NB: This is a pre-print. The full version is forthcoming in *From Contradiction to Defectiveness to Pluralism in Science: Philosophical and Formal Analyses* (Synthese Library) Bueno, O. and Martinez-Ordaz, M. (eds.) If you wish to cite, please refer to published version.

Classifying Data and Interpreting Theories: The Case of Generative Grammar

Gabe Dupre

Keele University

1 Introduction

There are, of course, many ways that data can be defective. Some data are defective in virtue of being vague or imprecise, false, or inaccurate, or even incoherent. I shall focus, however, on another, perhaps less drastic, way in which data can be defective: irrelevance. From the perspective of a given theory, many of our observations are simply irrelevant to confirmation, in that they carry minimal usable information about the target of interest, and can thus be safely ignored. However, determining which data are defective in this way and which are not is a difficult empirical question. I shall describe some general strategies used in making such a determination in generative linguistics, in which the question of which observations should be taken to be relevant to theory confirmation has been hotly debated over the past half century or so. I shall argue that the need to draw such a distinction places a heavy burden on possible interpretations of these theories. Thus, the question of which data are defective and which are not will turn out to have repercussions in one of the central debates in philosophy of linguistics, concerning the psychological reality of linguistic theories.

2 "Mere Performance Data"

As in any science, linguistics involves gathering observational data and using such data to select those hypotheses most likely to be true of the target domain. In linguistics,

this domain is, of course, language. In this domain, there are many different questions we could ask, and thus different kinds of hypotheses we could offer for testing against our observations. Paradigmatic questions include: what, if anything, are the universal constraints on what human language can be? And, within these constraints, in which specific ways do individual languages differ from one another? Any viable answers to such questions must be compatible with, and ideally would contribute to, accounts of how it is that humans acquire and use language.

One standard style of answer to such questions involves identifying the various possibilities and impossibilities made available by particular languages and by language in general. This approach bears a striking resemblance to traditional philosophical pictures of science as aimed at the discovery of *laws*: true, modally robust, generalizations. For example, it has been proposed that all human languages are structure-dependent. That is, rules within these languages make reference to abstract, structural properties of, and relations between, constituents, and never to the linear order of words. This provides a constraint on *possible* languages. Consider question formation. There are many conceivable ways that questions could be formed in a language, but not all of these are compatible with language being structure-dependent. We could, for example, imagine a language which formed questions from declaratives by placing heavy stress on the fourth word of a sentence, or by inverting the 3rd and 4th word, or by producing the entire sentence backwards, and so forth. The proposed structure-dependence constraint excludes such possibilities. And indeed, no such options are found when we investigate the languages of the world. Rather, we find question-formation procedures defined with reference to structural notions: invert the subject and the auxiliary, insert a question particle in place of the argument being questioned, etc. For illustration, here are the rules proposed, as applied to the declarative sentence "Moshe will want to see the data.":

1. Moshe will want TO see the data?

2. Moshe will to want see the data?

3. Data the see to want will Moshe?

4. Will Moshe want to see the data?

5. Moshe will want to see what?

No known language uses processes like those in 1-3, whereas 4 and 5 are widely attested. Note the modal status of this claim, however. It is not claimed merely that no languages have in fact adopted structure-independent options, but that they *could not* do so.

Moving down a level, linguists want an account of how languages differ from one another. And again, we can answer such questions by appeal to the modal properties of these languages. What is possible in one language may be impossible in another. Investigation of these issues uncovers interesting patterns of similarity and difference. This can be brought out in the comparison of negative sentences in English and French. Both languages require that whatever is responsible for marking the tense of the sentence intervenes between the subject and the negation. However, French differs from English in allowing lexical verbs to play this role. In English, lexical verbs always follow negative expressions, and so auxiliaries are required to indicate tense. Witness the following paradigm.

E1 Peter plays football.	**F1** Pierre joue au foot.
E2 *Peter plays not/not plays football.	**F2** Pierre (ne) joue pas au foot.
E3 Peter will not play football.	**F3** Pierre (ne) jouera pas au foot.
E4 Peter should not play football.	**F4** Pierre (ne) devrait pas jouer au foot.
E5 Peter does not play football.	

We see here in the first row that when there is no negation, tense is marked on the lexical verb. In the second row, when the negation is introduced without an auxiliary, in English, we seem to face a sort of word-ordering paradox: tense marking must appear before the negation, but the lexical verb, on which tense is marked, must appear after it, and so no option is feasible. In French, however, the latter constraint is not present, and

so the tense-marked lexical verb appears before the negation without issue.[1] In English, however, this paradox is resolved by the introduction of an auxiliary verb (E3-5), which can appear before the negation and indicate tense. No such option is needed in French (F3), but when an auxiliary is independently needed (F4), it patterns just as in English. This pattern of similarities and differences can then be used to explain other phenomena, such as the dummy verb 'do' found in E5: while this verb plays no semantic role in the sentence, it is needed to resolve the ordering paradox. As this paradox doesn't arise in French, we see no analogous phenomenon in this language.

Theoretically, we can explain these patterns by assuming that the ordering of tense, negation, and verb(s) is fixed, in that order, in both languages. The languages differ on whether lexical verbs can be moved from their initial position to the tense location. In French they can, and so we see sentences like F2, with the tensed verb before the negation. In English they cannot, resulting in E2's ungrammaticality. Auxiliary verbs, on the other hand, are free to move in both languages, and so we see auxiliary verbs mandatorily indicating tense in English, as the lexical verbs cannot, and doing so in French when there is an independent reason for them to be there. Once again, these rules govern what *can* happen in these languages, not merely what does happen.

Hopefully these brief examples of linguistic theorizing make clear the general strategies used for describing and explaining linguistic phenomena. Thus far, the epistemology of these scientific practices should likewise be relatively clear: linguistic observations are gathered, and rules are proposed which capture these observations. A good (i.e. well-confirmed) rule is capable of predicting both which expressions are acceptable in these languages and which are not.

