied to attested language changes, and having argued that such an approach is conceptually superior to one based on models of individuals' cognitions, chapter 5 turned its attention towards developments in languages which are not attested due to their pre-dating recognisable modern languages and happening in contexts which have no modern parallel (e.g. no contact with any language-using culture). Drawing on discussions of attested changes, consistent features of human life which could act as scaffolding guiding the development of communicative affordances were sought. For example, the ubiquitous use of expressions as labels for objects was argued to rest on the structuring of human life around concrete objects and, importantly, the ways in which human life structures objects into instances of the same or different *kind* of object (section 5.4.2). Such general developments were naturally presented in general terms, and numerous different routes could be envisaged (that is, language games in which expressions are used as labels for objects could develop from numerous different imaginable language games). This was one reason why these developments were not presented in terms of specific characterisation of earlier stages and the route by which modern language characteristics developed. Another reason, emphasised throughout earlier chapters, is that we lack a firm basis on which to construct accounts of the behaviours and forms of life engaged in by our ancestors, especially given the possibility of significant biological differences between them and us.

While the accounts offered in chapter 5 avoided relying on details of communicative practices in the distant past, nonetheless in engaging with questions of language evolution in the distant past they may be subjected to similar scrutiny to that applied to protolanguage theories. How well do these communicative–affordances–based accounts fit into describing–the–distant–past language games? The communicative affordances approach does not rely on any undiscovered theoretical entities (such as neural structures) but instead appeals to general principles of human/hominid/animal interactions and the affordances offered by the environment (both the social and physical environments). Accounts of language evolution based in communicative affordances can and should make clear that there are sound reasons for the assumptions they make about earlier forms of life. For example, that objects played a structuring role in the lives of our ancestors can be seen as a something of which we can be certain. This certainty does not rely on evidence about the distant past, but forms part of the framework within which we talk about the distant past (in Wittgenstein's terms, it is a proposition, fixed by the language game, which forms part of the river-bed through which flow other propositions, those that are empirical and hence fluid, *OC* §97).

More complex and empirically supported (and contestable) links to the distant past can be called on using the affordances approach. In chapter 6, it was suggested that the development in various languages of expressions with grammar similar to English colour terms could result from correlations between those colour contrasts which are relevant to human life (e.g. contrasts between ripe and unripe fruits). Evidence for such structuring effects across human communities, stretching far back into the past, was sought by considering the evolution of the trichromatic visual system. In this respect, the account engaged with a well established language game concerning the distant past.

While a range of possibilities concerning the route by which colour terms developed were held open in chapter 6, certain assumptions about the behaviour of individuals within structured colour environments were nonetheless relied upon (for example, that individuals would learn to exploit only parts of the affordances offered to them, see section 6.3.2). Such assumptions were not supported by observation that such behavioural dynamics do in fact occur, but were presented as reasonable and easy to imagine for human beings unacquainted with colour terms. However, these assumptions receive indirect support from the confirmation of predictions drawn from the account of colour term development (namely, the fit of WCS aggregate data to the biophysics of the human eye). The colour term account in particular and the affordances approach more generally would be bolstered by further empirical validation, for example matching differences between languages' colour term systems with differences in the visual ecology of those languages' speakers.

7.3.2 *Tackling the language gap*

The account of colour terms in chapter 6 demonstrates that an affordance-dynamics approach to language evolution can fruitfully address questions about language structure. However, this is not the only kind of question which interests researchers in the field of language evolution. Another set of questions (sometimes, depending on theoretical perspective, intertwined with questions about language structure) concern the biological evolution of human beings, and are focussed on the differences between us and other animals relevant to our ability (and their seeming inability) to use language. As Taylor (1997) argues, the difference between those who make sense (or make sense *as we do*) and those who do not, be it a difference between humans within and without a culture or between humans and other animals, has historically been a culturally significant phenomenon, stimulating an interest in the origins of language and leading to various origins myths. Taylor suggests these myths perform the function of justifying *why* "we" make sense and "they" do not.

Various differences between humans and other animals are identified as crucial by investigators in the field of language evolution, including differences in neural computational abilities (Pinker and Bloom 1990; Hauser, Chomsky, and Fitch

2002), differences in mental representations of others' internal lives (Tomasello, Carpenter, Call, Behne, and Moll 2005) and differences in altruistic tendencies (Fitch 2007; Dessalles 2007). In various ways, these differences, once identified, are taken as posing evolutionary puzzles: how could evolution by natural selection, gradually operating on small differences, result in these seemingly large differences between humans and other animals? This question is particularly acute when there is apparently no gradual scale between animals and humans which evolution could traverse: for example, a computational system either has recursive properties or it does not (Hauser, Chomsky, and Fitch 2002). Taylor (1997) notes that some theorists (including Darwin himself) have taken the problem of a "language gap" as threatening the foundations of Darwinian evolution, and requiring an overhaul of evolutionary theory.

As Taylor (1997) makes clear, the language gap is generally the product of the way theorists frame "the nature of the mind" and "the nature of language". Discussions in chapter 4 highlighted ways in which such framings could be (and often are) misunderstood as discoveries analogous to discovering that the human body has an appendix. For example, that language is a unified system of which it makes sense to talk about the origin is an assumption (or even a requirement) of many linguistic theories, but when we follow Wittgenstein's advice, "don't think, but look!" (*PI* §66), we find that what we call language is a variety of overlapping activities linked by family resemblances rather than a common core (see section 4.2.4). The illusion that language has a unified core may be encouraged by a focus on the fact that humans do but animals do not have language, as this statement may give the impression that a single entity (language) has been identified.

Taylor (1997) argues that to deal with the language gap and associated questions about the origin of language, a therapeutic approach is required, weaning the theorist off the temptation to see his own preconceptions as discoveries requiring evolutionary solutions. Much of chapter 4 of this thesis offers such therapeutic insights: by surveying how language games within cognitive science and linguistics operate, the temptation to succumb to conceptual confusion will hopefully be made less alluring. The affordances approach to language evolution also has the potential to have a therapeutic impact: by explicitly imagining developmental paths through which language games can develop, the requirement that some particular cognitive capacity is required (and that this represents an evolutionary puzzle) can be revealed as bogus. For example, the hypothetical development of role reversal imitation, explored in section 5.2.1, required no change to the constitution of individuals but was a (possible) product of affordance dynamics. In this way, affordance-dynamics accounts can wean the theorist off the temptation to view each way of describing a difference between humans and animals (e.g. we employ role reversal imitation while they do not) as *necessarily* corresponding to a biological difference which requires an evolutionary explanation.

