DOI: 10.1111/phpr.12896

Philosophy and Phenomenological Research

ORIGINAL ARTICLE

Reference and morphology*

Gabe Dupre

Keele University

Correspondence
Gabe Dupre, Keele University.
Email: G.g.dupre@keele.ac.uk

Abstract

The dominant tradition in analytic philosophy of language views reference as paradigmatically enabled by the acquisition of words from other speakers. Via chains of transmission, these words connect the referrer to the referent. Such a picture assumes the notion of a word as a stable mapping between sound and meaning. Utterances are constructed out of such stable mappings. While this picture of language is both intuitive and historically distinguished, various trends and programs that have developed over the last few decades in theoretical linguistics suggest an alternative. According to these approaches, the word, conceived of as a linguistic 'building block', has no special theoretical significance. While natural language systems generate structures mapping sounds onto meanings, they need not do so by composing elements which themselves specify such mappings. I shall describe some of these developments, and show how they pose a problem for traditional philosophical views of language and communication, before identifying an alternative approach to reference which does not rely on this common-sense picture of words.

*For early and insightful discussion of the ideas in this paper, I am grateful to Gabbrielle Johnson, Kevin Lande, and Jessie Munton. Gabby's written comments on a more complete draft, and those of John Dupré and an anonymous reviewer for this journal, also substantially improved the final product. Funding for this research was provided by the Leverhulme Trust.

1 | INTRODUCTION

Observations due to Kripke, Donnellan, Barcan Marcus, and others in the 1960s and 1970s drew attention to a fundamental distinction between referring to an entity and describing an entity. That someone was, in making a particular utterance, talking *about* some particular was shown to be compatible with (almost?) total ignorance concerning the traits or qualities of this particular. The familiar puzzle that this observation raised is: how is it possible for someone (say, me, in England in 2021) to pick out an individual arbitrarily distant temporally and spatially[1] (say, Elpinice, in 5th Century B.C.E. Greece) without being able to differentiate this individual from indefinitely many others (say, the numerous other noblewomen of the period)? The standard answer (see e.g. (Kripke, 1980), (Kaplan, 1989a,b)) is: by acquiring the word 'Elpinice'. This word was passed on to me by someone who themselves was able to use it to pick out Elpinice. And by adopting their word, I acquire their competency to use it in this way. Referential capacities are thus transmitted from one speaker to another. Tracing the entirety of this lineage, however, eventually leads to a usage of the expression whose referential properties are explained differently (e.g. the speaker is looking at Elpinice herself). So, while there must be some fact or facts which ensures that it is Elpinice, rather than anyone else, who is the referent of this expression, this constraint applies only to the introduction of the expression itself, not to its subsequent uses. These latter are able to simply piggyback on the referential capacities of the historical examples on which they are modeled. Such, in broad outline, is the philosophical orthodoxy.

As Kaplan (1990), one of the primary architects of this picture of reference, notes, this is not an ontologically neutral picture. Explaining reference in this way places substantial constraints on philosophical accounts of words and of their individuation and subsistence. His *continuant* model of words is designed precisely to answer the question: what must words be, if they are to play the role required of them by pictures of reference of this sort? Kaplan argued that words themselves should be viewed as constituted by these temporally extended chains of transmission, and thus that two utterances of words should be viewed as utterances of the same word (and thus referring to the same entity) only if they are part of the same lineage.

This style of argument, inferring from the observation that a philosophical picture requires a particular ontology to the adoption of just that ontology, is, I believe, a perfectly reasonable one. There is, however, another strategy for investigating the ontology of our world. Namely, empirical science. And there is much work in the empirical science of linguistics aimed precisely at discovering what words are and what kinds of properties they have, specifically within Morphology and its interfaces with other branches of linguistics. In this paper, I shall discuss some of this work, arguing that it makes very significant trouble for the picture of words described in detail by Kaplan and adopted widely within the philosophy of language.

The problem, I shall argue, is that several distinct but related research programs within generative linguistics (e.g. Distributed Morphology (Marantz, 1997); (Halle & Marantz, 1993, 1994); (Harley & Noyer, 2014), Exo-Skeletal Grammar (Borer, 2005a,b, 2013), and Nanosyntax (Baunaz et al., 2018)) suggest that theoretical linguistics does not make essential use of the notion 'word'. That is, theoretically significant processes and explanations do not rely on differentiating words from other kinds of linguistic structures, such as morphemes or phrases. As Julien (2007) puts it, "from the point of view of grammar, the word is merely an epiphenomenon." (p. 213). If this is correct, it forces us to re-evaluate philosophical accounts of reference which rely on an ontology of words.

[1] And, apparently, modally. Although this raised further worries.

The structure of this paper is as follows. I begin by spelling out, in more detail, the picture of natural language and natural language expressions such as words that has become orthodoxy within analytic philosophy of language. For reasons that will become clear, I shall call this picture the 'manifest image of language'. For my purposes, the central tenets of this picture are that sentences are composed of words, and that words are stable pairings of sound and meaning. After spelling this out, I shall outline the above-mentioned linguistic theories, identifying some empirical motivations as well as their core architectural assumptions, and show how they undermine the manifest image of language. To close, I shall attempt to show how, even without reliance on words and their transmission, the key insights of the referentialist tradition can be retained. The primary goal of the paper will be to establish the conditional claim: *if* these converging scientific traditions are on the right track, *then* some commonplace assumptions in analytic philosophy need rethinking. But a larger, secondary, goal is methodological: to exemplify the often surprising ways that philosophical theorizing can be dependent on specific and controversial results in the sciences, and thus to show that reliance on the manifest image generally can lead to mistaken theorizing.

2 | THE MANIFEST IMAGE OF LANGUAGE

One standard conception of the goals and strategies of philosophy, perhaps best exemplified in Strawson (1959), views philosophy as aimed at developing our everyday conception of the world so as to remove any unclarities or inconsistencies within it. The goal is to tell us what the world *as we experience it* is like, deviating from common-sense only when necessary to avoid internal contradiction. Sellars (1963) calls the result of such theorizing the 'manifest image'. I believe the philosophical picture of language largely adopted in analytic philosophy of language is best viewed as an instance of this style of theorizing, and so will call this picture 'the manifest image of language'.

Here are some seemingly "common-sense" observations about natural language, which I take to constitute the manifest image of language[2]:

- A language is a system of communication, shared by multiple speakers.
- A language prescribes the mapping of publicly observable signals (e.g. ink marks, sound waves, manual gestures) onto meanings.
- This mapping divides into basic elements (words) for which the mapping is stipulated, and complex expressions (e.g. sentences), for which the mapping is derived from that of the basic elements and combinatory rules.
- Complex utterances can thus be decomposed into more-or-less discrete simple constituents, i.e. words.
- The stipulated mappings associated with a particular word are learned from other competent users of the language.

[2] As will be discussed later, I do not view these observations as 'common sense' in the sense that they are universal, on the grounds that many of them apply much better to languages like English than to many other natural languages. Rather, I view them as highly intuitive aspects of the experience and understanding of language for someone who speaks a language like English.

While I will not here argue for this I believe that this collection of claims fairly accurately, although of course vaguely and incompletely, corresponds to a widespread view of what natural languages are (at least among English speakers). More to the point, it seems to be the picture adopted by most work in analytic philosophy of language.[3]

As noted above, to account for our ability to refer to people spatially and temporally distant from us, despite lacking sufficient knowledge to uniquely identify them, Kripke, Kaplan, and others appeal to the claim that acquisition of a word (paradigmatically for them, a proper name) involves the acquisition of the referential capacities made possible by this word. In this way, referential abilities can be transmitted from those who do have the (independently available) capacity to uniquely identify someone, say by perception, to those who otherwise would not. Chains of lexical transmission, combined with policies of deference (I intend to use the word in the way that the person from whom I acquired it did), massively expand our referential capacities by allowing my reference to depend on that of all those with whom I have linguistically interacted.

The Kripke-Kaplan proposal, I believe, owes much to the manifest image of language. Crucially, words, according to this picture, are shared, or at least sharable, items which enable their possessors to refer to individuals by producing a public token. As reference is identified with (or at least viewed as a central component of) meaning, this tradition takes words to provide mappings between sounds (or shapes, gestures, etc.) and meanings, and these mappings are acquired from other language users, just as in the manifest image of language.