One important point which these brief examples bring up is the interdependence between theory and categorization. When we propose hypotheses for explaining the observed data, we thereby introduce our theoretical predicates. The success of our theorizing thus depends on our having correctly identified the appropriate ways to classify our target. If our generalizations do not mesh with the data, one possible explanation is

[1] As 'ne' is largely optional in spoken French, I am assuming that the essential negative particle is 'pas'.

that the predicates we used to state them do not correctly classify our target phenomena. One kind of criticism of a proposed rule simply says that the rule is false. Another, more radical, criticism says that the very classification scheme adopted in the statement of the rule is misguided. For example, it is one thing to argue that there are some linguistic rules which appeal to linear order of constituents, and thus that structure-dependence is not a universal property of human language. It is another to say that viewing languages as systems composed of rules defined over constituents is a mistaken approach to linguistics.

With this preliminary model of linguistic theorizing on the table, we can turn now to the form of defective data I will be concerned with: mere performance data. These data arise when the above method breaks down. That is, when the simple comparison between the predictions suggested by hypothesized rules and observed linguistic behaviour becomes unsuitable for confirming linguistic hypotheses. In some circumstances, hypothesized rules are retained even in the face of apparently recalcitrant data, not merely because theorists don't know how to propose rules which don't face this problem, but because the data is taken to be simply irrelevant. Linguists call such data 'mere performance data'. One more example will demonstrate a typical pattern.

One of the core tools for generating expressions in any natural language is coordination. By coordinating two expressions, we can generate a larger expression with the two smaller expressions as constituents, and we can do so recursively. Expressions of many types can be coordinated, paradigmatically through conjunction: noun phrases ("Pilots and flight crew may use the lounge"), prepositions ("We walked up and down the staircase"), verbs (She was laughing and crying), etc. However, there are linguistic constraints on which expressions we can conjoin in this fashion. Most generally, it appears to be a rule of all languages that coordinators can combine only expressions of the same type. An adverb and an adverb may conjoin to form a complex adverbial phrase, for example, but we cannot conjoin an adjective and an adverb (*"I painted the house blue and quickly"), or a preposition phrase and a complementizer phrase (*"I believe in God and that snow is white").

Thus far, this theorizing seems parallel to that discussed earlier: a rule is proposed,

and the data we have thus far marshaled is fully compatible with it. However, further data collection throws up several problems. For example, in certain conditions, we seem able to coordinate linguistic expressions which are not even grammatical constituents. The generalization just proposed said that two expressions could be coordinated only if they were of the same grammatical type, but expressions are typed only when they are constituents. Consider, for example, sentences like "Marco loved, but Antonio hated, the paper", an example of what linguists call a "Right Node Raising" construction. *Prima facie* what is being conjoined here are the expressions "Marco loved" and "Antonio hated". But, on standard grammatical assumptions, these are not constituents of any sentence, and thus lack a grammatical category. In a simple SVO sentence, the verb and object form a constituent, and then this complex constituent combines with the subject: [Marco [[loved] the paper]$_{\text{VP}}$]$_{\text{S}}$. This expression's status as a non-constituent is further evidenced by the fact that it is very difficult to refer back to whatever such an expression would denote with a pronoun, or to move this expression around in the way that traditional transformational rules do. So, coordinate structures of this sort, apparently conjoining non-constituents lacking grammatical categories, seem inconsistent with the generalization that coordination can apply only to expressions of the same type.

There are, of course, several possible responses here. One could, of course, simply accept that these data falsify the generalization, and thus look for alternatives. The problem with this proposal is that it leaves all the data which did seem well accounted for by the rule unexplained. Further, this response is radical just in the way mentioned earlier: it doesn't merely require that this rule be false, it undermines the very classificatory scheme on which this linguistic approach is based. It is central to this style of linguistic theory that constituents are what rules can apply to. Lexical heads and complete phrases are the items picked out by every linguistic rule proposed within generative grammar. Opening up the doors to rules applying to non-constituents leaves linguistic theory massively unconstrained, and thus removes its ability to explain why the vast majority of conceivable linguistic structures are never attested in any human language.

There are, however, less radical strategies. One could, for example, claim that the

appearance of such sentences is misleading. While they appear to involve non-constituent coordination, they should really be analyzed as instances of different sorts of phenomena, such as ellipsis. "Marco loved but Antonio hated the paper" does not, on this proposal, really involve coordinating the non-constituents "Marco loved" and "Antonio hated". Rather, this sentence is an elided form of the perfectly kosher sentential coordination: "Marco loved the paper but Antonio hated the paper".[2]

Such proposals may work for many examples of this sort, but, for both empirical and theoretical reasons, cannot be the whole story. Ellipsis is itself a highly constrained linguistic mechanism, and so proposing ellipsis creates predictions of its own. In these cases, these predictions are not met. Consider the sentence "Marco loved, but Antonio hated, his haircut". This sentence is ambiguous: it can mean that Marco and Antonio held these attitudes towards their respective haircuts, or that both held them to Marco's haircut. On the other hand, the unelided sentential conjunction: "Marco loved his haircut but Antonio hated his haircut" is, to my ear, unambiguously interpretable on the latter reading.[3] As, on the account under consideration, the former is simply an elided version of the latter, what explains this difference?

Perhaps more significantly, going down this path leads again to the radical response to linguistic theory detailed above. Insisting that our linguistic rules capture all such data opens the door to kinds of data which no linguistic theory of this sort can, in principle, handle. Just as we can (apparently) conjoin non-constituent expressions, we can conjoin sub-expressions. Take the sentence "Marco was under-, but Anastasia was over-, whelmed by the play." What is conjoined here are not merely non-constituents, they aren't even complete words! Any grammatical theory which allowed for mechanisms to operate on sub-expressions in this way would be radically inconsistent with the vast majority of

[2]There are, as always, other options. Most prominently, these constructions have been argued to result not from ellipsis but from movement of the shared object of each clause to a position to the right of the conjoined expressions, hence the name 'right node raising' coined in Ross (1967). One major problem with such proposals is that there are empirically and conceptually motivated reasons to think that rightwards movement of this nature is generally prohibited (see **?**). See Sabbagh (2014) for a reasonably up-to-date survey of linguistic options here.