By disconnecting the categories of theoretical descriptions of language from biology, evolutionary paradoxes associated with the "language gap" may be avoided. The requirement of a biological evolutionary story corresponding to each facet of our metalinguistic language games is a mistake. However, this reorientation may serve to make the relationship between language and evolutionary biology *more* complicated. It would be fatuous to deny that there are numerous differences between humans and other animals (such as different degrees of facial mobility, see section 5.4.1) which impact on our abilities to engage with the language games afforded by human communities. Similarly, there is little reason to suppose that the biological evolution of these differences occurred entirely independently of the context of some form of communication (that is to say, in addition to affording certain interactions within individuals' lives, the existence of shared communicative practices may well have altered the fitness landscape leading to some form of biological adaptation to the communicative context). The relationships between genetic variation, the abilities to engage in the various activities we group together under the family resemblance term "language", and impacts on fitness are likely to be thoroughly complex. The study of the evolution of biological aspects of our abilities to use language will consequently be far more convoluted than "language gap" theorists suggest, but also far less paradoxical.

7.4 Beyond the individual

This thesis has argued for a reorientation of the field of language evolution studies. Models of the mechanisms internal to individuals which produce language have been shown to be empirically baseless, and founded on conceptual confusions. For this reason, I have attempted to go beyond the individual in the sense of investigating the evolution of language without recourse to such models of the workings of individuals. To do this, I adopted an approach that goes beyond the individual in a second sense, focusing on the affordances that exist within a community and the impact of individuals learning to exploit those affordances. Following this line, aspects of languages and tendencies for languages to change in certain ways are not seen as reflections of properties of individuals, but emerge from the dynamics of interaction within structured contexts. I advocate this approach to the evolution of language, not only because it avoids resting on the false foundations of a cognitive model, but also because, as demonstrated by the investigation into colour terms in chapter 6, it opens up a rich and fruitful line of research.

As mentioned in the introduction to this thesis, investigations into the evolution of language are often prompted by an interest in "what exactly it is that makes us human," (Christiansen and Kirby 2003b, p. 1). This way of asking the question suggests both we should discover a single "thing", and that the concept of "what makes us human" is far more determinate than it actually is. When we *look* at language, at the various activities which fall under the heading "using language", rather than assume what language *ought* to be like, we find a variety of activities, more or less related in various ways, with some clear and some not–so–clear examples. Likewise, the histories of languages are vast and have similarly indistinct boundaries. No single "thing" emerges as that which make us human or which is the basis of the evolution of language. Adopting an affordance-dynamics approach, the field of language evolution can embrace and explore this richness and diversity, and see it as a reflection of the richness and diversity of human life.

References

Albanese, E., E. Capitani, R. Barbarotto, and M. Laiacona (2000). Semantic category dissociations, familiarity and gender. *Cortex 36*(5), 733–746.

Andrews, E. (1995). Seeing is believing: visual categories in Russian lexicon. In E. Contini-Morava, B. S. Goldberg, and R. S. Kirsner (Eds.), *Meaning as Explanation: Advances in Linguistic Sign Theory*, pp. 363–377. Berlin: Mouton de Gruyter.

Arbib, M. A. (2003). The evolving mirror system: a neural basis for language readiness. See Christiansen and Kirby (2003a), Chapter 10, pp. 182–200.

Attrill, M. J., K. A. Gresty, R. A. Hill, and R. A. Barton (2008). Red shirt colour is associated with long-term team success in English football. *Journal of Sports Sciences 26*(6), 577–582.

van der Auwera, J. and V. A. Plungian (1998). Modality's semantic map. *Linguistic Typology 2*, 79–124.

Baker, G. P. and P. M. S. Hacker (1984). *Language, Sense and Nonsense*. Oxford: Blackwell.

Baker, G. P. and P. M. S. Hacker (1985). *Wittgenstein: Rules, Grammar and Necessity*. Oxford: Blackwell.

Barbarotto, R., E. Capitani, and M. Laiacona (2001). Living musical instruments and inanimate body parts? *Neuropsychologia 39*(4), 406–414.

Barton, R. A. and R. A. Hill (2005). Seeing red? Putting sportswear in context: Reply. *Nature 435*(293), E10–E11.

Bartsch, R. (1984). Norms, tolerance, lexical change, and context-dependence of meaning. *Journal of Pragmatics 8*(3), 367–393.

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy, and C. Knight (Eds.), *Approaches to the Evolution of Language: Social and Cognitive Bases*, Chapter 24, pp. 405–426. Cambridge: Cambridge University Press.

Bates, E. and J. C. Goodman (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language and Cognitive Processes 12*(5), 507–586.

Bates, E., S. Vicari, and D. Trauner (1999). Neural mediation of language development: perspectives from lesion studies of infants and children. In H. Tager-Flusberg (Ed.), *Neurodevelopmental Disorders: Contributions to a New Framework from the Cognitive Neurosciences*, pp. 533–581. Cambridge, MA: M.I.T. Press.

Bates, E., B. Wulfeck, and B. MacWhinney (1991). Cross-linguistic research in aphasia: An overview. *Brain and Language 41*(2), 123–148.

Belpaeme, T. and J. Bleys (2005a). Colourful language and colour categories. In A. Cangelosi and C. Nehaniv (Eds.), *Proceedings of the Second International Symposium on the Emergence and Evolution of Linguistic Communication (EELC 2005)*, Hatfield, pp. 1–8.

Belpaeme, T. and J. Bleys (2005b). Explaining universal color categories through a constrained acquisition process. *Adaptive Behavior 13*(4), 293–311.

Bennet, M. R. and P. M. S. Hacker (2003). *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.

Berko, J. and R. Brown (1960). Psycholinguistic research methods. In P. H. Mussen (Ed.), *Handbook of Research Methods in Child Development*. New York: Wiley.

Berlin, B. and P. Kay (1969). *Basic Color Terms: Their Universality and Evolution*. Berkley: University of California Press.

Bickerton, D. (1990). *Language and Species*. Chicago: University of Chicago Press.

Bickerton, D. (1995). *Language and Human Behaviour*. London: UCL Press.

Bickerton, D. (2000). How protolanguage became language. See Knight, Studdert-Kennedy, and Hurford (2000), Chapter 16, pp. 264–284.

Bickerton, D. (2002). Foraging versus social intelligence in the evolution of protolanguage. See Wray (2002b), Chapter 10, pp. 207–225.

Bickerton, D. (2007). Language evolution: A brief guide for linguists. *Lingua 117*(3), 510–526.

Bloom, K. (1988). Quality of adult vocalizations affects the quality of infant vocalizations. *Journal of Child Language 15*, 469–480.