Note that Kaplan relies on this picture of words as mappings even though he, along with many who follow him, is quite liberal about whether two distinct sound-meaning mappings can count as instances of the same word. Kaplan allows, for example, that he and Tom Brokaw are using the same word when he says '*million*' and Brokaw says '*mi-we-un*', on the grounds that these words share meanings and have suitably connected histories. One can accept, that is, that words are mappings between sounds and meanings, while denying that for two utterances to be utterances of the same words requires that they enact the same sound-meaning mapping. The former is a question of what words *are*, the latter is a question of word *individuation*. What matters for this picture, then, is not that the mapping that a word enacts between sound and meaning is stable over time or between speakers. As Kaplan stresses, we are typically perfectly able to view two utterances as utterances of the same word despite significant changes in their properties *qua* public signal.[4] What matters is that, in the process of constructing or interpreting an utterance, we make essential explanatory appeal to a specific signal-meaning mapping. E.g. in explaining why "the snakes are on the plane" has the properties it does, we might appeal to a *word* which maps /sneɪks/, in the dialect of the current speaker, to the set of snakes. It is the ability to convey this worldly meaning that is the central explanandum of these causal-chain theorists, and it is the sharing of words (however individuated) that serves as the central explanandum.

Although it will depend somewhat on the details of the view, it seems to be a corollary of the claim that referential capacities are transmitted from speaker to speaker that the semantic type and/or syntactic category of the words in question is likewise transmitted. If

[3] And indeed beyond analytic philosophy: it seems also to be reflected in the views of De Saussure (2011) and Aristotle (Modrak, 2001).

[4] In place of the somewhat clunky 'public signal', I will sometimes adopt the customary reference to 'sounds', simply for convenience, but it should be remembered that this expression is strictly inappropriate, as many languages are not produced for auditory consumption.

I acquire the word 'snakes' from you, and intend for my uses of it to refer to whatever your uses of it did, then it seems unavoidable that my word will be of the same semantic type as yours (i.e. a predicate) and thus of a syntactic category required of words with this type of meaning (e.g. a common noun).[5] Novel uses of expressions, as when 'snake' is used as a verb, are thus made to appear somewhat mysterious. This sympathy between syntax and semantics seems to be a strong prediction of this approach. Much more on this later.

This picture of words also relies on a strict demarcation between, on the one hand, words, which function as the elements over which linguistic processes are defined, and complex linguistic expressions like phrases and sentences. Kaplan (2011) is explicit here: "the basic elements of the language are earthly creations, but the compounds generated by syntactical rules … are structures -types- which may or may not have tokens." (p. 511). This is essential to this picture as, whatever the case with words, it is obvious that our capacity to use particular sentences, say, to mean what they do is not explained with reference to our having encountered other people using these very sentences in these ways. This is demonstrated by the *productivity* of natural language: we can meaningfully use indefinitely many sentences that we have never encountered before. The promise, then, of philosophical approaches like this, is that by encountering other people using various kinds of simple expressions (words) to pick out particular individuals, groups, processes, etc. in the world, we can thereby adopt these usages and pick out the same aspects of the world. But, crucially, the processes by which we combine these different referential capacities allow us to recombine them in novel ways, extending beyond those we have experience with. So, I can refer to snakes and planes in virtue of hearing you refer to snakes and planes with the words 'snakes' and 'planes'. But, to say that the snakes are on the plane, I need not have heard you put these words to this very use.

To summarize the story so far, our common-sense interpretation of what languages are and how they work appears to dovetail very nicely with our philosophical accounts of linguistic reference. These latter require that we be able to pick out entities with which we have no immediate experience, and of which we may have minimal knowledge. One way this could be achieved would be to imbue the words we use with the capacity to refer, in virtue of chains of transmission tracing back to a more substantive way of latching on to the referent. Complex linguistic behaviour then consists in combining such words with one another, and the sounds and meanings of complex linguistic expressions are a product of the sounds and meanings of these simple words. The manifest image of language thus fits the bill very nicely, in what descriptively-inclined philosophers might think was quite a nice result for theories of reference. Before turning to the empirical problems for this approach, it is important to make two quick clarifications about the nature of the words appealed to in these philosophical approaches.

Firstly, this story, I assume, purports to apply only to open-class, content words, and not to closed-class, functional expressions. While my ability to talk about Elpinice, or snakes, can be explained by appeal to my having heard about such creatures from others, it is not clear how to account in the same way for my use with conjunctions (*and, but*), auxiliaries (*will, must, should*), various forms of agreement marking, etc. For one thing, the former capacities are unevenly distributed among the speaking population (not everyone can refer to Elpinice, or even to snakes), while the latter are more-or-less species universal: if you can speak, you can distinguish the present from the past, can combine multiple expressions of the same type to create more complex expressions of that type, and so on. This suggests that the latter may be best appealed to

[5] See Mercier (1998) for discussion.

as aspects of a species-specific psychological capacity for language itself, rather than as expressions acquired on the basis of interactions with other speakers (although of course one must learn which particular sounds are used). For another, the referential model, while extending relatively naturally from individual reference to reference to collections or events[6], seems intrinsically unsuited for application to functional expressions of this sort: whatever meaning 'and' or 'will' have, it doesn't seem to sit at the end of a causal chain. Finally, as has long been noted in philosophy and linguistics, while the meaning of a complex expression may, in some sense, be determined by the meanings of its simple parts, it cannot simply consist in the set of such meanings. The meaning of "Anatoli and Sahir" may involve Anatoli and Sahir, but it is definitely not the set consisting of Anatoli, Sahir, and the referent of 'and'. In addition to expressions which pick out aspects of the world, we require some expressions to provide instructions for what to do with these referents. The traditional distinction between the open-, and closed-, class lexicon seems to track this division reasonably well, with the former enabling connection with worldly entities, and the latter structuring these entities in linguistically specific ways. So, when I talk about words being sound-meaning mappings, and being transmitted from one speaker to another, it should be remembered that I am interested here only in the open-class lexicon.

Secondly, the literature I have been discussing takes as its paradigm case proper names, words used to refer to particular individuals ('Elpinice', 'Manchester', etc.). There is a legitimate question about how far the referentialist picture of language is supposed to extend from these core cases. It is widely accepted that at least some common nouns (natural kind terms such as 'snake' or 'hydrogenase') are given similar treatment. Similarly, Burge (1986) applies analogous arguments to our competencies with artifactual terms like 'sofa', and Kaplan (1989a) expresses the view that we can manage to refer to love, grief, and the colour red in virtue of acquiring suitable expressions, even without direct experience of these phenomena (p. 604). More generally, Soames (2002) and Salmon (2005) present pictures of language which view the meanings of content expressions as pretty generally a matter of latching onto the world. And indeed, this is what you should expect, given the structure of argumentation for referentialist views. As Burge (1986) argues, the central grounds for viewing an expression as referential (in this extended sense) is that one can be deeply mistaken about the kind of thing one is talking about. But it is a pretty rare expression which does not open one up to such potential confusion, and so if these arguments are good, we should expect them to have very general lessons for the nature of language. So, I believe it is fair to assess this picture of language with investigation of a wide variety of open-class expressions. However, to make my case as strong as possible, I will at each stage use examples of the paradigmatic referential expression, the proper name. I shall argue that it is at best an idealization with fairly narrow application, and at worst simply false, that natural language expressions are composed out of words constituting stable sound-meaning mappings, even for the case of proper names. I shall argue, in line with various recent developments in linguistic theory, that open-class expressions in general display significant internal structure, and this is reflected in phonological, syntactic, and semantic variability. I conclude that the manifest image of language, and the philosophical views which presuppose it, are mistaken.

[6] I will be using the term 'reference' in this extended sense, to cover all word-world connections, throughout. Readers who wish to distinguish reference to entities from satisfaction by predicates, and so on, should feel free to translate into their own preferred terminology.

3 | THE UNSTABLE LEXICON I: MORPHOPHONOLOGY

As Kripke puts it, reference to distal entities is enabled by the communicative interactions through which a "name is spread from link to link as if by a chain." (Kripke, 1980 p.171). To play this role as links in communicative chains, as required by philosophical orthodoxy, there must be some relatively stable entity, the word, which is passed from one speaker to another. As Kaplan (1990) notes, such chains do not consist solely in *inter*personal links, communicative acts, but also "more mysterious *intra*personal stages" (p. 98) during which the word in question is stored in the memory or mental lexicon of the speaker. The problem I wish to raise for this view is that attention to the details of morphological theory demonstrates that these chains are far from homogenous. On Kaplan and Kripke's picture, we hear a word, store that very same word in memory, and then pass it on to others, in whom the process repeats. This is precisely to assume the manifest image doctrine that words are the atoms of language, the unchanging units out of which complex expressions can be creatively generated. But linguistic morphology is precisely the study of the ways in which words are *not* atomic in this way, but are creatively generated out of more basic units, just as are phrases and sentences.