[3]There are also readings where the possessive pronouns are interpreted deictically, to refer to salient males in the speech environment, and these males may even be Marco or Antonio. Such readings are not relevant to the point of this example.

theoretical linguistics from the past hundred or so years. It would also face massive problems with over-generation, without even a principled way to preclude grammatical operations applying to isolated syllables.

For these sorts of reasons, it is often better, rather than revising one's theory to account for all the observations, to simply exclude these data from the confirmation base of our theory. The traditional guise of such a move in linguistics goes under the banner of the 'competence-performance distinction'.

The competence-performance distinction (hereafter CP) is an intuitive one, but its significance is often under-appreciated. At core, it is the distinction between the rules of the language and actual linguistic behaviour. Just as we can distinguish the laws of the land from the actions of its citizens, and the two do not align perfectly, we can distinguish between the rules of the language and actual speech, and the two need not generally coincide. Stated this blandly, no-one could dispute that there is a distinction here. Most obviously, speakers can often recognize when their own speech deviates from the rules they know to govern their language. Slips of the tongue, spoonerisms, mispronunciations, malapropisms, and other linguistic errors are commonplace, and are typically easily identifiable even by their producers. So everyone recognizes that the laws governing the language are no more universally adhered to than the laws governing the highway.

However, as the majority of linguistic rules are not consciously available to even competent speakers, there is no requirement that disparities between competence and performance must be obvious to untrained speakers. Thus, there is no reason in principle why grammatical theory should not posit rules which exclude sentences normal speakers take to be unproblematic, and vice versa, intuitively terrible sentences can be perfectly compatible with the rules of grammar.[4] Linguistic competence is just one determinant of performance, and other influences can create all kinds of complicated mismatches between the two.

In the case described above, then, non-constituent coordination presents an option:

[4]Interesting examples of the latter phenomena may come from widely accepted, but false, prescriptive assumptions about language. Prescriptivists often report feelings of ill-formedness concerning perfectly standard linguistic phenomena, even when they themselves make use of them. See many of the essays in Liberman & Pullum (2006) for some amusing examples.

radically revise our linguistic theories, or view such linguistic behaviour as 'mere' performance, i.e. performance not reflective of genuine linguistic competence. If this reflects the influence of extra-linguistic factors on linguistic behaviour, it poses no more of a problem for our proposed linguistic rule than do slips of the tongue. Of course, the CP distinction does not provide *carte blanche* to linguists to dismiss apparently tricky counter-examples to their preferred theory. To propose that some linguistic behaviour is mere performance is to propose that its properties are explained by something other than competence, but these properties must still be explicable by something. And so such claims are empirical hypotheses in their own right.[5] But it is clear that this is at least an available option. When linguistic data doesn't match hypothesized theory, we can always deny that this behaviour is appropriately connected to the genuine theoretical target: competence.

To push our legal analogy, we can compare linguistics to the project of someone in a foreign land, attempting to infer the laws of the land from observed behaviour. Such a task is possible only if there is *some* correlation between the two. If no-one gave any heed whatsoever to the rules, this would be impossible. However, it does not require that everyone abide by the rules fully at all times. And indeed, even uniform behaviour does not guarantee a law governing such behaviour. What is needed is a detailed examination of observed behaviour, alongside an assortment of causal hypotheses concerning such behaviour. Some generalizations can be accounted for without reference to the laws: that most people go to sleep at night and wake up in the morning need not be mandated by law. But in other cases, behaviours that one might assume to be common in the absence of a law are rarely observed, making the positing of a rule prohibiting them plausible. And even behaviours which are only rarely observed may be attributable to a (lightly enforced) law if nothing else can explain their occurrence, such as the small number of cars which stick to the speed limit on the motorway.

Likewise, in linguistics, we need not expect perfect conformity between the rules of language and linguistic behaviour.[6] Rather, linguists perform complex pieces of causal

[5] See Dupre (2019) for extended discussion of this point.

[6] Indeed, as we shall see later, conformity between competence and performance is not merely empirically unfounded but is in fact a sort of category mistake. Competence governs a psychological system which cannot generate utterances by itself, but only in concert with myriad other psychological systems.

reasoning, seeking to account for why we say what we do, and why we don't say what we do not. No linguistic rule needs to be postulated to explain why no spoken sentence contains 100,000 words. However, that most, but not all, uses of coordination serve to combine linguistic constituents seems to call for a linguistic explanation. Of course, all such investigation is defeasible and subject to further testing, but so long as the possibility of explaining linguistic behaviour with reference to non-linguistic causes is a live one, we seem committed to the CP distinction.

One key disanalogy between the linguistic and legal case is that the latter is indisputably normative. Laws tell us what we *should* do (in some sense of the term), whereas I am assuming, with the majority of linguists and cognitive scientists, that linguistic rules are, *pace* many followers of Wittgenstein, merely descriptive. The CP distinction, however, relies on distinguishing between description of *language* and description of linguistic behaviour. On this purely descriptive view, the aim of a theory of competence is to describe the constitutive rules of natural language and natural languages. To violate these rules is not to do anything *wrong* or to fail in any way, which is fortunate, given how large a proportion of perfectly natural speech appears to do just this.

Having provided a sketch of the methods adopted in linguistic theorizing, centering on the distinction between the target of inquiry, competence, and the primary evidence base for confirmation, performance, I will turn now to one of the central debates in philosophy of linguistics: the psychological reality debate. I shall argue that while the opposing sides of this debate have claimed neutrality concerning the first-order work in linguistic science, only one position is in fact compatible with the methodological approach just sketched. Namely, one can only sensibly distinguish competence from performance if one views linguistic theories as theories of one aspect of the minds of human speakers.

3 The Psychological Reality Debate

The reaction to generative linguistics, from analytic philosophers, has been mixed. As a first-order theory, philosophers have not, for the most part, objected to this work.

Reasonably enough, philosophers have been content to let linguists answer questions about the structures of linguistic expressions, the rules governing them, the kinds involved in stating these rules, etc. without their input. However, second-order questions about how whatever theories linguists end up adopting are to be interpreted have long been disputed philosophical territory.

