Interpreting Mrs Malaprop: Davidson and Communication Without Conventions

By
Imogen Smith

A Thesis Submitted to
The University of Birmingham
For the degree of

Master of Philosophy

Department of Philosophy
College of Arts and Law
The University of Birmingham
December 2010

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Inspired by my reading of the conclusions of Plato's *Cratylus*, in which I suggest that Socrates endorses the claim that speaker's intentions determine meaning of their utterances, this thesis investigates a modern parallel. Drawing on observations that people who produce an utterances that do not accord with the conventions of their linguistic community can often nevertheless communicate successfully, Donald Davidson concludes that it is the legitimate intentions of speakers to be interpreted in a particular way that determine the meanings of their utterances.

This thesis investigates how we might interpret the non-standard utterances of another from the perspective of Davidson's anticonventionalist picture of language. It proceeds by investigating the possible roles for radical interpretation and triangulation in explaining successful day-to-day communication, where non-standard language use abounds, and argues that - even from a Davidsonian perspective - a full account of the linguistic skills and knowledge that a language user requires to interpret utterances must ultimately appeal at least to regularities word use and expressions, if not full-blown conventions.

For My Father

Acknowledgements

I would like to thank my supervisor, Dr. Darragh Byrne for his patience, comments and helpful discussion.

I would also like my mother, Prof. Margaret Smith, for her invaluable emotional and financial support during a very difficult time, and my partner Mr. Steven Allen for his patience, understanding and a thick skin during the period of writing.

Table of Contents

Chapter 0: Introduction **1**

0.1 A Hermogenean Frame *1*

0.1.1 The Outcome of the Debate. 8

0.2 Motivation *13*

0.3 General Focus of Thesis *15*

0.4 Scope and Limitations of the Thesis *20*

Chapter 1: The Necessity and Sufficiency of Convention for Communication **23**

1.0 Introduction *23*

1.1 "A Nice Derangement of Epithets" *23*

1.0.1 "Prior" and "Passing" Theories 26

1.0.2 Criteria for Successful Communication and Utterance Meaning 29

1.0.3 Lewis's Conventions and Davidson's "Anticonventionalism" 32

1.2 "There is No Such Thing as a Language" and the Necessity and Sufficiency of Knowledge of Conventional Meanings *39*

1.2.1 Malaprops, generalisability and First Meaning 39

1.3 *Lepore and Ludwig's Formulations of Davidson's Anticonventionalist Claims* *43*

1.3.1 Weak and Strong Sufficiency Claims 45

1.3.2 Weak and Strong Necessity Claims 48

1.3.3 Argument for 2W 49

1.3.4 Argument for *(2S)* 54

1.4 Summary of Discussion of Lepore and Ludwig's Defence of Davidson *57*

1.5 Conclusions: Sharpening the Terms of the Debate *59*

Chapter 2: Radical Interpretation and Innovative Utterances **61**

2.0. Introduction: *61*

2.1. Radical Interpretation *61*

2.1.1 Radical Interpretation and the Procedure of Radically Interpreting a Language 62

2.1.2 The Principle of Charity and the Procedure of the Radical Interpreter 64

2.2 Radical Interpretation, Conventions, and Mrs Malaprop *71*

2.2.1 Equivocations on Radical Interpretation 72

2.2.2 The Resources and Strategies of the Malaprop Interpreter 76

2.2.3 Limitations of the Radical Interpreter 83

2.3 Applying the Principle of Charity *85*

2.4 Further Difficulties for Radical Interpretation *90*

2.4.1 Timing issues: Humpty and Alice 90

2.4.2 The Observer Effect: Radical Interpreters and Speaker's Intentions 91

2.5 Conclusions *93*

2.5.1 Limitations: What Determines Meaning and What is Required for Interpretation? 95

Chapter 3: Speaker Meaning and Utterance Meaning **97**

3.0 Nice Arrangements and Nice Knock Down Arguments: Speaker Meaning and Word Meaning, First Meaning and Literal Meaning *97*

3.1. "Whose Meaning is it Anyway?" *98*

3.2 Davidson on Donnellan *100*

3.2.1.Does "Glory" Mean A Nice Knock Down Argument? 105

3.3 Do a Speaker's Dispositions Fix the Meanings of his Words? *107*

3.4 The Perspective of the Interpreter *112*

3.5 Conclusion *115*

Chapter 4: Triangulation and Interpretation **119**

4.0 Introduction *119*

4.1 Triangulation: Preliminary Remarks *121*

4.2 The Ambitious Project *123*

4.2.1 Triangulation and Getting Interpretation "Off the Ground" 123

4.2.2 Individuation of Content 126

4.2.3. Triangulating Non-occasion Sentences 130

4.2.4 A Supplementary Account of Triangulation 131

4.3. The Modest Project *135*

4.4 Conclusions *141*

Chapter 5 Conclusions **145**

5.1. Davidsonian Communication *145*

5.1.1 Radical Interpretation 145

5.1.2 Triangulation 148

5.2 General Conclusions *149*

Appendix: Plato, Davidson and Dialogue **152**

A.1 Habit and agreement in Cratylus *152*

A.1.1. The necessity of dialogue for interpreting others on the Hermogenean view 152