Bloom, K. and N. Masataka (1996). Japanese and Canadian impressions of vocalising infants. *International Journal of Behavioral Development 19*(1), 89–100.

Bloom, K., A. Russell, and K. Wassenberg (1987). Turn taking affects the quality of infant vocalizations. *Journal of Child Language 14*, 211–227.

Bloom, L. (1973). *One Word at a Time: The Use of Single Word Utterances Before Syntax*. The Hague: Mouton.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge, MA: M.I.T. Press.

Booth, A. E. and S. R. Waxman (2002). Word learning is 'smart': evidence that conceptual information affects preschoolers' extension of novel words. *Cognition 84*(1), B11–B22.

Borg, I. and P. J. F. Groenen (1997). *Modern Multidimensional Scaling: Theory and Applications*. New York: Springer-Verlag.

Borgo, F. and T. Shallice (2003). Category specificity and feature knowledge: Evidence from new sensory-quality categories. *Cognitive Neuropsychology 20*(3), 327–353.

Bornstein, M. H. (1975). Qualities of color vision in infancy. *Journal of Experimental Child Psychology 19*, 401–419.

Bornstein, M. H., W. Kessen, and S. Weiskopf (1976). Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception and Performance 2*(1), 115–129.

Botha, R. (2006a). On the windows approach to language evolution. *Language & Communication 26*(2), 129–143.

Botha, R. (2006b). Pidgin languages as a putative window on language evolution. *Language & Communication 26*(1), 1–14.

Botha, R. (2007). On homesign systems as a potential window on language evolution. *Language & Communication 27*(1), 41–53.

Botha, R. P. (2003). *Unravelling the Evolution of Language*. Language & Communication Library. Oxford: Elsevier.

van Brakel, J. (2005). Color is a culturalist category. *Behavioral and Brain Sciences 28*(4), 507–508.

Brambati, S. M., D. Myers, A. Wilson, K. P. Rankin, S. C. Allison, H. J. Rosen, B. L. Miller, and M. L. Gorno-Tempini (2006). The anatomy of category-specific object naming in neurodegenerative diseases. *The Journal of Cognitive Neuroscience 18*(10), 1644–1653.

Bréal, M. (1900). *Semantics: Studies in the Science of Meaning*. London: William Heinemann.

Briscoe, T. (Ed.) (2002). *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge: Cambridge University Press.

Broca, P. (1861). Remarques sur la siège de la faculté de la parole articulée, suivies d'une observation d'aphémie (perte de parole). *Bulletin de la Société d'Anatomie 36*, 330–357.

Buonomano, D. V. and M. M. Merzenich (1998). Cortical plasticity: From synapses to maps. *Annual Review of Neuroscience 21*, 149–186.

Burrows, A. M., B. M. Waller, L. A. Parr, and C. J. Bonar (2006). Muscles of facial expression in the chimpanzee (*Pan troglodytes*): descriptive, comparative and phylogenetic contexts. *Journal of Anatomy 208*(2), 153–168.

Bybee, J. L., R. Perkins, and W. Pagliuca (1994). *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Byrne, A. and D. R. Hilbert (2003). Color realism and color science. *Behavioral and Brain Sciences 26*, 3–64.

Calvin, W. H. and D. Bickerton (2000). *Lingua ex Machina*. Cambridge, MA: M.I.T. Press.

Capitani, E. and M. Laiacona (2005). An illusory illusion? *Cortex 41*(6), 854–855.

Capitani, E., M. Laiacona, B. Mahon, and A. Caramazza (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cognitive Neuropsychology 20*(3), 213–261.

Caramazza, A. and B. Mahon (2006). The organisation of conceptual knowledge in the brain: the future's past and some future directions. *Cognitive Neuropsychology 23*(1), 13–38.

Caramazza, A. and J. R. Shelton (1998). Domain-specific knowledge systems in the brain: the animate-inanimate distinction. *Journal of Cognitive Neuroscience 10*(1), 1–34.

Caramazza, A. and E. B. Zurif (1976). Dissociation of algorithmic and heuristic processes in language comprehension: evidence from aphasia. *Brain and Language 3*(4), 572–582.

Carey, S. and E. Bartlett (1978). Acquiring a single new word. *Papers and Reports on Child Language Development 15*, 17–29.

Carroll, E. and P. Garrard (2005). Knowledge of living, nonliving and "sensory quality" categories in semantic dementia. *Neurocase 11*(5), 338–351.

Carter, R. L., M. K. Hohenegger, and P. Satz (1982). Aphasia and speech organization in children. *Science 218*(4574), 797–799.

Caselli, C., P. Casadio, and E. Bates (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language 26*(1), 69–111.

Caselli, M. C., E. Bates, P. Casadio, J. Fenson, L. Fenson, L. Sanderl, and J. Weir (1995). A cross-linguistic study of early lexical development. *Cognitive Development 10*(2), 159–200.

Changizi, M. A., Q. Zhang, and S. Shimojo (2006). Bare skin, blood and the evolution of primate colour vision. *Biology Letters 2*(2), 217–221.

Chao, L. L., J. V. Haxby, and A. Martin (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience 2*(10), 913–919.

Chao, Y.-R. (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica 4*(4), 363–397. Reprinted in Readings in Linguistics, ed. Martin Joos, p38-54.

Choi, S. and A. Gopnik (1995). Early acquisition of verbs in Korean: a cross-linguistic study. *Journal of Child Language 22*(3), 497–529.

Chomsky, N. (1959). A review of B.F. Skinner's verbal behavior. *Language 35*(1), 26–58.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: M.I.T. Press.

Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: M.I.T. Press.

Chomsky, N. (2007). Of minds and language. *Biolinguistics 1*, 9–27.

Christiansen, M. H. and S. Kirby (Eds.) (2003a). *Language Evolution*. Studies in the Evolution of Language. Oxford: Oxford University Press.

Christiansen, M. H. and S. Kirby (2003b). Language evolution: the hardest problem in science? See Christiansen and Kirby (2003a), Chapter 1, pp. 1–15.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy 78*(2), 67–90.

Clark, E. V. (2003). *First Language Acquisition*. Cambridge: Cambridge University Press.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, S. A., T. Allard, W. M. Jenkins, and M. M. Merzenich (1988). Receptive fields in the body-surface map in adult cortex defined by temporally correlated inputs. *Nature 332*(6163), 444–445.

Corina, D. P., S. L. McBurney, C. Dodrill, K. Hinshaw, J. Brinkley, and G. Ojemann (1999). Functional roles of broca's area and SMG: Evidence from cortical stimulation mapping in a deaf signer. *NeuroImage 10*(5), 570–581.