Chomsky's position, developed most explicitly in Chomsky (1965, 1980, 1986), has consistently been that linguistic theories are psychological theories. That is, a true generative grammar describes the workings of some aspect of human psychology, which plays a central causal role in the acquisition and use of natural language. If English differs in some way from Mandarin, this difference in grounded in differences in the minds of speakers of these respective languages. Just as a vision scientist aims to discover the rules governing the visual system's transitions from retinal simulation to 3-D representation of the distal scene, a linguist aims to discover the rules governing those aspects of the human mind which enable us to acquire and use language. This thesis is typically stated in shorthand as the claim that linguistic rules, as stated in linguistic theories, have, or at least purport to have, *psychological reality*.[7] Following standard practice, I will describe psychologically real interpretations of generative grammars as 'cognitivist'.

This cognitivist interpretation has proved highly controversial in philosophy. While various arguments have been made against viewing linguistics as a branch of psychology, and various alternatives have been proposed, the common core of all such worries is that linguistics underdetermines psychology.[8] Consider the methodology described earlier, involving comparing the expressions generated by a collection of hypothesized rules with the expressions taken to be legitimate by a native speaker. In this way, the confirmation of a linguistic theory, it is argued, can thus distinguish between two theories only when they differ in the set of expressions they generate. However, for any such set of expressions,

[7] Within the camp of those who accept this claim, there is further debate about how it is itself to be interpreted. In particular, it has been vigorously debated whether these rules are themselves *represented*, or whether they merely govern the processing of other representations. The former proposal is commonly associated with Jerry Fodor (e.g. Fodor (1981, 2001)), while the latter has recently become more standard in philosophy of linguistics (see e.g. Collins (2014) and Pereplyotchik (2017)). See Stabler (1983) for a canonical discussion of the issues here.

[8] To my knowledge, the first place this worry was developed is Quine (1970), although it clearly traces back to Quine's earlier work on indeterminacy.

there will be numerous distinct sets of rules capable of generating this very set. Thus, there is nothing in linguistic theory to choose between such distinct, but 'extensionally equivalent' (equivalent in the set of expressions they license or generate), theories. But, it is assumed, there is a fact of the matter which of these distinct grammars is psychologically real. Given that standard linguistic methodology can thus not justify the selection of one such theory as psychologically real over any extensionally equivalent alternative, it is best to simply say that linguistics is not a science of psychological reality.

This core argument has been developed in several ways. Soames (1984) and Devitt (2006) argue that we should identify the goals of a science with reference to the kinds of evidence that they utilize and the kinds of claims that evidence of this sort will be liable to justify. As linguists tend to rely on behavioural evidence, their theories cannot be such that behavioural evidence will not discriminate between them. And the above underdetermination argument then purports to show that such behavioural evidence (i.e. that native speakers judge certain sentences to be well-formed, ambiguous, etc.) cannot distinguish between theories of which extensionally equivalent grammar is psychologically real. Thus, linguistics is not to be viewed as a branch of psychology.

Katz (1977, 1980, 1984), Soames (1984), and Devitt & Sterelny (1989) each take the argument a step further, and argue not merely that linguistics *can't* discriminate between psychologically distinct, but extensionally equivalent, grammars, but that even if it could, it *ought* not do so. Just as the functionalist movement in philosophy of mind has convinced many that psychological science requires a certain level of granularity in its classificatory schemes, such that mere neurobiological difference is insufficient grounds for distinguishing two mental states, so this line of argument claims that differences in psychological implementation are insufficient for distinguishing amongst the proper targets of linguistics: languages. Linguistics aims to describe the structural properties of languages (e.g. whether a language allows lexical verbs to occur before negative markers or not), abstracting away from the complexities involved when particular human beings produce and interpret these languages. One standard intuition-pump for this line of reasoning, found in various versions in the works just cited, has come to be known as the 'Martian

Argument'. Here is an early, and clear, statement of the argument from Katz (1977) (p. 576): "If linguists were saddled with [the commitment to view linguistic theories as psychological theories], their account of a natural language, say English, would have to reflect the idiosyncratic features of the speech production and comprehension mechanisms of human English speakers. Then, if there were creatures, Martians or perhaps porpoises, with sufficiently different speech production and comprehension mechanisms, but with whom we could communicate in English as easily as with any human English speaker, we would have to make the absurd claim that they do not speak English."[9]

This cluster of arguments has the negative goal of showing what linguistics is *not* (i.e. psychology), but different theorists have proposed different alternative accounts of what linguistics is. Defenders of these alternative views typically distinguish between our knowledge of language and the language itself. The two primary competitors, Platonism and nominalism, differ in what they view as the language itself, but agree that this differs from our psychological states that relate us to this object.

Nominalism, as defended by Michael Devitt (see especially Devitt (2006) and Devitt & Sterelny (1999, 1989).)[10], is the view that linguistic theories are about physical linguistic tokens: ink marks on paper, pixels distributed on computer screens, sound waves in the air, motions of the hands, etc. Linguistic rules, according to the nominalist, don't (necessarily) govern the psychological systems involved in the production and interpretation of these symbols, but govern the symbols themselves, specifying which such symbols are grammatical, what they mean, etc. The central analogy here is to chess: the rules of chess can be studied independently of the psychological processes involved in learning and playing chess. And creatures with very different psychologies could play chess if they could learn to act in accordance with these rules. Likewise with language.

Platonism, on the other hand, as defended most famously by Katz (1980, 1984), views

[9]The similarity between this argument and arguments for psychological functionalism involving Martians, robots, octopuses, etc. should be obvious.

[10]As Devitt notes (2006, p.30) it maybe be something of a misnomer to call his view nominalistic. While Devitt restricts linguistics to the study of public, physical objects, as in classical nominalism, he is not particularly concerned to deny that these physical objects have properties in some metaphysically robust sense. Despite this, the label has stuck, and is widely used both by Devitt (e.g. he refers to "my sort of nominalism" (fn. 13 on p.27 *op. cit.*)) and his opponents, and so I will continue to use it.

linguistics as a theory of *abstract* objects, akin to pure mathematics. There are, on the Platonist view, psychological and empirical questions about which languages we in fact speak, but the languages themselves, the proper targets and truth-makers of linguistics, are independent of us and our behaviour. The central model here is logic or pure mathematics. It would be a serious confusion to confuse the truths of the formal sciences (e.g. Gödel's Incompleteness Theorems) with some facts about human psychology. Likewise, the Platonist warns against confusing truths about language (e.g. that "Anton broke the bed" entails that the bed broke) with facts about language *users* (e.g. that English speakers are indeed prone to draw such an inference).