A.2. Dialogue and the Possibility of Disagreement *154*

A.2.1 Getting the philosophical discussion off the ground. 156

A.3. Avoiding Linguistic Anarchy: Triangulation and The Principle of Charity. *158*

Abbreviations **163**

Bibliography **165**

Psammetichus, when he was in no way able to learn by inquiry which people had first come into being, devised a plan by which he took two newborn children of the common people and gave them to a shepherd to bring up among his flocks. He gave instructions that no one was to speak a word in their hearing; they were to stay by themselves in a lonely hut, and in due time the shepherd was to bring goats and give the children their milk and do everything else necessary. Psammetichus did this, and gave these instructions, because he wanted to hear what speech would first come from the children, when they were past the age of indistinct babbling. And he had his wish; for one day, when the shepherd had done as he was told for two years, both children ran to him stretching out their hands and calling "Bekos!" as he opened the door and entered. When he first heard this, he kept quiet about it; but when, coming often and paying careful attention, he kept hearing this same word, he told his master at last and brought the children into the king's presence as required. Psammetichus then heard them himself, and asked to what language the word "Bekos" belonged; he found it to be a Phrygian word, signifying bread. Reasoning from this, the Egyptians acknowledged that the Phrygians were older than they

Herodotos II, ch. 2, Trans. A.D. Godley.

Chapter 0: Introduction

'There's glory for you!'
'I don't know what you mean by "glory,"' Alice said.
Humpty Dumpty smiled contemptuously. 'Of course you don't – 'till I tell you. I meant "there's a nice knock-down argument" for you!'
'But "glory" doesn't mean a "nice knock-down argument,"' Alice objected.
'When I use a word,' Humpty Dumpty said in a rather scornful tone, 'it means just what I choose it to mean – neither more nor less.'
'The question is,' said Alice, 'whether you can make words mean so many different things.'
'The question is,' said Humpty Dumpty, 'which is to be master – that's all.'
Lewis Carroll Through the Looking Glass.

0.1 A Hermogenean Frame

Plato's *Cratylus* is ostensibly a discussion about the right way to characterize the relation between words and the world. It pits two opposing accounts of names against each other; one account of names proposes that it is in some sense "up to us" what names we can use to pick out objects: the other holds that in order for a name to be an object's name, there must exist some natural fit, or correctness, between the two.

One interlocutor, Cratylus, holds the "naturalist" position, and a second, Hermogenes, the "up to us" account.[1] Socrates, as is his wont, claims not to know the truth of the matter, but promises to help both in their discussion. The structure of the dialogue appears *prima facie* to present Socrates as first rejecting Hermogenes' view, then rejecting Cratylus naturalist view resulting in a classic *aporia*: we must all go away and consider the issues further. However, the conclusions are more nuanced than this simple gloss would suggest. In order to see why this is, it is necessary to look in more detail at the positions being advanced.

[1] Hermogenes' position has been frequently labelled "conventionalist" but for reasons I will discuss shortly I avoid this terminology.

Because of the limitations of word count, beyond giving and outline of the naturalist position held by Cratylus (as fleshed out by Socrates) below, I will restrict my discussion to Hermogenes' position. At any rate, the naturalism of Cratylus is dead in the water as a viable philosophical account of communication and language, in addition to being of no relevance to this thesis. Hermogenes' view is often erroneously labeled "conventionalist", but this is anachronistic. He espouses no modern notion of conventionalism whereby the meanings of utterances are governed by the rules or norms of one's speech community.

When Hermogenes first explains the contrast between his own and Cratylus' views, there appears to be a fairly straightforward *nomos-phusis* ("convention"-nature) dichotomy between them. Cratylus' view "that there is a naturally correct name by nature for each of the things that are" is contrasted with the "*nomos*" orientated view which Hermogenes espouses: Cratylus holds that names are "not whatever some people call a thing by "making a compact" (*συνθέμενοι*) to call it so" (383a6-7).