Both terminology and theory are hotly contested, but the distinction between two kinds of linguistic entity is presupposed by all morphological theorizing. There are, on the one hand, basic linguistic units 'morphemes', which must somehow be stored in the memory of the language user. And on the other, there are complex units composed out of the former type. Words are in this latter category. The pronunciation and meaning of the word will be a product then of the morphemes involved in its construction and the structures in which these morphemes are found. Whether these categories overlap (i.e. whether some words are morphemes) is debated, but that they are not equivalent (that some morphemes are not words, and some words are not morphemes) is undisputed. What will be central for our purposes is that words are what is pronounced. When a linguistic expression is transmitted from speaker to speaker, this expression is a word (or a larger expression composed of words). However, words *need not be* what is stored. When I hear a sentence, I need not retain this whole sentence as a unit for later use, but can instead simply store the component words, and reconstruct the sentence as necessary. Just so with words: when I hear a word, I may not store it as a unit, but can instead decompose it into its component morphemes, and recompose these at a later date. These processes of decomposition and recombination mean that the relationship between morpheme and word is many-to-many.

It is traditional (although controversial) to distinguish between inflectional and derivational morphology. Traditionally, inflection distinguishes between words with the same grammatical category and meaning as found in different grammatical contexts, while derivation marks variation which encodes changes to syntactic category or meaning. So, in English, alternation between *kick* and *kicks* (due to grammatical person), *kick* and *kicked* (due to tense), and so on is an inflectional process, while the difference between *inhabit*, *uninhabited*, *inhabitant*, *inhabitation*, and so on are derivational. Whatever the theoretical status of this distinction, these processes are widespread in natural language, and point to worries for any account of communication which relies on stable entities being passed from one speaker to another. The problem is that, while there are distinct *words* here, they are composed out of shared morphemes. And these morphemes are extractable upon hearing these words, and can be used in the generation of these words in future usage. The standard, Kripke-Kaplan, picture of communicative chains predicts that (for most users) referential success with a word depends on having heard someone use *that very word* in the past. But the productive processes of morphology show that this is not the case. By hearing

you use one member of a morphological paradigm, I can often extract the relevant morphemes, and use them to generate and then successfully use, the rest of the paradigm, even though I have never heard anyone use these additional words. It seems perfectly possible for me to hear you use the term 'uninhabited', and base my use of 'inhabited' on this usage. If someone then hears me, and bases their subsequent use of 'inhabitant' on my usage, there is no communicative chain of uses of a particular word to appeal to, for each member of the chain used different words, with different pronunciations. So, contrary to this standard picture, while what is produced and transmitted is a word, what is stored is a morpheme.

While there are a variety of distinct proposals for the metaphysics of words in the literature, fortunately my argument allows us to bypass these debates. On any plausible account of the nature of words, 'inhabit' and 'inhabitable' are distinct words. Most obviously, a theory of words which individuates words according to their linguistic (phonological, semantic, syntactic) properties, such as Miller's (2021) bundle theory of words, will immediately exclude them. But even more permissive accounts, such as Kaplan's, which allow two instances of the same word have quite different phonological properties, so long as the speakers are intending to produce instances of the same word (e.g. if one speaker intends to produce an instance of the same word that they heard the other produce) will, and indeed must, be compatible with the relevant judgements of identity and difference. When I produce 'inhabitant', I may be dependent on your production of 'inhabitable', but there is no sense in which I am even attempting to produce the same word. Whatever words are, these words are different. But nonetheless they seem to support exactly the kinds of communicative chains appealed to in standard stories about referential dependence. Therefore, such communicative chains need not consist of repeated productions of a particular word.

The natural proposal, at this stage, is to appeal to common features in each word which, in virtue of 'appearing' in each communicative interaction constituting the communicative chain, can play the role required of words in the standard picture. Such common elements are dubbed 'roots' in morphophonology.[7] And it is indeed (definitionally) true that every member of a derivationally or inflectionally derived set of words will share a root. But the problem is that roots are not words, i.e. are not what is produced in speech. Minimally, there are some roots which are never uttered without undergoing processes of derivation or inflection. Our example of 'inhabit' provides an example: in the relevant sense, 'habit' is not a word of (modern) English.[8] Certainly, most people who use the verb 'inhabit' and the adjectival participle 'inhabited' have never heard anyone use 'habit' as a verb. And so appeal to roots simply cannot play the role attributed to words by the manifest image of language, on the grounds that roots are not pronounced.

English can be misleading in this respect, as given its relatively impoverished inflectional system, it can be hard to distinguish roots from their appearance in words. But this is not the case for most languages. In romance languages like Spanish or Italian, many words are never pronounced without some form of inflection (e.g. Spanish *corro* vs. *corre* (I run vs. he runs) and *alumna* vs. *alumno* (female vs. male student)). And in some languages, such as Modern Hebrew and other semitic languages, the majority of roots are not even pronounceable, with roots specifying only the consonants in a word, with the vowels filled in by a finite stock of standard patterns. Further, for reasons we will see later in the discussion of syntax and semantics, there are strong reasons to

[7] I shall use the term 'root' in much the same way as Borer (2005a,b). A root is an open-class morpheme, associated with extra-linguistic conceptualization and perception. Roots contrast with functional morphemes, such as tense and aspect markers, determiners, and auxiliary verbs, which are closed-class and linguistically specific.

[8] Note that 'inhabit' cannot itself be a root, on pain of missing out on the relation between, on the one hand, 'inhabit' and 'habitable', and, on the other, between 'inhabit' and 'inscribe', 'invite', 'influx', etc.

think that even in English, roots are *never* able to play any role in a linguistic structure without the help of functional items and structure, undermining the idea that roots play the role of the 'atoms of language' required by the manifest image of language.

Another possibility for defending the standard, manifest, picture would be to reduce its scope. It was originally centrally motivated by a few key examples, such as proper names, and perhaps should be viewed as an account specifically of how expressions of this sort are used. While in general, language generates quite complex mappings between sounds and meanings, perhaps expressions like proper names work more simply. And indeed, in English proper names are even less prone to (phonologically detectable) morphological processes than other expressions.[9]

However, once again this reflects a contingent feature of English, not a fact about natural language in general. In Finnish, for example, proper names undergo inflection for number and case, and this can result in changes to the pronunciation of the stem.[10] Finnish phonologically distinguishes between short and long ('geminate') consonants. For example, 'korpi' means *wilderness*, while 'korppi' means *crow*.[11] So, word-stems are phonologically distinguished by the presence or absence of short or long consonants. And many traditional Finnish names do contain geminates (e.g. 'Antti', 'Hilkka'). A general phonological process in Finnish, consonant gradation, reduces plosive geminate consonants (/pp/, /tt/, and /kk/) to short consonants in certain contexts, particularly when the syllable containing the vowel immediately following the geminate consonant ends in a consonant. So 'pappi' (*priest*) is geminated, but when the possessive suffix '-n' is added, this geminate consonant is reduced: 'papin'. Notably, the same is true for proper names: 'Antti' becomes 'Antin' in the possessive, as 'Hilkka' becomes 'Hilkan' (*'Anttin' and *'Hilkkan' are deemed incorrect). Just like the consonants, the vowels in proper names can be changed by morphophonological processes: When the partitive plural suffix '-ja' is added to a noun ending in a vowel, this (now penultimate) vowel is rounded. The plural partitives for 'Antti' and 'Hilkka' are 'Antteja' and 'Hilkkoja'.

What these examples show is that even for proper names, there is, in general, no phonological property that must be shared among all words derived from the same root. Further, inflection and derivation do not merely involve adding material to repeated expressions, but manipulate the way that even the phonological core of the word sounds. Given these processes, what sense can be made of the idea that a core component of communication involves acquiring individual words from others and repeating them, with the intention of conveying what they conveyed? It seems clear that the ability to refer to a previously referred to individual *does not* require that I use the same word that I have heard used to refer to that individual. If you use the word 'Hilkkoja', knowledge of Finnish would enable me to calculate that you were referring to an individual called 'Hilkka', and I can subsequently use this name to refer to her.[12] But this is not explained by appealing to our shared usage of a word. It may be explained by appealing to our shared usage of a *root*,

[9] Such processes are not, however, absent entirely. Clear cases of derivational morphology involving names used in abnormal (e.g. verbal, adjectival) contexts will be discussed in the next section. Diminutives (James → Jim/Jimmy, William → Will/Bill/Billy, Jessica → Jess/Jessy, etc.) provide another clear case. Extreme cases of near-suppletion can be found as well, especially when the derived term is modeled on a translation of the original: followers of King James are called 'Jacobites', people from Liverpool are called 'Liverpudlians', philosophical theories modeled on those of David Lewis are (sometimes, somewhat playfully) called 'Ludovicean', etc.