Neither the nominalist nor the Platonist denies that there is a role for psychological theories of language (i.e. for psycholinguistics). They just deny that linguistics in general aims at this project.[11] Further, these non-psychological alternatives are not supposed to conflict with the first-order theorizing in linguistics. That is, both the theoretical results and the methodological approaches of practicing linguists (although not, of course, the truth of the meta-theoretical claims made by linguists concerning their enterprise, which often express commitments to cognitivism) ought to survive these revised interpretations of what linguistic science is about. Thus, if we cannot make sense of some aspect of linguistic theorizing without viewing linguistics as a branch of psychology, then such philosophical proposals about what linguists are doing are in trouble. I shall now turn to the argument that this is indeed the case, and that appeal to the competence-performance distinction is the sticking point.

4 Competence is Not Idealized Performance

The arguments above, against psychological interpretations of linguistic theory, stem, I believe, from a very common misinterpretation of the relationship between competence and performance. Once this misinterpretation is cleared up, it will be clear both why these arguments fail, and why the alternative interpretations they have been used to

[11] Again, the analogy to functionalism in philosophy of psychology is illuminating. Functionalists need not deny that neuroscience is discipline wort pursuing. They claim rather that theorizing in cognitive psychology is, to a significant degree, independent of neuroscientific results.

bolster cannot succeed. The central mistake made in these arguments, and in much of the philosophical discussion of linguistics, is to overestimate the closeness of fit between competence and performance. Given the widespread acceptance of this mistaken view, I will spend a little time spelling out how this mistake has been developed in the literature, what is wrong with it, and why it matters.

One standard suggestion is that competence is an idealized, or cleaned-up, version of performance. On this view, if we take the set of performance data, especially the set of utterances judged to be acceptable, and then clean up the data by excluding some observations as reflective of extra-grammatical influence (memory limitations, considerations of pragmatic appropriateness, etc.), we are left with what might be called 'competence data'. A competence theory then consists in the simplest account of the patterns extractable from this set.

This conception, taken literally, is an obvious non-starter. Competence is, and always has been, an aspect of a psychological capacity. Performance is, and always has been, behaviour. Viewing the former as an idealized, simplified, version of the latter, akin to the relationship between raw data and some sort of statistical summary shorn of anomalies, is a category mistake. Given the conceptual error involved in this proposed relation between competence and performance, I assume that participants in these discussions do not really endorse it.[12]

There is, however, a nearby view, mistaken but conceptually coherent, which is widely endorsed. This is the view that competence is a psychological capacity which would cause the production of competence data, were it not for extra-linguistic influences causing performance 'errors'. Understanding the competence-performance distinction in this way is perfectly understandable, having been suggested by Chomsky himself in early methodological discussions of generative linguistics. In Chomsky (1965), the canonical such discussion, Chomsky claims that under certain idealizations (perfect memory, no shifts in attention, errors in applying linguistic knowledge, etc.), "performance [is] a direct reflec-

[12] Although Stich (1973) seems to endorse this view of linguistic methodology when he claims that "The motivation for separating acceptability and grammaticality is *broad theoretic simplicity*." (p.803, emphasis in original).

tion of competence" (p. 4). Again in Chomsky & Halle (1968), another inquiry-defining text, it is claimed that "We may, if we like, think of the study of competence as the study of potential performance of an idealized speaker-hearer who is unaffected by such grammatically irrelevant factors." (p.3). But despite these endorsements, such a proposal is in no sense a commitment that generative grammarians need to make, and indeed they would be better off not making it.

The flaw in this proposal is the insistence on viewing the disparity between competence and performance as reflective of 'errors'. Competence is viewed as a specification of a complete natural language which, due to inherent and temporary limitations, our performance reflects in a degraded form, only allowing us to produce a finite subset of expressions generated by the competence, and including in addition a host of deviant expressions.

Despite Chomsky's apparent endorsement of this view, we can see that on his own preferred interpretation of linguistics, there is no reason to accept it. Competence, on his standard cognitivist reading, is one contributing cause of performance among many. There is no reason to single out this cause as providing the norm against which the output is to be compared. That is, there is no reason to view disparities between competence and performance as necessarily reflecting imperfections of some sort. Consider the case we discussed earlier, of non-constituent coordination. Utterances of sentences like "Marco loves, but Antonio hates, linguistics" are not in any sense *errors*, reflective of memory limitations, distraction, etc. But it does not follow from this that expressions with analogous structures *must* be generable by the rules of one's linguistic competence.[13] Of course, they may be, but it is a perfectly plausible empirical hypothesis that our ability to use such expression types reflects something beyond our grammatical competence. The mapping between generable grammatical structures and linguistic behaviour need not be at all simple, and so the inference from the seeming impeccability of an utterance to the claim that expressions of this type are grammatical is a risky one. Indeed it is perfectly

[13]Another way to put this point, in the technical terminology of Chomskian approaches, is that neither *observational adequacy* nor *descriptive adequacy* provide even weak constraints on grammatical theories.

possible that many of the properties of our everyday speech are neither errors nor simple reflections of grammatical competence.[14]

A simple thought experiment should drive home the point that there is no *in principle* reason why rules governing competence need not be reflected in performance. A group of speakers, perhaps dedicated anti-generativists, could, without much difficulty, agree to use language in violation of any proposed linguistic hypothesis. They could, for example, agree to ask one another questions by inverting the order of words of the corresponding declarative sentence.[15] If these speakers were vigilant in following this convention, their performance could contain only questions of this sort. Such a community should not, however, cast any doubt on claims about universal grammar (UG). When generativists talk about UG as determining the boundaries of 'possible languages' (see e.g. Moro (2016)), they are not making a claim about possible *performances*, but about possible competences.