But as Hermogenes continues, we see that he does not believe that some form of inter-personal agreement presiding over a whole language community governs what a word means (or what it can be used to mean). Strangely enough, it seems that one person alone can adopt and use a Hermogenean name successfully.[2]

> Her: Indeed, Socrates, I have already, on many occasions, held discussions with this man and many others and am convinced that there is no other correctness of names beside compact and agreement (*συνθήκη καὶ ὁμολογία*). This is because it seems to me that whatever name someone applies (*θῆται*) to a thing, this is the correct one. Furthermore, if he

[2] Lewis (1969: 182) perhaps would not have a problem with this: he at least is prepared to consider the possibility of a Crusoe language, although the user of a Crusoe language would have to have to secure some sort of "convention" between time-slices of himself.

> substitutes another name for himself (*μεταθῆται*), and no longer calls it by the former one, the later name will be no less correct than the earlier one - just as when we change the names of our household servant for ourselves (*μετατιθέμεθα*). For there is no correct name naturally existing for any one thing except in the "*nomos*" and habit (*νόμῳ καὶ ἔθει*) of those who (first) used it and those who continue to call it thus (*τῶν ἐθισάντων τε καὶ καλούντων*). If things are otherwise, then I am ready to listen and learn: not only from Cratylus, but also from anyone else.
>
> 384c10-e2

Before continuing, it is important to get clear on the terminology being used in this passage. Terms such as *νόμος*, *συνθήκη*, *ἔθος*, *ὁμολογία,* have all been translated *loosely* as denoting convention (albeit pre-theoretical), but to unreflectively translate them in this way is surely to pre-empt the purpose of the dialogue, part of which is to clarify the interlocutors' notions of such concepts. "*Sunthéké*" has at its root the meaning of "compounding" or compact, *homologia*, agreement; *nomos* can have a normative force, being translated as "law", "custom", "convention", but need not do (as Hermogenes' use here implies), and *ethos* (often translated "custom)" may be taken to imply a regularity, habit or wont on the part of an individual or community.[3]

It would be incorrect to saddle Hermogenes with endorsing the possibility of a Wittegensteinian private language (Wittgenstein, 1950: §258ff.). Hermogenes has no intention of suggesting that his own "language" cannot be made public. Certainly the

[3] Given the obvious allusions in this dialogue to the ubiquitous *nomos-phusis* debate of 5th and 4th century Athens, one might want to take *nomos* understood as defined in opposition to *phusis* at this point. The term *phusis* is broader than our term nature. On *nomos*, Guthrie's (1971: 55-6) definition is helpful here: he describes *nomos* as something that "presupposes an acting subject – a believer, practitioner or apportioner – a mind from which the *nomos* emanates." *Phusis* understood in opposition to this (as the two terms were often considered to be mutually exclusive) would then be something that did not emanate from the mind of any apportioner, hence if naming should fall into the realm of *phusis*, names would not be allotted by judgments or fiats of mankind. Interestingly, Socrates carves out a special significance for the use of *nomos* in this dialogue, playing upon its wide application in Greek usage, and this *does* appear to have normative force; however, in the final passages, the idea that *nomos* is responsible for the meanings of words (as opposed to mere *sunthéké* or *homologia*) is notably absent. See below for further discussion.

translatability of Hermogenes' "idiolect" into the names in common use by his peers is acknowledged by both Hermogenes and Socrates.

> Soc: What then? If I call one of the things that exist by a name, for example if I call that thing which we now call a "man" a "horse", and conversely that which we now call a "horse", a "man", would the name for the same thing be "man" in public, and "horse" in private? And in turn the name for the same thing will be "man" in private and "horse" in public? Are you saying this kind of thing?[4]
> Her: It seems that way to me.
>
> 385a6-b1

Not only can we see from this passage that both men acknowledge that the "private" language can be translated into the "public" language. Hermogenes may also be endorsing the notion that language use can be made appropriate to context: here perhaps the "public" and the "private" spheres (one can alter one's speech for the occasion). However, Hermogenes' view does seem to lack underpinning, the requirement for making such a language publicisable; we might suggest a co-operation requirement - some notion of the need or intent to make oneself understood. Hermogenes needs to factor in the point of using names and expressions: to make himself understood by, and so convey information to, those with whom he intends to communicate. It is not until later in the dialogue that he appears to come to understand as much:

> Soc: What do we do when we weave? Do we not separate out (*διακρίνομεν*) the warp and the woof that has been collected together?
> Her: Yes.
> Soc: And you'd be able to answer in the same way about the bore and the other things.
> Her: Of course.
> Soc: Can you say something in a similar fashion about naming? Since a name is a tool, what do we do when we name with it?

[4] Unlike its English derivations (such as "idiolect"), the Greek word translated as "in private", *ἰδίᾳ*, here need not have the implications of "for one person alone", very often in Greek usage, the private-public opposition is framed in terms of public life outside the household, private life inside the household and various permutations thereof. In the conversation following Socrates makes clear that he takes it in the sense of being particular to an individual, but this may be unfairly twisting Hermogenes' initial claim.