[10] Thanks to Antti Hiltunen for help with these data.

[11] English speakers: think of difference between the middle consonant in 'peanut' (short) and 'pine nut' (long), or 'booking' vs. 'bookcase'.

[12] Note similar phenomena in English: if I hear you use the expression 'Marxism', I can wonder (*de re*) who Marx was. But, this is not explained by you using the word 'Marx'.

but this root is not something (is not the *kind* of thing) that can be uttered. Utterances of 'Hilkka', 'Hilkan', and 'Hilkkoja' are related to one another etiologically: they are all causally dependent on the same mental symbol. But there is no relevant sense in which these utterances, *qua* public objects, share constituents.[13] Roots play a role in determining which words are spoken, but the fact that words containing the same root can differ in their pronunciation shows that it is not roots themselves that are pronounced. Further, as we saw with Hebrew, there are roots that cannot be pronounced, and as we will see in the next section, there is good reason to think that even when words appear with minimal phonological inflection, to be semantically interpretable a root must be 'packaged' with some grammatical structure. For these reasons roots cannot play the role words were supposed to in standard referentialist pictures.

This exemplifies the dilemma faced by any account that requires that words be passed from one speaker to another. The ability to use a word does not require having heard other speakers use that very word. The root (and whatever other 'pieces of inflection' are used) is all that is needed, but this can be extracted from any number of morphologically (and phonologically) distinct words. However, this root is not what is produced and processed, as roots may lack the properties required to be produced and processed (e.g. a phonological representation compatible with the rules of the language). Roots should instead be viewed as ingredients which determine what is produced (on the speaker side), and products of the processing of what is produced (on the hearer side). So, what unifies a reference-supporting communicative chain is the stable presence of a given root. But these roots are not what is transmitted from speaker to speaker in links of this chain. The traditional picture conflated these two distinct processes.

Before moving onto the instability of semantics, it is worth briefly reflecting on how such basic facts have been overlooked in these discussions which have been absolutely central to the philosophy of language. One reason, I suspect, is the dominance of English within analytic philosophy. English is morphologically highly non-representative of natural languages generally. Within Morphological Typology, dating back at least to Friedrich Schlegel and initiated in its modern form in Greenberg (1960), languages are classified according to the ways that meaning is 'parceled out' into individual words. Analytic languages, like English or Yoruba, largely convey discrete elements of meaning with distinct words. Synthetic languages, like Italian or Malayalam, compress multiple semantic elements into single expressions, differentiating these semantic complexes by phonological inflection. So, for example, person and tense in English are often indicated by the presence of distinct expressions in addition to the main verb ("we will run") whereas in Spanish they are standardly indicated by different inflected forms of the verb on its own ("Correremos").[14] Clearly, this is a difference in degree, rather than in kind: English does have some inflectional and derivational processes ("We run" vs. "We ran"), and Spanish does sometimes parcel out meaning across multiple words ("Nosotros habiamos corrido": *we had run*), but it is mostly agreed that there is a real difference here between ways of packaging meaning. This is important because the plausibility of viewing language as consisting in stable sound-meaning mappings is boosted by focusing on largely analytic languages, in which specific words appear with similar pronunciations across a wide range of grammatical and communicative contexts. The more synthetic a language is, however, the clearer it is that words are products of linguistic processes, generated anew by the combination of distinct linguistic atoms in different conversational contexts, just like

[13] There is an *irrelevant* sense in which they share a phonetic segment /hilk/, but this does not isolate these related expressions from unrelated expressions like 'hilkulla' (*close/almost*).

[14] This discussion is both massively simplified, and somewhat controversial. The point is, of course, not to provide an introduction to Morphological Typology, but just to give an indication of its relevance here.

sentences. This is why I say that the manifest image is at best an idealization as applied to English, for which the derivational processes are subtle and somewhat marginal, but radically misplaced as applied to a language like Finnish or Hebrew for which such morphophonological operations are absolutely central. Any philosophical account of how language must work which assumes that words have stable pronunciations is thus in need of revision.

One other potential explanation for the assumed centrality of words in reference-preservation has been that much work in this tradition has assumed that there *must* be a high degree of inter-personal stability in our uses of language in order to make communication possible.[15] Stemming from Lewis (1969), it has widely been thought that there must be regularities in the way that we use a particular sound to convey a particular meaning, in order to mutually understand our fellow speakers. If what I have said above is correct, it seems that language does not place such a constraint: the same root can lead to quite different pronunciations in different linguistic contexts. But if sound and meaning are as dissociable as I am here arguing, how is communication possible? A full answer here would require more space, but the answer is interesting and important enough to at least gesture to it. I believe the requirement that a root be predictable from the words it comprises is a genuine one, but applies at the diachronic level of language acquisition and change, not the synchronic level of the computational system of the language. That is, there are no morphophonological constraints which preclude radically different pronunciations of the same root in different contexts. But such mappings will be particularly difficult for novice speakers to learn, and so are unlikely to remain stable in a linguistic population without strong pressures in their favour. This explains why there is a strong correlation between word frequency and morphophonological variation, as witnessed by the fact that the verb 'to be' typically corresponds to far more phonological forms than any other. Common words are heard frequently enough to learn that they are exceptional, while if an irregular word is rare enough that it may not be heard by the child, it is liable to simply be usurped by a regular counterpart.

4 | THE UNSTABLE LEXICON II: SYNTAX AND SEMANTICS

Hopefully the preceding has convinced that competence with a particular word does not require having encountered someone else using that word previously. Lexical competence requires competence with a *root*, but each application of such a competence may result in words with quite different pronunciations. But what about the other half of the sound-meaning pairing traditionally identified as the nature of a word? Meaning, in the sense of reference or that which determines reference, has been at the centre of philosophical discussion of words, and for obvious reasons. As detailed above, the inter-personal transmission of words has been appealed to primarily to explain our ability to refer to non-local entities or states of affairs. And indeed those making such an appeal (e.g. Kaplan) have often been explicit about their reliance on the stability of reference, even while pronunciations and the like may change radically. It may be thought that, so long as the *reference* of the items in the chain (alternately, roots and words) remains stable, the spirit of the traditional picture is vindicated, even if the letter is mistaken. In this section I shall argue that reference is just as susceptible to modulation and modification within the computational system of language as sound.

[15] Ruth Millikan's picture of the relations between mind and language has centered the relationships between stable patterns of inter-speaker coordinations of sound and meaning and the possibility of successful communication. See Millikan (2008) for representative discussion.

Before making this case, however, it will be useful to make a few distinctions in order to clarify somewhat the notoriously murky notion of the meaning of a word. Firstly, as noted above, the targets here are accounts of *reference* that appeal to transmission of words. For this reason, I shall be assuming that meaning is, or at least is determinative of, reference. If the meaning of a word did not determine its referent, then acquiring a word from someone would not explain how one manages to refer to non-local referents, and so the above account of this capacity would be short-circuited. If meaning doesn't determine reference, the story doesn't even get off the ground. I am also perfectly happy to allow that meaning is *more* than reference. Perhaps lexical meaning includes conceptions (in the sense of Rey, 1983), or inferential connections, or whatever. I here remain neutral on such debates. What is crucial for my purposes is that the referential capacities we acquire when we encounter the words of others outstrip the referential uses these words were put to in the communicative exchanges in which we acquired them. Specifically, we can acquire a single root from encounters with a range of different words, with different referential capacities, and then put this root to novel referential uses. Thus, the many-to-many relationship between roots and words is found just as much for reference as for pronunciation.

To make this case, I will also be relying on the standard distinction between semantic *type* and semantic *value*. The semantic type of an expression is the kind of semantic contribution it makes to an utterance: proper names are traditionally considered to be of type *e*, referring to individual objects, intransitive verbs are of type <*e,t*>, referring to sets of individuals or their characteristic functions, and so on. Semantic values differentiate between expressions of the same semantic type. So 'Hans Zimmer' and 'John Williams' are expressions of the same semantic type, referring to individuals, but with different values, one referring to Hans Zimmer, the other referring to John Williams. Compositional semantic theories have insightful things to say about semantic types, but typically little to say about specific semantic values (see Glanzberg, 2014). What matters for my purposes is that sharing a semantic type is necessary (although not sufficient) for shared meaning. There is no sense to be made, within standard semantic theory, of the idea that a type *e* and a type <*e,t*> expression could mean the same thing.