It is crucial here to separate the epistemology from the metaphysics. If linguists discovered a community which asked questions in this way, this would be good *evidence* that the proposed theory of UG which precluded such operations was mistaken. However, this is because the assumption is that this rule is acquired in the normal way. If it was discovered that this rule was a *learned exception*, a productive idiom, the connection between performance and competence would be broken and the hypothesis that such a rule is not compatible with UG, despite being utilized, would be untouched.[16]

A helpful analogy can be made to the relationship between genotype and phenotype in biology. It had often been assumed that there was some sort of simple relationship between genotype and phenotype, even that the genotype serves as a 'blueprint' for the phenotype: with the genome specifying the macro-level traits of the organism. For various reasons (see e.g. Godfrey-Smith (1999) and Lewontin (2001)) this picture has been widely rejected. Genes are one determinant of behaviour among many, and the

[14]See Dupre (2020) for discussion of the idea that *everything* that differentiates one spoken language from another falls into this category.

[15]Of course, if Chomsky et al. are correct, this rule would not be acquired, at least not in the typical way, by children raised in such an environment.

[16]The famous 'Pirahã controversy' (see Everett (2005) and the response by Nevins et al. (2009)) seems to be plausibly analyzed in just this way.

inference from "x is a phenotypic trait" to the claim that "the structure/properties of x are specified by the genome" is, at best, a risky one (at worst, it is meaningless). I believe the assumption of a simple mapping between competence and performance should go the same way. Linguistic competence does not specify features of performance, and no special interference is needed to cause disparities between the two. Rather, linguistic competence provides one causal influence on performance, and any inference from one to the other requires empirical assumptions concerning a wide range of psychological systems.

5 Why Competence Must Be Psychological

Hopefully the above has shown the mistake involved in treating performance as merely a degraded reflection of competence. Competence contributes to performance, alongside many other influences, but need play no specially privileged role in doing so. This reconceptualization of the distinction thus opens the door for a much broader disparity between the two than is typically assumed. I shall now show how this traditional, but incorrect, view of the CP distinction has motivated non-cognitive views of linguistic science, and how such views cannot be maintained in the face of this revised conception.

I'll start with nominalism, as this view has received a lot of discussion in recent years, due centrally to its revival by Michael Devitt.[17] As discussed above, the nominalist views linguistic theory as a theory of publicly observable symbols. Such symbols can be grammatical, ambiguous, etc., and the job of a linguistic theory is to predict and explain which such symbols have which such properties. Devitt claims that whichever grammatical theory best explains/predicts the properties of public symbols, within some given language, provides an *hypothesis* for a psychological theory of a speaker's competence

[17]Devitt's proposal has come under heavy criticism. For example, Rey (2020b,a) claims that focus on public symbols would prevent linguistic theory from providing robust *explanations* for why these symbols are as they are. That is, patterns of (e.g.) grammaticality and ungrammaticality are grounded in the ways that human minds represent linguistic expressions, and so excluding psychological claims from linguistic theorizing leaves us able (at best) to describe these patterns, while of course the aim is to explain them. Similarly, Collins (2020) argues that while public expressions of language may vary radically from utterance to utterance, it is only in the underlying psychological system that robust invariances are found, and so linguistic theorizing must target the latter. I believe my argumentation here, centering on the idea that psychological theorizing is required to even identify the relevant linguistic phenomena, is distinct from, though complementary to, these objections.

in this language, but the argument from underdetermination shows that this hypothesis could be mistaken, and would require specifically psycholinguistic evidence (from reading time experiments, clinical studies of aphasia, etc.) to confirm.

It is clear that Devitt is assuming the traditional distinction between competence and performance, and that doing so is essential to his project. A representative quote from Devitt (2006): "So a theory of the outputs of a competence is automatically, to that extent, a contribution to the theory of the competence, for it tells us about the outputs the production of which is definitive of the competence. And we can say that a competence and its processing rules must "respect" the nature of the appropriate output in that, performance errors aside, the processing rules must produce outputs that have that nature." (p. 22). There are two crucial mistakes in this quote. Firstly, Devitt refers to public symbols (utterances, inscriptions, etc.) as 'outputs of a competence'. Secondly, he claims that a theory of these 'outputs' must be 'respected' by the competence.

The first point to note is that public symbols are no more outputs of linguistic competence than motion is an output of an internal combustion engine. Linguistic competence generates linguistic behaviour only in concert with many other cognitive systems, many of which are in no way language-specific. Ignoring the contribution these other systems make is precisely to assume the traditional picture of linguistic competence, according to which it is only when things interfere ("performance errors aside") that performance will differ from competence. But as we saw earlier, this is an unwarranted and implausible assumption.

This leads on to the next point: respect. The claim that psychological theories must respect grammatical theories, i.e. that whichever rules a theory of public linguistic symbols uncovered constrain a theory of linguistic psychology by ensuring that the 'outputs' of the latter theory have the properties specified by the former theory, is absolutely central to Devitt's conception. But again, this constraint makes sense only if we assume that the mapping between competence and performance is very tight. Once the influence of all the additional, extra-grammatical, systems on performance is acknowledged, there is no reason to assume that generalizations covering linguistic performance (even

modulo traditional performance 'errors') will be respected by the rules of the competence grammar. We saw above an example of this. Non-constituent coordination is frequent enough in linguistic performance, i.e. in public linguistic symbols. But it doesn't follow from this that a theory of our psychological grammatical competence must also allow for constructions of this form. That is, our competence theory *need not respect* these features of our public linguistic behaviour. If such public symbols are legitimized by these extra-grammatical influences, constraining our theories of competence in this way would lead to worse theories. Without the assumption that competence and performance are closely aligned, modulo some imperfections in the latter, Devitt's respect constraint is unmotivated, and inference from performance to competence requires empirical assumptions concerning the causal explanation of why this public symbol has the properties it has.

The nominalist could retreat to a weaker position by rejecting the respect constraint and simply allowing that there are two distinct projects here: describing public symbol systems, and describing psychological capacities. Each may be illuminating about the other, but there is no deep reason to expect this in general. This pluralist approach can be interpreted in several ways, however.