Cree, G. S. and K. McRae (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General 132*(2), 163–201.

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Croft, W. and D. A. Cruse (2004). *Cognitive Linguistics*. Cambridge textbooks in linguistics. Cambridge: Cambridge University Press.

Damasio, H., D. Tranel, T. Grabowski, R. Adolphs, and A. Damasio (2004). Neural systems behind word and concept retrieval. *Cognition 92*(1-2), 179–229.

Deacon, T. W. (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*. New York: W.W. Norton.

DeCasper, A. J., J.-P. Lecanuet, M.-C. Busnel, C. Granier-Deferre, and R. Maugeais (1994). Fetal reactions to recurrent maternal speech. *Infant Behavior and Development 17*(2), 159–164.

Dehaene-Lambertz, G. and S. Dehaene (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature 370*, 292 – 295.

Dessalles, J.-L. (2007). *Why We Talk*. Oxford: Oxford University Press.

Devlin, J. T., L. M. Gonnerman, E. S. Andersen, and M. S. Seidenberg (1998). Category-specific semantic deficits in focal and widespread brain damage: a computational account. *Journal of Cognitive Neuroscience 10*(1), 77–94.

Devlin, J. T., R. P. Russell, M. H. Davis, C. J. Price, H. E. Moss, M. J. Fadili, and L. K. Tyler (2002). Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia 40*(1), 54–75.

Diesendruck, G., L. Markson, and P. Bloom (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science 14*(2), 164–168.

Dixon, M. J., G. Desmarais, C. Gojmerac, T. A. Schweizer, and D. N. Bub (2002). The role of premorbid expertise on object identification in a patient with category-specific visual agnosia. *Cognitive Neuropsychology 19*(5), 401–419.

Dixon, R. M. W. (1994). *Ergativity*, Volume 69 of *Cambridge Studies in Linguistics*. Cambridge: Cambridge University Press.

Dixson, A. F. (1998). *Primate Sexuality*. Oxford: Oxford University Press.

Dowman, M. (2003). Explaining colour term typology as the product of cultural evolution using a bayesian multi-agent model. In R. Alterman and D. Kirsh (Eds.), *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, Mahwah, N.J. Lawrence Erlbaum Associates.

Dowman, M. (2007). Explaining color term typology with an evolutionary model. *Cognitive Science 31*(1), 99–132.

Drachman, D. A. (2005). Do we have brain to spare? *Neurology 64*(12), 2004–2005.

Dronkers, N. F., D. P. Wilkins, R. D. Van Valin, B. B. Redfern, and J. J. Jaeger (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition 92*(1-2), 145–177.

Emanatian, M. (1992). Chagga 'come' and 'go': metaphor and the development of tense-aspect. *Studies in Language 16*(1), 1–33.

Evans, V. and M. Green (2006). *Cognitive Linguistics: an Introduction*. Edinburgh: Edinburgh University Press.

Everett, D. L. (2005). Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology 46*(4), 621–646.

Farah, M. J. and J. L. McClelland (1991). A computational model of semantic memory impairment: modality specificity and emergent category specificity. *Journal of Experimental Psychology: General 120*(4), 339–357.

Feldman, J. and S. Narayanan (2004). Embodied meaning in a neural theory of language. *Brain and Language 89*(2), 385–392.

Fitch, W. T. (1994). *Vocal Tract Length Perception and the Evolution of Language*. Ph. D. thesis, Brown University, Providence, RI.

Fitch, W. T. (2007). Evolving meaning: the roles of kin selection, allomothering and paternal care in language evolution. In C. Lyon, C. Nehaniv, and A. Cangelosi (Eds.), *Emergence of Communication and Language*, Chapter 2, pp. 29–51. London: Springer Verlag.

Fodor, J. A. (1974). Special sciences (or: the disunity of science as a working hypothesis). *Synthese 28*(2), 97–115.

Fodor, J. A. and J. J. Katz (1971). What's wrong with the philosophy of language? In C. Lyas (Ed.), *Philosophy and Linguistics*. London: Macmillan and St Martin's Press.

Forde, E. M. E. and G. W. Humphreys (1999). Category specific recognition impairments: a review of important case studies and influential theories. *Aphasiology 13*(3), 169–194.

Franklin, A. and I. R. L. Davies (2004). New evidence for infant colour categories. *British Journal of Developmental Psychology 22*(3), 349–378.

Funnell, E. and P. De Mornay-Davies (1996). JBR: A reassessment of concept familiarity and a category-specific disorder for living things. *Neurocase 2*, 461–474.

Gallese, V. and G. Lakoff (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology 22*(3–4), 455–479.

Gentner, D. (1978). On relational meaning: the acquisition of verb meaning. *Child Development 49*(4), 988–998.

Gentner, D. (1982). Why are nouns learned before verbs: linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language, Thought and Culture*, Volume 2 of *Language Development*, pp. 301–334. Hillsdale, NJ: Lawrence Erlbaum.

Gentner, D. and L. Boroditsky (2001). Individuation, relativity and early word learning. In M. Bowerman and S. C. Levinson (Eds.), *Language Acquisition and Conceptual Development*, Chapter 8, pp. 215–256. Cambridge: Cambridge University Press.

Gershkoff-Stowe, L. and L. B. Smith (2004). Shape and the first hundred nouns. *Child Development 75*(4), 1098–1114.

Gibson, J. J. (1977). The theory of affordances. See Shaw and Bransford (1977), Chapter 3, pp. 67–82.

Gillette, J., H. Gleitman, L. Gleitman, and A. Lederer (1999). Human simulations of vocabulary learning. *Cognition 73*(2), 135–176.

Goldsmith, T. H. (1990). Optimization, constraint, and history in the evolution of eyes. *The Quarterly Review of Biology 65*(3), 281–322.

Goossens, L. (2000). Patterns of meaning extension, "parallel chaining", subjectification, and modal shifts. In A. Barcelona (Ed.), *Metaphor and Metonymy at the Crossroads: a Cognitive Perspective*, pp. 149–169. Berlin: Mouton de Gruyter.

Greene, J. D. W. (2005). Apraxia, agnosias, and higher visual function abnormalities. *Journal of Neurology, Neurosurgery, and Psychiatry 76*(supplement 5), pp. v25–v34.

Grice, H. P. (1957). Meaning. *The Philosophical Review 66*(3), 377–388.

Gross, M. (1979). On the failure of generative grammar. *Language 55*(4), 859–885.

Harms, W. F. (2004). Primitive content, translation, and the emergence of meaning in animal communication. In D. K. Oller and U. Griebel (Eds.), *Evolution of Communication Systems: A Comparative Approach*, pp. 31–48. Cambridge, MA: M.I.T. Press.