The problem then, for the idea that lexical meaning is stable across chains of communication, is that a single root can apparently be used to convey meanings of quite different semantic types. Human language-users display a remarkable ability to extend their lexical competencies beyond their experience. While we may have the intuition that certain words are specified as members of a particular grammatical or semantic category, as transitive verbs or mass nouns, say, and thus as applying to worldly referents of specific ontological categories, in practice we are prone to use, and happy to interpret, such expressions in a much wider variety of ways. Consider, for example, the following sentences:

1. Water is essential for life.
2. There is water all over the floor.
3. We'll have three waters.
4. The plants need watering.
5. This beer is pretty watery.

English speakers competent with the expression 'water' can thereby interpret sentences 1–5. But in each sentence, this expression is put to a quite different semantic use: as a kind term, a mass noun, a count noun, a transitive verb, and an adjective. It is debatable here whether these are examples of the same word being used in different ways, or of different words. Neither option, however, will help defenders of the traditional picture. In the latter case, the upshot is the same as

in the previous discussion: we can use words perfectly competently even if we have not heard others use those very words. In the former case, we can use words with semantic/referential properties which we did not acquire from our prior experience with that word.

As above, the same phenomena can be observed with proper names:

6. Houdini died in 1924.
7. The Houdinis were devastated by his death.
8. I Houdini'd my way out of the closet.[16]
9. He must have some Houdini in him, to escape like that.
10. That touchdown was Houdini-esque.

Here again we see the same expression being used, with varying degrees of detectable morphological processing, to make singular reference, as a plural noun, as a verb, as an abstract noun, and as an adjective. Our experience with individual instances of these uses seems, at least in principle, to ground our ability to produce and interpret other, semantically distinct, uses. When we use 'Houdini' as a verb, we do not (linguistically) refer to Houdini, and we need not have heard anyone else do. Thus again we see that the creative mechanisms of morphology expand the semantic potential of our language, in ways unexpected on the traditional account.

Yet further, expressions which are *not* acquired as proper names will be interpreted as if they are (i.e. with singular, rigid, reference to an individual) when found in the grammatical context characteristic of names[17]:

11. Cat! What are you doing on the bed?
12. Hey, Sunglasses, watch where you're going.

As Borer (2005a,b) argues, the best predictor of the semantic type of an expression is thus not the lexical item in question itself, but its grammatical (syntactic/morphological) context. While we might conceptualize water as a mass, we find it very easy to interpret as applying to specified portions of water, in a 'count-like' way. However, given a particular grammatical context, we see no such flexibility: Sentence 2 can *only* be used to talk about a mass of water (not, e.g., a situation in which there are many cups of water on the floor). She infers from this that we should view open-class expressions like 'water' as intrinsically syntactically (and therefore semantically) uncategorized. Given that interpretation is determined by grammatical context, it would be redundant to mark lexical items as themselves categorized. For example, assuming that 'Houdini' is specified in the lexicon as a proper name (type *e*), this information will be redundant in the context of sentence 6, as anything in this position will have to be of this type, and will be over-ridden in sentences 7–10 so as to allow for proper composition. This argumentative strategy, it is worth noting, is a fairly natural continuation of the tradition of identifying grammatical categories by their distribution: If nouns are identified by occurring in particular grammatical contexts, then marking the expression as nominal *in addition to* its occurrence in such environments is redundant. It is better to simply say that there is nothing more to being a noun than occurring in these environments.

[16] This example is from Clark and Clark (1979).

[17] Gleitman et al. (1977) provide an example of a child (7) recognizing this grammatical fact in the following dialogue, in which the interviewer (LG) is investigating the child's (CG) intuitions about linguistic acceptability.LG: How about this one: *Boy is at the door*?CG: If his name is Boy…

The point is not that a linguistic theory could not be made to work which posited basic syntactic/semantic categories for lexical items, and then posited various kinds of coercion and type-shifting machinery to make things go smoothly when expressions are found in unpredicted environments. Rather, the point is that such machinery is unnecessary and unmotivated. We need such machinery only on the supposition that words are pre-specified as nouns, adjectives, etc., with their characteristic semantic types. The above kinds of data, however, suggest that a simpler linguistic theory views semantic structure as syntactically imposed on intrinsically uncategorized open-class morphemes. This view is adopted by, among others, Borer (2005a,b, 2013), Harley (2014), and Acquaviva and Panagiotidis (2012).[18]

If this view is correct, however, it follows that the meaning (i.e. referent) of an expression in one person's usage need not be shared with the expression as used by the speaker from whom the expression was acquired. While I may have learned the word 'Houdini' from hearing others refer to the individual, Harry Houdini, my own uses need not share this referent. If I use this expression verbally or adjectivally, my uses are not even of the same semantic type as that on which my usage depends. And, going in the other direction, when I hear a sports commentator utter 10, and then wonder (or even ask) "Who was Houdini?", I manage to refer to this individual, despite the fact that the utterance from which I acquired the word did not. Thus, it cannot be that our capabilities to refer are essentially explained by appeal to chains of transmission of words with the same referential properties.

In the specific case of proper names, much of this literature, modifying a proposal by Longobardi (1994), has it that singular reference is conditioned by entire Determiner-phrases, not individual lexical items themselves. So, to get a proper name to refer, it must be found within a larger expression alongside various functional elements such as an explicit determiner expression, (as in Greek: "O Giannis") or unpronounced expressions which specify the larger structure as definite (see Borer (2005a) for the most extensive account of nominal structure along these lines). When a proper name is then found in other sorts of grammatical contexts (i.e. in construction with other sorts of functional items), as in 7–10, its semantic role is quite different. This shows that what we acquire when we become competent with a name is not a 'device for reference' as often stated (see e.g. (García-Carpintero, 2018): "names, like indexicals, are de jure constitutively referential expressions" p. 1126). Instead, competence with a name, like any other expression, involves possessing a root which contributes an abstract component of meaning, which must be integrated into a grammatical structure for full interpretation.[19] So, the root of 'Houdini' indicates that some connection to the famous escape artist is being appealed to, but how this root interacts with functional grammatical structure will determine what use this connection will be put to: referential in a DP context, as in sentence 6; identifying events with certain properties drawn from my concept of said escape artist as in sentence 7; and so on. What constraints there are on cross-categorial uses of lexical items are then accounted for semantically, rather than grammatically: "He Houdini'd his sandwich" sounds unacceptable, not because it features what is intrinsically nominal in a verbal context, but simply because we lack the knowledge of Houdini necessary to determine what

[18] Of course, there is much more to say here, and much could be said in defense of the 'typeshifting' approach to semantics more familiar to many philosophers (stemming from seminal work such as Partee (1986) and Partee and Rooth (1983)). The point in this paper is not to establish that this approach to lexical meaning, featuring acategorical roots and grammatically enforced semantic typing, is the correct view. Just to point out that it is a live and plausible option, and draw out some philosophical implications of it if true.

[19] If anything is a "device for reference", it is the entire DP structure, into which many expressions other than names can be placed in order to generate a referring expression.

could be meant. But once we acquire such knowledge (say, we discover that Houdini preferred his sandwiches deep fried), interpretation becomes relatively easy.[20]

The linguistic approaches that incorporate this sort of picture differ in various ways, especially on the properties of roots. However, they share a broad architecture suggestive of a picture of language quite unlike the manifest image. Sentences are not merely composed out of words, which have been collected through interactions with other speakers. Rather, the computational engine of syntax generates representations, hierarchically structured by abstract, functional items specifying how to interpret minimally informative open-class roots. Rules of morphophonology then specify that a particular 'chunk' of grammatical structure, composed out both open-class and functional items, has a specific pronunciation.[21] Syntactic/semantic rules specify that particular sorts of structure are semantically typed in predictable ways (e.g. a definite DP calls for a referent). The concept associated with the root can then enable the identification of the entity/property/event under discussion, either as a product of the information contained in an associated "mental file" or in a more externalist way, through causal connections to the world.[22] These two interpretive processes (at the morphology-phonology interface, and at the syntax-semantics interface) operate distinctly from one another. There is no input to the grammatical process which specifies a mapping between sound and meaning. What we acquire from our interactions with others are thus neither the pronunciations nor meanings of words, which are not atoms but products of this computational system, but the properties of these roots, which stand in highly complex relations to these outputs. In the words of Marantz (1995): "Words like sentences are the interpreted output of the computational system of grammar." (p. 17). There thus seems to be no place for the traditional notion of words assumed in typical discussions in philosophy of language.[23]

While I take the view developed here to be fairly revisionary, at least within philosophy, it is important to stress its limits. This paper discusses a particular *meta-semantic*, *diachronic* proposal for how linguistic expressions come to have their semantic, specifically their referential, properties. What is being denied is that a speaker's words at a given time refer to what they refer to in virtue of these same words being used to make this same reference by earlier speakers. This is not to challenge any (semantic, synchronic) claims about the reference that a particular use of a word

[20] This capability is also somewhat difficult to square with the standard referentialist line that the meaning of a name is 'exhausted' by its referent, as in such non-referential uses it is clear that it is broader encyclopedic knowledge that determines what is meant (although this could be accounted for pre-semantically).