On one, weak, reading, the claim is just that linguists ought to describe both the publicly observable patterns in linguistic performance, and the psychological systems partially responsible for these patterns. On this reading, this claim is more-or-less uncontroversial: even cognitivist linguistics begins by describing linguistic performance and then inferring on the basis of these descriptions a plausible underlying system capable of explaining these observations. That is, the public symbols serve as the evidence base for the real theorizing, concerning psychology.

The stronger reading treats these public symbols not only as providing evidence, but as providing phenomena to be theorized about in their own right. This, I take it, is Devitt's view. The problem with this proposal is that it is not clear what such a theory would look like, or why we would want one. Given the diversity of causes of linguistic behaviour (not merely competence and error, as traditionally assumed), it is unclear that there

is anything very unified about such symbols. Some public symbols are the way they are because they reflect genuine grammatical rules, others reflect societal conventions (and may be consistent with grammatical rules or not, as we have seen), others reflect behavioural habits, historical quirks, features of memory or perceptual systems, and so on. For such a motley bunch of facts, we may be able to describe them, and perhaps even propose some rough taxonomies, but the prospects for a deep and interesting theory seem less than promising.[18] As we saw above, our grammatical theories can achieve a level of simplicity and generality (crucial for any theory of language *acquisition*) in virtue of our excluding apparently problematic data, such as non-constituent coordination, from their scope. But this move is justified precisely by the focus on just one causal influence of our linguistic performance, namely competence. When we shift the focus to the performance itself, this move, and thus the theorizing it allows, becomes unavailable.

Further, Devitt frequently motivates the need for such a public-language-centric approach to linguistics on the grounds that it is presupposed by the cognitivist account (see e.g. p.28 of Devitt (2006)). To theorize about competence, he argues, we must first have a theory of what competence produces, just as a psychological theory about how humans play chess presupposes knowledge of what is involved in playing chess, i.e. the rules. But as we have seen, there is no interesting sense in which these public symbols are products of competence, and so there is no reason to think a theory of public symbols is a key first step towards a theory of competence. So if the targets of such a theory, public symbols, are unlikely to form a natural kind, and a theory of these targets seems unlikely to illuminate competence, nothing seems to be gained in viewing public symbols as targets of, rather than mere data for, theorizing.

What all of this suggests is that, if the goal is developing a *science* of language, public language may just not be a very interesting object. Performance is an interaction effect of numerous different systems, and is thus unlikely to be susceptible to illuminating analysis. The significance of the CP distinction is thus not merely that such a target is a bit rough

[18] Millikan (1998, 2008) advocates for an account of (public) language as a 'system of criss-crossing conventions', which seems susceptible to this sort of description and taxonomy without much in the way of generality.

around the edges, and in need of cleaning up by removing anomalies and errors. Rather, it points to the fact that the goal here is to use this heterogenous mass of data, and isolate one causal factor relevant to it. This factor need not, and is indeed unlikely to, bear any strong resemblance to the complex set of observations we began with.

Platonism fails for a very similar reason. Namely, it provides no way for linguists to identify the kinds pertinent to their distinctive theorizing. This has long been a concern for the Platonist, as in their view the target of linguistic theorizing is abstract and thus causally isolated from our observations. This may not be a problem for Platonist accounts of mathematics, as it is widely assumed that there is just one set of mathematical objects, which all competent mathematicians are in contact with.[19] But in the case of (public) languages, there are, of course, many possible options. What, then, determines that the Platonic object of my linguistic knowledge is English (or British English, or British Standard English, or whatever), rather than, say, Dyirbal (or Warrgamay, or whatever)? For many commentators, it has seemed that this is where the action will lie (see e.g. Fodor (1981)).

One possibility, developed in Lewis (1975), is that the language we speak is determined by the conventions we operate under. In particular, conventions that I will only utter certain soundforms (or produce certain bodily motions, etc.) if I intend to communicate a particular thought, and likewise that I will take soundforms produced by others to communicate particular thoughts. This will, however, have exactly the same problems as the nominalist account. We can adopt conventions of all sorts in our linguistic behaviour. Remember our community who asked questions by reversing declaratives. Stipulating that any such conventions automatically become suitable targets for linguistic theorizing runs the risk of precluding any interesting theory construction in linguistics. It is only by focusing attention on the language faculty, which operates according to different principles from the rest of cognition and is plausibly largely uniform across the human population, that deep, simple, and general theories can be constructed. So again we see that the

[19] Although analogous worries pose perhaps the most famous problem for Mathematical Platonism. See Benacerraf (1965). It is also worth noting that the assumption that mathematics is itself monistic has recently been challenged in philosophy of mathematics. See e.g. Priest (2013).

directive to develop theories of psychology is not simply an expression of interest on behalf of Chomskian linguists, but a claim about where the suitable targets of inquiry are liable to be. Public symbols, governed by social convention, are simply too variable, and responsive to too many distinct influences, to be suitable targets for scientific inquiry.

What we want is a way of distinguishing observations which reflect an underlying, unified, target, suitable for abstract theorizing, from those which reflect myriad non-target influences and sources of noise. Theorizing about the latter is unlikely to get us anywhere. If conventionality cannot be used to draw this distinction, what can? One obvious answer is: psychology. One could propose a Platonism of this sort: linguistics aims to describe the abstract linguistic system which corresponds to the set of grammatical rules which constitute native speakers' linguistic competences. The problem with this proposal is equally obvious. The abstract object here is simply a third wheel, and the difficult and interesting work of linguistic science will be in figuring out which rules are operative in our linguistic competences. But this is just the cognitivist approach.

So neither conventions nor psychology will provide a suitable answer to the question of which (Platonic) language speakers know/speak. But without such an answer, the Platonist approach is empty. There are indefinitely many conceivable languages. What makes linguistics interesting is finding out which small subset of these are, or could be, spoken by actual humans. So, unless some other proposal can be made, and I know of no viable options, the Platonist approach seems to be a non-starter.