Harris, M., M. Barrett, D. Jones, and S. Brookes (1988). Linguistic input and early word meaning. *Journal of Child Language 15*, 77–94.

Harris, R. (1981). *The Language Myth*. London: Duckworth.

Harris, R. (1996). *Signs, Language and Communication*. London: Routledge.

Harris, R. (1998a). The integrationist critique of orthodox linguistics. See Harris and Wolf (1998), Chapter 2, pp. 15–26.

Harris, R. (1998b). Three models of signification. See Harris and Wolf (1998), Chapter 8, pp. 113–125.

Harris, R. and G. Wolf (Eds.) (1998). *Integrational Linguistics: A First Reader*, Volume 18 of *Language & Communication Library*. Oxford: Pergamon.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics 42*, 25–70.

Haspelmath, M. (2007). Pre-established categories don't exist: consequences for language description and typology. *Linguistic Typology 11*(1), 119–133.

Hauk, O., I. Johnsrude, and F. Pulvermüller (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron 41*(2), 301–307.

Hauser, M. D., N. Chomsky, and W. T. Fitch (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science 298*, 1569–1579.

Hawkey, D. J. C. (2008). Do individual preferences determine case marking systems. In A. D. M. Smith, K. Smith, and R. Ferrer i Cancho (Eds.), *The evolution of language: proceedings of the 7th international conference (EVOLANG7)*, pp. 147–154. World Scientific.

Hebb, D. O. (1949). *The Organization of Behavior: a Neuropsychological Theory*. New York: Wiley.

Heider, E. R. (1971). "Focal" color areas and the development of color names. *Developmental Psychology 4*(3), 447–455.

Hill, R. A. and R. A. Barton (2005). Red enhances human performance in contests. *Nature 435*(7040), 293.

Holzman, M. (1972). The use of interrogative forms in the verbal interaction of three mothers and their children. *Journal of Psycholinguistic Research 1*(4), 311–336.

Hopper, P. (1987). Emergent grammar. *Berkeley Linguistics Conference (BLS) 13*, 139–157.

Hopper, P. J. and E. C. Traugott (2003). *Grammaticalization* (2$^{\text{nd}}$ ed.). Cambridge textbooks in linguistics. Cambridge: Cambridge University Press.

Horgan, T. and J. Woodward (1985). Folk psychology is here to stay. *The Philosophical Review 94*(2), 197–226.

Hovers, E., S. Ilani, O. Bar-Yosef, and B. Vandermeersch (2003). An early case of color symbolism: ochre use by modern humans in Qafzeh cave. *Current Anthropology 44*(4), 491–522.

Hughlings-Jackson, J. (1878). On affections of speech from disease of the brain. *Brain 1*(3), 304–330.

Humphreys, G. W. and E. M. E. Forde (2001). Hierarchies, similarity, and interactivity in object recognition: "category-specific" neuropsychological deficits. *Behavioral and Brain Sciences 24*, 453–476.

Hurford, J. R. (2003). The neural basis of predicate-argument structure. *Behavioral and Brain Sciences 26*, 261–316.

Hurford, J. R. (2007). *The Origins of Meaning*, Volume 1 of Language in the Light of Evolution. Oxford: Oxford University Press.

Huttenlocher, J. and P. Smiley (1987). Early word meanings: The case of object names. *Cognitive Psychology 19*(1), 63–89.

Ibarretxe-Antuñano, B. I. (1999). *Polysemy and metaphor in perception: a cross-linguistic study*. PhD thesis, Edinburgh University.

Imai, M. and D. Gentner (1997). A cross-linguistic study of early word meaning: universal ontology and linguistic influence. *Cognition 62*(2), 169–200.

Ishai, A., L. G. Ungerleider, A. Martin, J. L. Schouten, and J. V. Haxby (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences 96*(16), 9379–9384.

Jackendoff, R. (2002). *Foundations of Language*. Oxford University Press.

Jones, S. S. and L. B. Smith (2002). How children know the relevant properties for generalizing object names. *Developmental Science 5*(2), 219–233.

Jones, S. S., L. B. Smith, and B. Landau (1991). Object properties and knowledge in early lexical learning. *Child Development 62*(3), 499–516.

Joseph, J. E. (2000). *Limiting the Arbitrary: Linguistic Naturalism and its Opposites in Plato's Cratylus and Modern Theories of Language*. Amsterdam studies in the theory and history of linguistic science. Amsterdam: John Benjamins.

Kaiser, P. K. (1984). Physiological response to color: A critical review. *Color Research & Application 9*(1), 29–36.

Kay, P. and L. Maffi (1999). Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist 101*(4), 743–760.

Kay, P. and C. K. McDaniel (1978). The linguistic significance of the meanings of basic color terms. *Language 54*(3), 610 EP – 646.

Kay, P. and T. Regier (2003). Resolving the question of color naming universals. *PNAS 100*(15), 9085–9089.

Kemmerer, D. (2006). Action verbs, argument structure constructions, and the mirror neuron system. In M. A. Arbib (Ed.), *From Action to Language via the Mirror Neuron System*, Chapter 10, pp. 347–373. Cambridge University Press.

Kenny, A. J. P. (2004). *Ancient Philosophy*, Volume 1 of *A New History of Western Philosophy*. Oxford: Oxford University Press.

Kirby, S. (1999). *Function, Selection and Innateness*. Oxford: Oxford University Press.

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. See Briscoe (2002), Chapter 6, pp. 173–204.

Kirby, S., M. Dowman, and T. L. Griffiths (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences 104*(12), 5241–5245.

Kirby, S. and J. Hurford (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi and D. Parisi (Eds.), *Simulating the Evolution of Language*, Chapter 6, pp. 121–148. London: Springer Verlag.

Knight, C., M. Studdert-Kennedy, and J. R. Hurford (Eds.) (2000). *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge: Cambridge University Press.

Kobayashi, H. and S. Kohshima (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *Journal of Human Evolution 40*(5), 419–435.

Lakoff, G. and M. Johnson (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, G. and M. Johnson (1999). *Philosophy in the Flesh*. New York: Basic Books.

Lambon Ralph, M. A., K. Patterson, P. Garrard, and J. R. Hodges (2003). Semantic dementia with category specificity: a comparative case-series study. *Cognitive Neuropsychology 20*(3), 307–326.

Landau, B., L. B. Smith, and S. S. Jones (1988). The importance of shape in early lexical learning. *Cognitive Development 3*(3), 299–321.

Laws, K. R. (2005). "Illusions of normality": a methodological critique of category-specific naming. *Cortex 41*, 842–851.