[21] The process does not necessarily end there. Between morphophonology and actual speech, there may be many further systematic or unsystematic processes influencing what is actually produced, ranging from phonological adjustment rules (e.g. ensuring that subsequent phonemes are not too similar or dissimilar for ease of production) to "performance effects" like mid-sentence hiccoughs.

[22] Special meanings can also be associated with such grammatical structures, as in idiomatic phrases.

[23] Note that merely rejecting this approach to theoretical linguistics does not necessarily help much for those who rely on the manifest image. Lexicalist approaches to linguistics (Di Sciullo and Williams (1987), and Williams (2007)) which view the word as a grammatically significant unit of its own, agree that words are typically a product of substantial generative procedures, not simply sound-meaning pairings stored in memory to replicate previous experience. This idea that meaning is predictable not on the basis of individual words, but only on words in specific grammatical contexts is widespread in linguistics. Fisher et al. (1991) argue that language must be organized in this way, for it to be learnable by children.While it violates the letter of the manifest image, its spirit may be more closely aligned with work in Construction Grammar (Goldberg, 2003) which, though it denies the centrality of words, does view language use as a matter of recycling acquired 'chunks' of language. For, inter alia, reasons discussed in Adger (2013) and Lidz and Williams (2009), I doubt that defenders of the manifest image should place their faith in the success of Construction Grammar as a complete story about human language. But for my purposes I am happy to rest with the conditional conclusion: to the extent that the generativist approaches discussed are correct, the manifest image of language is mistaken.

has. Indeed, I am taking it for granted that words can be used to enact reference. The question is just: what enables this?

This distinction permits, and may even motivate, the standard strategy within formal semantics of treating individual words as semantic atoms. While the work in morphology here canvassed strictly denies this, it is a reasonable idealization for the purposes of modeling compositional semantics. On my picture, the word 'Houdini' is not an atom, but has internal morphosyntactic structure. That this word refers to Houdini is thus not a basic, ground-level stipulation, as in most semantics texts, but is derived from this internal structure combined with the causal history of this structure's root. But this additional complexity is strictly irrelevant for most work within compositional semantics, and can thus be ignored. This is thus strictly analogous to the way that linguists annotate trees with triangles to cover up irrelevant internal structure (e.g. treating the DP 'the magician' as a unified whole). As is typical in any scientific endeavour, what is relevant for one inquiry may not be for another. That a component of a system has a relevant property (e.g. that a given word has a given grammatical category and meaning) is what matters for much theorizing, not what makes this the case (e.g. the presence of a given root in a particular kind of functional structure). In describing how, say, quantifier scope works, all that is likely to be relevant is that some expression is of type *e*, and so this can simply be stipulated within the lexicon of the semanticist's toy model. But, for other purposes, the facts that have been idealized or abstracted away from may become crucial (e.g. as when semanticists develop theories of tense or aspect, which requires positing meanings for sub-lexical morphemes), and of course a "complete" semantic theory must de-idealize and show how the semantics of words is derived from their internal structure and parts. My claim is that for a philosophical metasemantics accounting for how these expressions acquire these referential capacities, the distinction between a stable linguistic atom, and an unpronounced but repeatable component of many phonologically and semantically distinct structures is a crucial one, even if much first-order work in semantics can ignore this.

5 | REFERENCE WITHOUT WORDS

If the above discussion is on the right track we do not manage to refer to distant entities in virtue of acquiring words which themselves refer to distant entities. How, then, do we do this? One possible response to this sort of picture, defended by Chomsky (2000), is to adopt a highly skeptical attitude towards the entire notion of linguistic reference. For both empirical and theoretical reasons, I would rather try and salvage some of the substantial insights of the referentialist tradition. In this section I will sketch an alternative picture which, I believe, can do this without relying on the manifest image of language, and the ontology of words featured within it.

While I think the extent to which using language is a matter of latching onto objective features of the world can sometimes be over-stated, it is almost undeniable that we do, at least sometimes, manage to pick out worldly entities. Putnam (1975) and Burge (1979, 1982) establish this through a range of thought experiments. More generally, however, it seems to be a core commitment of any minimally rational view of the world that we sometimes manage to say (objectively, mind-independently) true things. And the only way I know of to make sense of this fact is that our words sometimes manage to latch onto the mind-independent world. When I say things about Gandhi, whether what I say is true or false depends, at least in part, on how Gandhi was, and not on how anyone with arbitrarily similar properties to Gandhi was, or how my or your mental representation of Gandhi is. This word-world connection needs explaining.

However, it is not clear that though this is, in some sense, a linguistic phenomenon, it requires a strictly linguistic solution. I believe that the mistake made by Kripke, Kaplan, and others was to assume that a linguistic capacity like reference must be traced to the properties of the public linguistic expressions used in acts of referring. This obscures the ways that our linguistic system is embedded in our broader cognitive system. Many apparent properties of linguistic expressions (such as reference) are liable to be explicable only by appreciating this broader context. Assuming that words themselves refer, rather than provide, in concert with this broader cognitive make-up, the tools for enacting reference, limits the tools available for a philosophical theory of reference.

Identifying and keeping track of individual referents is a cognitively and evolutionarily deep capacity. Pylyshyn (2007) discusses the ways that the earliest stages of perceptual processing involve identifying external objects, distinguishing them from one another, and tracking the ways they evolve. Burge (2007) contains rich discussion of the ways that linguistic reference is grounded by perceptual referential capacities. Cheney and Seyfarth (2008, 2018) discuss the ways that non-human primates store information about particular conspecifics and other ecologically relevant agents, and communicate with one another to coordinate such information. And the literature on "mental files" (see e.g. Recanati (2012) and Green & Quilty-Dunn (2020)) posits locations in the memory systems of human and non-human animals specifically dedicated to keeping tabs on specific referents. So, there are several strands of research pointing to the ways that cognition is geared towards latching onto extra-mental entities.

I think this paves the way towards thinking of linguistic reference as a sort of interface phenomenon between these general cognitive mechanisms and the uniquely human language faculty. We can use the capacities made available by the language faculty to publicly identify individuals in ways that allow others to latch onto them. But language is not, unlike most animal communication systems (Armstrong, 2021), a specialized device for doing this, and the relationship between these general cognitive capacities and the structures made available by language is highly messy. Philosophy and cognitive science would be furthered by not assuming any simple relationship here, as is assumed when it is claimed that linguistic items are tools specifically for reference. Better to view reference as a complex phenomenon made possible only by the coming together of general cognitive skills for tracking with specific kinds of abstract linguistic structures. Syntax-semantics interface rules specify that a structure requires an interpretation of a specific ontological sort (e.g. of an event, or an individual), and then the roots involved in these structures are associated with concepts which provide clues as to how to construct such an interpretation. Sometimes, such clues will be pretty decisive– the concept 'Houdini' used in a definite DP refers to Harry Houdini– while sometimes interpretation will require more substantial reliance on background knowledge, as in sentences 8–10 above. But such word-world connections are always enacted as a complex interaction between grammatical structures and more general features of cognition.[24]

This proposal does not preclude the centrality of individual reference in our cognitive systems. I think much of the above-cited work does indeed suggest that keeping track of specific entities is absolutely core to the kinds of minds we have. I deny that the chains of communication that underwrite our referential capacities need to be chains of individual reference. But this does not mean that individual reference is not explanatorily central to any story here. There is generally an asymmetrical dependence of the non-individually-referring uses of a root on the individually referential uses: we could refer to Houdini without knowing that he was an escape artist, but

[24] See Pietroski (2018) and Glanzberg (2018) for related ideas about how lexical structure packages meaning.

we could not use the verb 'Houdini' to talk about an act of elusiveness without such knowledge. But this centrality must, I believe, be understood in terms of interactions between our linguistic system, which provide ways of packaging linguistic units so as to call for interpretations of particular kinds, and our broader cognitive system, which provides the raw materials for specifying the specific interpretations called for.