Both the nominalist and Platonist approaches are doomed to fail, as the underdetermination argument which bolsters them misconceives the relationship between competence and performance. These arguments assume that we first identify, by suitably cleaning up performance, the rules of the language. We can then ask what psychological systems enable us to know/speak a language of this sort, and the underdetermination argument consists in pointing out that multiple different rule systems would do this job equally well. But as we have seen, linguistic behaviour is a product of many systems, some distinctively linguistic, and others much more general. The goal of linguistics is to identify the former. This means that we cannot identify the rules of the language independently

of our hypotheses about how the mind works. That structures like "Marco hated, but Antonio loved, his haircut" are ungrammatical is not evidenced in our linguistic behaviour or judgements, but it may nonetheless be a deep fact about language. This fact, if it is indeed a fact, will be identified only by developing theories of the causes of our linguistic behaviour, and using these to classify our observations. So psychological theorizing is not some additional step, to be completed after the best linguistic theory has been identified, as assumed by nominalists and Platonists, but is involved at every stage of theorizing, from gathering and classifying data through evaluating and confirming theories. Only a cognitivist approach can make sense of this methodology.

6 Conclusion

In this paper, I have argued that generative linguistics provides an excellent case study for philosophers of science interested in the role defective data play in the sciences. Generative linguists have paid unusually high levels of attention to the classification of observations, in particular with reference to whether observations are indicative/reflective of competence, the target of their studies, or mere performance effects. In this way, the classification of data as defective in one way (anomalous/counter-exemplary) rather than another (mere performance) places a central role in guiding which theories are to be adopted. Further, attention to the role of this distinction in linguistic practice guides our philosophical theorizing about how exactly such theories are to be interpreted. Specifically, this methodology, of classifying data according to its causal antecedents, necessitates a cognitivist interpretation of grammatical theory.

References

Benacerraf, P. (1965). What numbers could not be. *The Philosophical Review*, *74*(1), 47–73.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. The MIT Press.

Chomsky, N. (1980). *Rules and Representations*. Columbia University Press.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. Harper and Row, New York.

Collins, J. (2014). Representations without representata: Content and illusion in linguistic theory. In P. Stalmaszczyk (Ed.) *Semantics and Beyond: Philosophical and Linguistic Inquiries*, (pp. 27–64). Walter de Gruyter.

Collins, J. (2020). Invariance as the mark of the psychological reality of language. In A. Bianchi (Ed.) *Language and Reality from a Naturalistic Perspective*, (pp. 7–44). Springer.

Devitt, M. (2006). *Ignorance of language*. Oxford University Press on Demand.

Devitt, M., & Sterelny, K. (1989). Linguistics: What's wrong with "the right view". *Philosophical perspectives*, *3*, 497–531.

Devitt, M., & Sterelny, K. (1999). *Language and reality: An introduction to the philosophy of language*. MIT Press.

Dupre, G. (2019). Linguistics and the explanatory economy. *Synthese*, (pp. 1–43).

Dupre, G. (2020). What would it mean for natural language to be the language of thought? *Linguistics and Philosophy*, (pp. 1–40).

Everett, D. (2005). Cultural constraints on grammar and cognition in pirahã: Another look at the design features of human language. *Current anthropology*, *46*(4), 621–646.

Fodor, J. (2001). Doing without what's within: Fiona cowie's critique of nativism. *Mind*, (pp. 99–148).

Fodor, J. A. (1981). Some notes on what linguistics is about. In N. Block (Ed.) *Readings in the Philosophy of Psychology*, vol. 2, (pp. 197–207). Harvard University Press Cambridge, Mass.

Godfrey-Smith, P. (1999). Genes and codes: lessons from the philosophy of mind? In V. Hardcastle (Ed.) *Where Biology Meets Psychology: Philosophical Essays*, (pp. 305–330). MIT press.

Katz, J. J. (1977). The real status of semantic representations. *Linguistic Inquiry*, *8*(3), 559–584.

Katz, J. J. (1980). *Language and Other Abstract Objects*. Rowman and Littlefield Publishers.

Katz, J. J. (1984). An outline of platonist grammar. In T. G. Bever, J. M. Carroll, & L. A. Miller (Eds.) *Talking minds: The study of language in cognitive science*, (pp. 17–48). MIT press.

Lewis, D. (1975). Languages and language. In K. Gunderson (Ed.) *Minnesota Studies in the Philosophy of Science*, (pp. 3–35). University of Minnesota Press.

Lewontin, R. C. (2001). *The triple helix: Gene, organism, and environment*. Harvard University Press.

Liberman, M., & Pullum, G. K. (2006). *Far from the madding gerund and other dispatches from language log*. William, James & Company.

Millikan, R. G. (1998). Language conventions made simple. *The Journal of Philosophy*, *95*(4), 161–180.

Millikan, R. G. (2008). A difference of some consequence between conventions and rules. *Topoi*, *27*(1-2), 87–99.

Moro, A. (2016). *Impossible languages*. MIT Press.

Nevins, A., Pesetsky, D., & Rodrigues, C. (2009). Pirahã exceptionality: A reassessment. *Language*, (pp. 355–404).

Pereplyotchik, D. (2017). *Psychosyntax: The nature of grammar and its place in the mind*, vol. 129. Springer.

Priest, G. (2013). Mathematical pluralism. *Logic Journal of the IGPL*, *21*(1), 4–13.

Quine, W. (1970). Methodological reflections on current linguistic theory. *Synthese*, *21*(3-4), 386–398.

Rey, G. (2020a). Explanation first! the priority of scientific over "commonsense" metaphysics. In A. Bianchi (Ed.) *Language and Reality from a Naturalistic Perspective: Themese from Michael Devitt*, (pp. 299–327). Springer.

Rey, G. (2020b). *Representation of language: Philosophical issues in a Chomskyan linguistics*. Oxford University Press.

Ross, J. R. (1967). *Constraints on variables in syntax.*. Ph.D. thesis, Massachusetts Institute of Technology.

Sabbagh, J. (2014). Right node raising. *Language and Linguistics Compass*, *8*(1), 24–35.

Soames, S. (1984). Linguistics and psychology. *Linguistics and Philosophy*, *7*(2), 155–179.

Stabler, E. P. (1983). How are grammars represented? *Behavioral and Brain Sciences*, *6*(3), 391–402.

Stich, S. P. (1973). Grammar, psychology, and indeterminacy. *The Journal of Philosophy*, *69*(22), 799–818.