Laws, K. R., T. M. Gale, V. C. Leeson, and J. R. Crawford (2005). When is category specific in Alzheimer's disease? *Cortex 41*, 452–463.

Laws, K. R. and G. Sartori (2005). Category deficits and paradoxical dissociations in Alzheimer's disease and herpes simplex encephalitis. *The Journal of Cognitive Neuroscience 17*(9), 1453–1459.

Leavens, D. A., W. D. Hopkins, and K. A. Bard (2005). Understanding the point of chimpanzee pointing: epigenesis and ecological validity. *Current Directions in Psychological Science 14*(4), 185–189.

Levelt, W. J. M., P. Praamstra, A. S. Meyer, P. Helenius, and R. Salmelin (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience 10*(5), 553.

Levy, E. and K. Nelson (1994). Words in discourse: a dialectical approach to the acquisition of meaning and use. *Journal of Child Language 21*, 367–389.

Lichtheim, L. (1885). On aphasia. *Brain 7*, 433–484.

Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Galaway and B. J. Richards (Eds.), *Input and Interaction in Language Acquisition*. Cambridge: Cambridge University Press.

Lindsey, D. T. and A. M. Brown (2006). Universality of color names. *Proceedings of the National Academy of Sciences 103*(44), 16608–16613.

Lucy, J. A. (1997). The linguistics of color. In C. L. Hardin and L. Maffi (Eds.), *Color Categories in Thought and Language*. Cambridge: Cambridge University Press.

Mace, W. M. (1977). James J. Gibson's strategy for perceiving: ask not what's inside your head, but what your head's inside of. See Shaw and Bransford (1977), Chapter 2, pp. 43–65.

Mahon, B. Z. and A. Caramazza (2003). Constraining questions about the organisation and representation of conceptual knowledge. *Cognitive Neuropsychology 20*(3), 433–450.

Malt, B. C., S. A. Sloman, S. Gennari, M. Shi, and Y. Wang (1999). Knowing versus naming: similarity and the linguistic categorization of artifacts. *Journal of Memory and Language 40*(2), 230–262.

Markman, E. M. and J. E. Hutchinson (1984). Children's sensitivity to constraints on word meaning: taxonomic versus thematic relations. *Cognitive Psychology 16*(1), 1–27.

Markson, L. and P. Bloom (1997). Evidence against a dedicated system for word learning in children. *Nature 385*(6619), 813–815.

Martin, A. and L. L. Chao (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology 11*(2), 194–201.

Martin, A., C. L. Wiggs, L. G. Ungerleider, and J. V. Haxby (1996). Neural correlates of category specific knowledge. *Nature 379*, 649–652.

Masataka, N. (1993). Effects of contingent and noncontingent maternal stimulation on the vocal behaviour of three- to four-month-old Japanese infants. *Journal of Child Language 20*, 303–312.

Masataka, N. (1995). The relation between index-finger extension and the acoustic quality of cooing in three-month-old infants. *Journal of Child Language 22*, 247–257.

Masataka, N. (2003). *The Onset of Language*. Cambridge studies in cognitive and perceptual development. Cambridge: Cambridge University Press.

Masataka, N. and K. Bloom (1994). Acoustic properties that determine adults' preferences for 3-month-old infant vocalizations. *Infant Behavior and Development 17*(4), 461–464.

Masur, E. F. (1997). Maternal labelling of novel and familiar objects: implications for children's development of lexical constraints. *Journal of Child Language 24*(02), 427–439.

McGinn, M. (1997). *Wittgenstein and the Philosophical Investigations*. Routledge Philosophy Guidebooks. London: Routledge.

McGurk, H. and J. MacDonald (1976). Hearing lips and seeing voices. *Nature 264*(5588), 746–748.

Mervis, C. B., J. Catlin, and E. Rosch (1975). Development of the structure of color categories. *Developmental Psychology 11*(1), 54–60.

Narayanan, S. S. (1997). *Karma: knowledge-based active representations for metaphor and aspect*. Ph. D. thesis, University of California, Berkeley, California. (http://www.icsi.berkeley.edu/˜snarayan/thesis.pdf).

Nathans, J. (1999). The evolution and physiology of human color vision: Insights from molecular genetic studies of visual pigments. *Neuron 24*, 299–312.

Nielsen, J. M. (1946). *Agnosia, Apraxia, Aphasia: Their Value in Cerebral Localization* (2$^{\text{nd}}$ ed.). New York: Hoeber.

Ninio, A. and J. S. Bruner (1978). The achievement and antecedents of labelling. *Journal of Child Language 5*(1), 1–15.

Núñez, R. E. and E. Sweetser (2006). With the future behind them: convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science 30*(1), 1–49.

Oller, D. K., R. E. Eilers, and D. Basinger (2001). Intuitive identification of infant vocal sounds by parents. *Developmental Science 4*(1), 49–60.

Oller, D. K., L. A. Wieman, W. J. Doyle, and C. Ross (1976). Infant babbling and speech. *Journal of Child Language 3*(1), 1–11.

Özgen, E. (2004). Language, learning, and color perception. *Current Directions in Psychological Science 13*(3), 95–98.

Parr, L. A., B. M. Waller, and S. J. Vick (2007). New developments in understanding emotional facial signals in chimpanzees. *Current Directions in Psychological Science 16*(3), 117–122.

Peichl, L., H. Künzle, and P. Vogel (2000). Photoreceptor types and distributions in the retinae of insectivores. *Visual Neuroscience 17*(6), 937–948.

Penke, M. and G. Westermann (2006). Broca's area and inflectional morphology: Evidence from Broca's aphasia and computer modeling. *Cortex 42*, 563–576.

Petitto, L. A., R. J. Zatorre, K. Gauna, E. J. Nikelski, D. Dostie, and A. C. Evans (2000). Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. *Proceedings of the National Academy of Sciences 97*(25), 13961–13966.

Pinker, S. and P. Bloom (1990). Natural language and natural selection. *Behavioral and Brain Sciences 13*(4), 707–784.

Pinker, S. and M. T. Ullman (2002). The past and future of the past tense. *Trends in Cognitive Sciences 6*(11), 456–463.

Poizner, H., U. Bellugi, and E. S. Klima (1990). Biological foundations of language: clues from sign language. *Annual Review of Neuroscience 13*, 283–307.

Pulvermüller, F. (2002). *The Neuroscience of Language: on Brain Circuits of Words and Serial Order*. Cambridge: Cambridge University Press.

Pulvermüller, F., Y. Shtyrov, and R. Ilmoniemi (2005). Brain signatures of meaning access in action word recognition. *The Journal of Cognitive Neuroscience 17*(6), 884–892.