One way in which human linguistic capacities (construed broadly) are unique is in their ability to form networks of speakers spanning vast distances, both spatial and temporal, and to allow some speakers' utterances to depend, semantically, on others within the network. The standard "chains of communication" picture was correct to highlight these long-distance dependencies as essential in explaining how linguistic reference is possible. However, the continuity of a chain of communication linking me with Elpinice does not require that there be some item, a word, which is exemplified by each link in the chain. What matters is simply that each link in the chain enables subsequent speakers to acquire the capacity for *de re* thought about the same entity as previous speakers. Thus, we can say that the referential potential of the root can be passed on from one speaker to another, even if the words uttered by each speaker are quite different, both in their meaning and in their pronunciation. The manifest image, reflective of contingent facts about English rather than language generally, suggested that the links in the chain, words, are the focus point in such a story, but this was a mistake. Reference is conveyed along causal/informational chains, structured by the organizational principles of our minds, whether or not the links of these chains feature the same words.

I believe this change in perspective points towards solutions for many puzzling features of natural language. For example, if natural language lexical meaning is as closely connected to reference as is assumed by the manifest image of language, the apparent meaningfulness of non-referential expressions poses a deep worry (see e.g. Kripke (2013) and Everett (2013) for discussions of the problem with very different and elaborate solutions). However, from the perspective of the language faculty, the existence or lack thereof of a referent for a (referential use of a) given word seems irrelevant. 'Harry Potter' behaves linguistically in exactly the same way that 'J.K. Rowling' does. By viewing reference as the product of specific linguistic structures interacting with broader aspects of cognition, we can explain referential (and thus truth-conditional) failures without positing any linguistic difference.[25]

My admittedly quite sketchy proposal here is thus in line with the "pragmatist" approaches to reference pioneered by Strawson (1950) and Donnellan (1972, 1966), and developed by e.g. Bach (1987), Perry (2001), and Cumming (2021), according to which reference is, first-and-foremost, a matter of "having something in mind", and then using whatever linguistic (or even non-linguistic) resources are available to attempt to get your audience to have the same entity in mind. Thus, reference is sharply demarcated from meaning, the latter of which can then be treated as a genuinely language-internal phenomenon, perhaps in the internalist manner of Pietroski (2017), or in terms of providing linguistically encoded constraints on reference, as in Harris (2020). The slogan, from Strawson, is that words don't refer, people refer (1950 p. 326). Hopefully the above has shown how such a view can be integrated within, and motivated by, work from contemporary morphological theory.

[25] I say this "points towards" solutions, not that it solves these problems in their entirety. There may still be a question of how we manage to "think about" non-existent entities. But showing that a problem is not a problem for linguistic theory to deal with seems like progress here.

6 | CONCLUSION

The manifest image of language offers up the following explanation for how an agent's reference to temporally and spatially distant referents is possible: the agent is equipped with a *word* which itself refers to such a referent, and the agent is able to utilize this word to enact such an act of reference. This immediately raises the further question: how does this word come to have this referential capacity? And the standard story, from Kripke, Kaplan and others, is that this word was acquired from some other speaker who themselves used this word to enact a reference to this referent, and the current agent's usage is now deferential to that of whomever the word was acquired from. Such deferential and referential links can be summed, creating a chain stretching back to some primordial use of the word which referred to this referent, but whose referential capacities were explained in some other way (e.g. the word was used to refer to some perceptually available object). While this picture is admittedly very elegant, it makes certain empirical assumptions. Specifically, it assumes (i) that there is some element common to each communicative exchange constituting the causal chain, and (ii) this common element is public, i.e. it is both produced by the speaker and perceived by the audience. Morphological theory, by distinguishing words from roots, precludes the identification of anything meeting both of these conditions. Each communicative exchange will involve the production and perception of words, but not necessarily the same words in each exchange. In principle, each link could feature distinct words (e.g. speaker 1 says 'inhabit', speaker 2 says 'inhabitable', speaker 3 says 'uninhabited' and so on). On the other hand, while each such exchange must involve the same root (in order to count as links in the same causal chain), these roots are not what is pronounced or perceived. As we saw above, in many cases roots do not provide enough phonological information to specify a pronunciation (e.g. the triconsonantal radicals of Hebrew), and even when they do the production of just this material is often not legitimate (e.g. in Spanish where gender marking is obligatory on many nouns and verbs).

The mainstream picture of communication assumed within philosophy of language treats words as the atoms of language. One word can be used by a given speaker, and then acquired by another to be used in a similar way in the future. But several approaches within modern morphological theory reject this picture. Uttering a word is a creative process, just like uttering a sentence. And parsing a sentence requires going beyond decomposing it into its component words. The components of these, i.e. roots, must often themselves be extracted. Chains of communication can then involve putting these roots to quite different purposes, so that each link in the chain may feature words with quite different phonological and semantic properties. This revised, and perhaps for English speakers counter-intuitive, picture of language calls for similarly substantial revisions to our philosophical understanding of phenomena such as communication and reference.

Bromberger (2011) closes his commentary on Kaplan's theory of words by beseeching philosophers to "[T]urn their metaphysical binoculars around to see the facts more closely. Not abstract facts, but concrete ones. Not how we talk about words or might prefer to talk about words to avoid puzzles, but what words come to in the actual world." (p. 503). I hope that the above discussion can be viewed as an initial attempt to do just this. It is easy to assume that natural language is more-or-less how it appears to us, both perceptually and as part of our common-sense view of the world. And philosophical theories are just as liable to rely on such a picture as anything else. But decades of work in theoretical linguistics has shown us beyond doubt that natural languages are radically unlike this manifest image. Words are not the atomic constituents of sentences, but are computationally derived from a variety of dissimilar simpler parts. The atoms of language do not

come pre-packaged with a meaning and a pronunciation, but mean different things, and sound different ways, in different grammatical contexts. And words are not simply tools that we can acquire from others, and then pick up off the shelf, to be used as they were when we first encountered them, but like sentences are actively reconstructed on different usages. Viewing language in these ways calls for substantial revisions to our philosophical accounts of language, communication, thought, and our place in the world. But it also suggests solutions, or dissolutions, to some puzzles that have been plaguing us for decades if not centuries. At a minimum, I hope the above has convinced people that it pays to look to this concrete work in the linguistic sciences in order to assess the foundations on which our philosophical theories are built.

ORCID

Gabe Dupre https://orcid.org/0000-0001-9312-4691

REFERENCES

Acquaviva, P., & Panagiotidis, P. (2012). Lexical decomposition meets conceptual atomism. *Lingue e linguaggio*, 11(2), 165–180. https://doi.org/10.1418/38784

Adger, D. (2013). Constructions and grammatical explanation: Comments on Goldberg. *Mind & language*, 28(4), 466–478. https://doi.org/10.1111/mila.12027

Armstrong, J. (2021). Communication before communicative intentions. *Noûs*. https://doi.org/10.1111/nous.12396

Bach, K. (1987). *Thought and reference*. Oxford: Oxford University Press.

Baunaz, Lena., Haegeman, Liliane., De Clercq, Karen., & Lander, Eric. (Eds.) (2018). *Exploring nanosyntax*. Oxford University Press.

Borer, H. (2005a). *Structuring Sense: Volume I: In Name Only*. Oxford University Press.

Borer, H. (2005b). *Structuring Sense: Volume II: The Normal Course of Events*. Oxford University Press on Demand.

Borer, H. (2013). *Structuring Sense: Volume III: Taking Form*. Oxford University Press.

Bromberger, S. (2011). What are words? comments on Kaplan (1990), on Hawthorne and Lepore, and on the issue. *The Journal of Philosophy*, 108(9), 486–503. https://doi.org/10.5840/2011108925

Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy*, 4(1), 73–122. https://doi.org/10.1111/j.1475-4975.1979.tb00374.x

Burge, T. (1982). Other bodies. In A. Woodfield (Ed.) *Thought and Object: Essays on Intentionality*, (pp. 97–121). Oxford: Oxford University Press.

Burge, T. (1986). Intellectual norms and foundations of mind. *The Journal of Philosophy*, 83(12), 697–720. doi: jphil198683121

Burge, T. (2007). *Foundations of Mind*. Oxford University Press.

Cheney, D. L., & Seyfarth, R. M. (2008). *Baboon metaphysics*. University of Chicago Press.

Cheney, D. L., & Seyfarth, R. M. (2018). *How monkeys see the world: Inside the mind of another species*. University of Chicago Press.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press.

Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, (pp. 767–811). https://doi.org/10.2307/412745

Cumming, S. (2021). *What is reference?* UCLA Ms.

De Saussure, F. (1916/2011). *Course in general linguistics*. Columbia University Press.

Di Sciullo, A., & Williams, E. (1987). *On the definition of word*. Cambridge, MA: MIT Press.

Donnellan, K. S. (1966). Reference and definite descriptions. *The philosophical review*, 75(3), 281–304. https://doi.org/10.2307/2183143

Donnellan, K. S. (1972). Proper names and identifying descriptions. In D. Davidson, & G. Harman (Eds.) *Semantics of natural language*, (pp. 356–379). Springer.

Everett, A. (2013). *The nonexistent*. Oxford University Press.

Fisher, C., Gleitman, H., & Gleitman, L. (1991). On the semantic content of subcategorization frames. *Cognitive psychology*, 23(3), 331–392. https://doi.org/10.1016/0010-0285(91)90013-E

García-Carpintero, Manuel. (2018). The Mill-Frege theory of proper names. *Mind*, 127(508), 1107–1168. https://doi.org/10.1093/mind/fzx010

Glanzberg, M. (2014). Explanation and partiality in semantic theory. *Metasemantics: New essays on the foundations of meaning*, (pp. 259–292). https://doi.org/10.1093/acprof:oso/9780199669592.003.0011

Glanzberg, M. (2018). Lexical meaning, concepts, and the metasemantics of predicates. In D. Ball, & B. Rabern (Eds.) *The Science of Meaning: Essays on the Metatheory of Natural Language Semantics*, (pp. 197–225). Oxford University Press. https://doi.org/10.1093/oso/9780198739548.003.0007

Gleitman, L. R., Gleitman, H., & Shipley, E. F. (1977). The emergence of the child as grammarian. In *Topics in cognitive development*, (pp. 91–117). Springer. https://doi.org/10.1007/978-1-4613-4175-8_8

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in cognitive sciences*, 7(5), 219–224. https://doi.org/10.1016/S1364-6613(03)00080-9

Green, E. J., & Quilty-Dunn, J. (2020). What is an object file? *The British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axx055

Greenberg, J. H. (1960). A quantitative approach to the morphological typology of language. *International journal of American linguistics*, 26(3), 178–194. https://doi.org/10.1086/464575

Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale, & S. J. Keyser (Eds.) *The View From Building 20: Essays in Honor of Sylvain Bromberger*, (pp. 111–176). MIT Press.

Halle, M., & Marantz, A. (1994). Some key features of distributed morphology. *MIT working papers in linguistics*, 21, 275–288.

Harley, H. (2014). On the identity of roots. *Theoretical linguistics*, 40(3-4), 225–276. https://doi.org/10.1515/tl-2014-0010

Harley, H., & Noyer, R. (2014). Distributed morphology. In *The second Glot International state-of-the-article book*, (pp. 463–496). De Gruyter Mouton. https://doi.org/10.1515/9783110890952.463

Harris, D. W. (2020). Semantics without semantic content. *Mind & Language*. https://doi.org/10.1111/mila.12290

Julien, M. (2007). On the relation between morphology and syntax. In G. Ramchand, & C. Reiss (Eds.) *The Oxford handbook of linguistic interfaces*, (pp. 209–238). Oxford University Press Oxford. https://doi.org/10.1093/oxfordhb/9780199247455.013.0008

Kaplan, D. (1989a). Afterthoughts. In J. Almog, J. Perry, & H. Wettstein (Eds.) *Themes From Kaplan*, (pp. 481–563). Oxford University Press.

Kaplan, D. (1989b). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.) *Themes From Kaplan*, (pp. 481–563). Oxford University Press.

Kaplan, D. (1990). Words. *Aristotelian Society Supplementary Volume*, 64(1), 93–119. https://doi.org/10.1093/aristoteliansupp/64.1.93

Kaplan, D. (2011). Words on words. *The Journal of Philosophy*, 108(9), 504–529. https://doi.org/10.5840/2011108926

Kripke, S. A. (1980). *Naming and Necessity*. Harvard University Press.

Kripke, S. A. (2013). *Reference and Existence: The John Locke lectures*. Oxford University Press.

Lewis, D. (1969). *Convention: a philosophical study*. Cambridge, Harvard University Press.

Lidz, J., & Williams, A. (2009). Constructions on holiday. *Cognitive Linguistics*, 20(1), 177–189. https://doi.org/10.1515/COGL.2009.011

Longobardi, G. (1994). Reference and proper names: A theory of nmovement in syntax and logical form. *Linguistic inquiry*, (pp. 609–665).

Marantz, A. (1995). "cat" as a phrasal idiom: Consequences of late insertion in distributed morphology. MIT Ms.

Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania working papers in linguistics*, 4(2), 14.

Mercier, A. (1998). On communication-based de re thought, commitments de dicto, and word individuation. In K. Murasugi & R. Stainton (Eds.) *Philosophy and Linguistics*, (pp. 85–111). Boulder, CO: Westview Press.

Miller, J. T. M. (2021). A bundle theory of words. *Synthese*, 198(6), 5731–5748. https://doi.org/10.1007/s11229-019-02430-3

Millikan, R. G. (2008). A difference of some consequence between conventions and rules. *Topoi*, 27(1-2), 87–99. https://doi.org/10.1007/s11245-008-9026-3

Modrak, D. K. W. (2001). *Aristotle's theory of language and meaning*. Cambridge University Press.

Partee, B., & Rooth, M. (1983). Generalized conjunction and type ambiguity. In R. Bauerle, C. Schwarze, & A. von Stechow (Eds.) *Meaning, Use and Interpretation of Language*. Walter de Gruyter, Berlin. https://doi.org/10.1515/9783110852820.361

Partee, B. H. (1986). Noun phrase interpretation and type-shifting principles. In J. Groenendijk, D. de Jongh, & M. Stokhof (Eds.) *Studies in discourse representation theory and the theory of generalized quantifiers*, (pp. 115–144). Dordrecht: Foris. https://doi.org/10.1515/9783112420027-006

Perry, J. (2001). *Reference and reflexivity*. CSLI Stanford.

Pietroski, P. (2017). Semantic Internalism. In J. McGilvray (Ed.) *The Cambridge companion to Chomsky*. Cambridge University Press (pp. 196–216). https://doi.org/10.1017/9781316716694.010

Pietroski, P. (2018). *Conjoining Meanings: Semantics Without Truth-Values*. Oxford University Press.

Putnam, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science*, 7, 131–193. Reprinted in H. Putnam (2010) *Mind, Language and Reality: Philosophical Papers Volume 2*. Cambridge University Press. https://doi.org/10.1017/CBO9780511625251.014

Pylyshyn, Z. W. (2007). *Things and places: How the mind connects with the world*. MIT press.

Recanati, F. (2012). *Mental files*. Oxford University Press.

Rey, G. (1983). Concepts and stereotypes. *Cognition*, 15(1), 237–262. https://doi.org/10.1016/0010-0277(83)90044-6

Salmon, N. (2005). Are general terms rigid? *Linguistics and Philosophy*, 28(1), 117–134. https://doi.org/10.1007/s10988-004-2430-2

Sellars, W. (1963). Philosophy and the scientific image of man. In W. Sellars, *Science, perception and reality*, (pp. 35–78). Ridgeview Publishing Company.

Soames, S. (2002). *Beyond rigidity: The unfinished semantic agenda of naming and necessity*. Oxford University Press, USA.

Strawson, P. F. (1950). On referring. *Mind*, 59(235), 320–344. https://doi.org/10.1093/mind/LIX.235.320

Strawson, P. F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. Routledge.

Williams, E. (2007). Dumping lexicalism. In G. Ramchand, & C. Reiss (Eds.) *The Oxford handbook of linguistic interfaces*, (pp. 353–382). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199247455.013.0012

How to cite this article: Dupre, G. (2022). Reference and morphology. *Philosophy and Phenomenological Research*, 1–22. https://doi.org/10.1111/phpr.12896