Pyers, J. E. (2006). Constructing the social mind: Language and false-belief understanding. In N. J. Enfield and S. C. Levinson (Eds.), *Roots of Human Sociality: culture, Cognition and Interaction*, Chapter 7, pp. 207–228. New York: Berg.

Quine, W. V. O. (1960). Translation and meaning. In *Word and Object*, Chapter 2, pp. 26–79. M.I.T. Press.

Recanzone, G., C. Schreiner, and M. Merzenich (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience 13*(1), 87–103.

Regan, B. C., C. Julliot, B. Simmen, F. Viénot, P. Charles-Dominique, and J. D. Mollon (2001). Fruits, foliage and the evolution of primate colour vision. *Philosophical Transactions of the Royal Society of London B 356*(1407), 229–283.

Regier, T., P. Kay, and N. Khetarpal (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences 104*(4), 1436–1441.

Reiss, N. (1989). Speech act taxonomy, chimpanzee communication, and the evolutionary basis of language. In J. Wind, E. G. Pulleybank, E. DeGrolier, and B. H. Bichakjian (Eds.), *Studies in Language Origins*, Volume 1, pp. 283–304. Amsterdam: John Benjamins.

Roberson, D., I. Davies, and J. Davidoff (2000). Color categories are not universal: replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General 129*(3), 369–398.

Rogers, T. T. and D. C. Plaut (2002). Connectionist perspectives on category-specific deficits. In E. M. E. Forde and G. W. Humphreys (Eds.), *Category-Specificity in Brain and Mind*, Chapter 9, pp. 251–290. Hove, East Sussex: Psychology Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch and B. B. Lloyd (Eds.), *Cognition and Categorization*, Chapter 2, pp. 27–48. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rowe, C., J. M. Harris, and S. C. Roberts (2005). Seeing red? Putting sportswear in context. *Nature 435*(293), E10.

Russell, B. (1912). *The Problems of Philosophy*. London: Home University Library.

Sagona, A. (1994). The quest for the red gold. In A. Sagona (Ed.), *Bruising the Red Earth: Ochre Mining and Ritual in Aboriginal Tasmania*, pp. 8–38. Melbourne: Melbourne University Press.

Sakai, K. L., Y. Tatsuno, K. Suzuki, H. Kimura, and Y. Ichida (2005). Sign and speech: amodal commonality in left hemisphere dominance for comprehension of sentences. *Brain 128*, 1407–1417.

Samuelson, L. K. and L. B. Smith (1999). Early noun vocabularies: do ontology, category structure and syntax correspond? *Cognition 73*(1), 1–33.

Sandhofer, C. M. and L. B. Smith (2001). Why children learn color and size words so differently: evidence from adults' learning of artificial terms. *Journal of Experimental Psychology: General 130*(4), 600–620.

Sandhofer, C. M., L. B. Smith, and J. Luo (2000). Counting nouns and verbs in the input: differential frequencies, different kinds of learning? *Journal of Child Language 27*, 561–585.

Saunders, B. A. C. and J. van Brakel (1997). Are there nontrivial constraints on colour categorization? *Behavioral and Brain Sciences 20*, 167–228.

Saunders, B. A. C. and J. van Brakel (2002). The trajectory of color. *Perspectives on Science 10*(3), 302–355.

de Saussure, F. (1966). *Course in General Linguistics*. New York: McGraw-Hill Education.

Schieffelin, B. B. and E. Ochs (1986). Language acquisition and socialization: Three developmental stories and their implications. In R. A. Shweder and R. A. LeVine (Eds.), *Culture Theory: Essays on Mind, Self and Emotion*, Chapter 11, pp. 276–319. Cambridge: Cambridge University Press.

Searle, J. R. (1979). *Expression and meaning*. Cambridge: Cambridge University Press.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.

Sharpe, L. T., A. Stockman, W. Jagla, and H. Jägle (2005). A luminous efficiency function, v*(λ), for daylight adaptation. *Journal of Vision 5*(11), 948–968.

Sharrock, W. and J. Coulter (1998). On what we can see. *Theory & Psychology 8*(2), 147–164.

Shaw, R. and J. Bransford (Eds.) (1977). *Perceiving, Acting and Knowing: Toward an Ecological Perspective*. New Jersey: Lawrence Erlbaum Associates.

Skinner, B. F. (1957). *Verbal Behaviour*. New York: Appleton-Century-Crofts.

Smith, K. (2004). The evolution of vocabulary. *Journal of Theoretical Biology 228*(1), 127–142.

Smith, L. B., S. S. Jones, B. Landau, L. Gershkoff-Stowe, and L. Samuelson (2002). Object name learning provides on-the-job training for attention. *Psychological Science 13*(1), 13–19.

Soja, N. N. (1994). Young children's concept of color and its relation to the acquisition of color words. *Child Development 65*(3), 918–937.

Soja, N. N., S. Carey, and E. S. Spelke (1991). Ontological categories guide young children's inductions of word meaning: object terms and substance terms. *Cognition 38*(2), 179–211.

Sperber, D. and D. Wilson (1995). *Relevance: Communication and Cognition* (2^{nd} ed.). Blackwell.

Steels, L. and T. Belpaeme (2005). Coordinating perceptually grounded categories through language: a case study for colour. *Behavioral and Brain Sciences 28*, 469–529.

Steels, L. and F. Kaplan (2002). Bootstrapping grounded word semantics. See Briscoe (2002), Chapter 3, pp. 53–74.

Stockman, A. and L. T. Sharpe (2000). The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research 40*(13), 1711–1737.

Stockman, A., L. T. Sharpe, and C. Fach (1999). The spectral sensitivity of the human short-wavelength sensitive cones derived from thresholds and color matches. *Vision Research 39*(17), 2901–2927.

Studdert-Kennedy, M., A. M. Liberman, K. S. Harris, and F. S. Cooper (1970). Motor theory of speech perception: a reply to Lane's critical review. *Psychological Review 77*(3), 234–249.

Sumner, P. and J. D. Mollon (2000a). Catarrhine photopigments are optimised for detecting targets against a foliage background. *Journal of Experimental Biology 203*, 1963–1986.

Sumner, P. and J. D. Mollon (2000b). Chromaticity as a signal of ripeness in fruits taken by primates. *Journal of Experimental Biology 203*, 1987–2000.

Sumner, P. and J. D. Mollon (2003). Colors of primate pelage and skin: objective assessment of conspicuousness. *American Journal of Primatology 59*, 67–91.

Sweetser, E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science 12*(1), 49–100.

Tardif, T. (1996). Nouns are not always learned before verbs: evidence from Mandarin speakers' early vocabularies. *Developmental Psychology 32*(3), 492–504.

Tardif, T., S. A. Gelman, and F. Xu (1999). Putting the "noun bias" in context: a comparison of English and Mandarin. *Child Development 70*(3), 620–635.

Taylor, T. J. (1997). The origin of language: Why it never happened. In *Theorizing Language*, pp. 241–260. Pergamon.

Tinbergen, N. (1963). On aims and methods in ethology. *Zeitschrift für Tierpsychologie 20*, 410–433.

Tomasello, M. (2000). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2003). *Constructing a Language. A Usage-Based Theory of Language Acquisition*. Harvard University Press.

Tomasello, M. (in press). *Origins of Human Communication*. Cambridge, MA: M.I.T. Press.

Tomasello, M. and N. Akhtar (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development 10*, 201–224.

Tomasello, M., M. Carpenter, J. Call, T. Behne, and H. Moll (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences 28*, 675–735.

Torgerson, W. (1952). Multidimensional scaling 1: theory and method. *Psychometrika 17*(4), 401–419.

Traugott, E. C. (1989). On the rise of epistemic meanings in English: an example of subjectification in semantic change. *Language 65*(1), 31–55.

Traugott, E. C. and R. B. Dasher (2002). *Regularity in Semantic Change*. Cambridge studies in linguistics. Cambridge: Cambridge University Press.

Turner, V. W. (1966). Colour classification in Ndembu ritual: a problem in primitive classification. In M. Barton (Ed.), *Anthropological Approaches to the Study of Religion.*, pp. 47–84. London: Tavistock.

Twaddell, W. F. (1935). On defining the phoneme. *Language 11*(1), 5–62.

Ullman, M. T., S. Corkin, M. Coppola, G. Hickok, J. Growdon, W. Koroshetz, and S. Pinker (1997). A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience 9*(2), 266–276.

Vargha-Khadem, F., E. Isaacs, and V. Muter (1994). A review of cognitive outcome after unilateral lesions sustained during childhood. *Journal of Child Neurology 9*(Suppl 2), 2S67–2S73.

Vargha-Khadem, F., A. M. O'Gorman, and G. V. Watters (1985). Aphasia and handedness in relation to hemispheric side, age at injury and severity of cerebral lesion during childhood. *Brain 108*(3), 677–696.

Vick, S.-J., B. Waller, L. Parr, M. Smith Pasqualini, and K. Bard (2007). A cross-species comparison of facial morphology and movement in humans and chimpanzees using the facial action coding system (FACS). *Journal of Nonverbal Behavior 31*(1), 1–20.

Vogt, P. (2003). Iterated learning and grounding: From holistic to compositional languages. In S. Kirby (Ed.), *Proceedings of Language Evolution and Computation Workshop/Course at ESSLLI*, Vienna, pp. 76–86.

Vorobyev, M. (2004). Ecology and evolution of primate colour vision. *Clinical and Experimental Optometry 87*(4/5), 230–238.

Waismann, F. (1965). *The Principles of Linguistic Philosophy*. London: Macmillan and St Martin's Press.

Warner, A. (1993). *English Auxiliaries: Structure and History*. Cambridge: Cambridge University Press.

Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology 27*, 635–657.

Warrington, E. K. and R. A. McCarthy (1987). Categories of knowledge: further fractionations and an attempted integration. *Brain 110*(5), 1273–1296.

Warrington, E. K. and T. Shallice (1984). Category specific semantic impairments. *Brain 107*(3), 829–853.

Webster, M. A. and J. D. Mollon (1994). The influence of contrast adaptation on color appearance. *Vision Research 34*(15), 1993–2020.

Wernicke, C. (1874). *Der Aphasische Symptomencomplex. Eine Psychologische Studie auf Anatomischer Basis*. Breslau: Kohn und Weigert.

West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford: Oxford University Press.

Wilkins, W. K. and J. Wakefield (1995). Brain evolution and neurolinguistic preconditions. *Behavioral and Brain Sciences 18*, 161–182.

Wittgenstein, L. (1967). *Philosophical Investigations* (3rd ed.). Oxford: Blackwell.

Wittgenstein, L. (1969). *The Blue and Brown Books* (2nd ed.). Oxford: Blackwell.

Wittgenstein, L. (1974). *Philosophical Grammar*. Oxford: Blackwell.

Wittgenstein, L. (1975). *On Certainty*. Blackwell.

Wittgenstein, L. (1977). *Remarks on Colour*. Oxford: Blackwell.

Wittgenstein, L. (1978). *Remarks on the Foundations of Mathematics* (3rd ed.). Oxford: Blackwell.

Wittgenstein, L. (1981). *Zettel* (2nd ed.). Oxford: Blackwell.

Woods, B. T. (1983). Is the left hemisphere specialized for language at birth? *Trends in Neurosciences 6*, 115–117.

Woods, B. T. and H.-L. Teuber (1978). Changing patterns of childhood aphasia. *Annals of Neurology 3*(3), 273–280.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language & Communication 18*(1), 47–67.

Wray, A. (2000). Holistic utterances in protolanguage: the link from primates to humans. See Knight, Studdert-Kennedy, and Hurford (2000), Chapter 17, pp. 285–302.

Wray, A. (2002a). Dual processing in protolanguage: performance without competence. See Wray (2002b), Chapter 6, pp. 113–137.

Wray, A. (Ed.) (2002b). *The Transition to Language*. Studies in the Evolution of Language. Oxford: Oxford University Press.

Wray, A. and G. W. Grace (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua 117*(3), 543–578.

Wyszecki, G. and W. S. Stiles (1982). *Color science: concepts and methods, quantitative data and formulae* (2nd ed.). New York: Wiley.

Yoshida, H. and L. B. Smith (2003). Shifting ontological boundaries: how Japanese- and English-speaking children generalize names for animals and artifacts. *Developmental Science 6*(1), 1–18.

Yoshida, H. and L. B. Smith (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science 16*(2), 90–96.

Zahn, R., P. Garrard, J. Talazko, M. Gondan, P. Bubrowski, F. Juengling, H. Slawik, P. Dykierek, B. Koester, and M. Hull (2006). Patterns of regional brain hypometabolism associated with knowledge of semantic features and categories in Alzheimer's disease. *The Journal of Cognitive Neuroscience 18*(12), 2138–2151.

Zeki, S. (1983). Colour coding in the cerebral cortex: the reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience 9*(4), 741–765.

Zuberbühler, K., R. Noë, and R. M. Seyfarth (1997). Diana monkey long-distance calls: messages for conspecifics and predators. *Animal Behaviour 53*(3), 589–604.