Her: I don't know how to answer that.
Soc: Don't we teach something to one another and separate out (διακρίνομεν) things as they are?
Her: Of course.
Soc: So a name is an instrument for teaching and separating out essences, just as a shuttle separates the threads on a loom.

388b1-c1

When one "uses names", one's successful use of them depends on the ability of someone to be "taught" by the names one uses. But does this depend upon one's using the naturally "right", or correct name? Socrates initially appears to think so, arguing as follows:

1) For any object, that object's nature exists independently of an individual/agent/perceiving subject. (386d8-e4)
2) The success of all actions (being such objects that stand in relation to objects in the world) is objectively determined by the objects in the world they stand in relation to. (386e6-387a9)
3) Speaking is a class of action and so determined by the constraints specified in (2) (387-b8-c4)
4) Using names is a species of speaking and so determined by the constraints specified in (2). (387d2-8)

A weak reading of this may merely suggest that there are objective criteria determining whether speaking (and by this let us say speaking meaningfully) has occurred. However, such a reading here would not be faithful to the text. It is clear that in what follows Socrates has in mind the idea that the object that a particular action is oriented to has a stronger constraining role on the success of that action. For example, suppose during an utterance I were to use a particular word to pick out an object, say, a dog. Now what constitutes a dog is not up to me, but then nor is the

linguistic activity required to pick out a dog. For Socrates, it is the objective nature of a dog that determines what name I can use to pick it out. If the long central etymological section is intended as an attempt to illustrate this claim, the determination goes right down to the phonetic composition of the word itself: if the letters “k y n o s” so concatenated are not appropriate to the creature in question, then one will fail in one’s attempt to use that word to communicate anything about that object – even if that is the word everyone uses for a dog.[5]

At any rate, Hermogenes refuses to be convinced by Socrates (although he does not know why). His scepticism of Socrates’ purported refutation of his own position forces Socrates to provide empirical evidence for the naturalist alternative.

> Her: Socrates, I can’t think of a way to counter you. Perhaps indeed it is not easy to be persuaded (from one’s position) so quickly. I think I would be more easily persuaded by you if you showed me exactly what you say this natural correctness of names is.
>
> 390e10-391a2

Thus at 391b Socrates begins to set out in detail a naturalist account of naming which proposes that the forms of names are determined by the natures of the objects that they name. There follows an etymological analysis of some 110 names, which can be catalogued into groups covering many aspects of the Greek conceptual scheme: proper names, gods, natural entities, ethical concepts and so on. The etymological investigation aims to provide empirical support for the naturalist thesis by demonstrating that names can be analysed into component names that give up a description of their nominata. Yet, fascinating, and occasionally plausible, as the etymological investigation is, it fails to prove by itself what Socrates intends; it does not demonstrate that names are naturally correct for their nominata. This is because

[5] See Fine (1978) for a weaker reading of the claim that objects themselves determine what names we can use to pick them out.

it does not show that names are related to their nominata at all: one name is resolved into a set of further names which must then themselves be subject to the same analysis, this process continuing *ad infinitum*. The most that the etymological investigation could show us would be that there is a certain consistency within language whereby the meaning of each name appears to be consistent with the meaning of its components. Socrates has a solution to this; however, this solution has seemed remarkably implausible to many. He suggests that at some point in our analyses of names we reach certain "primary" or "atomic" names which can no longer be subjected to etymological analysis. We have to account for how these names relate to their nominata in a different way. The relationship that holds between atomic names and their nominata, Socrates suggests, is a mimetic one: the sounds or phonemes that make up a name *imitate* the qualities that go to make up the essence of the nominatum.[6] A name can communicate or disclose an object to another because it is an imitation of its object (or its object's essential properties) in sound.

The mimetic thesis, then, proposes that a name picks out or refers to an object in virtue of being like it. Thus it should be clear that the component atomic names must *be like* their nominata if the resultant compound name is not only to be the "right" name for its nominatum, but also *to qualify for being a name at all*, since the principle of likeness (*mimesis*) exhausts the sign relation.

---

[6] Soc: Whenever we want to communicate (δηλοῦν) something with our voices, tongues or mouths, do we not successfully communicate by these means whenever an imitation (μίμημα) of an object arises through them?||Her: Necessarily, it seems. Soc: So it's likely that a name is an imitation in sound of that object which it imitates; and the one who imitates whatever he imitates in sound is using names (ὀνομάζειν). (423b4-10) Soc: If someone is able to imitate this very thing, the *ousia* of each object, in letters and syllables, will he be communicating each thing, what-it-is, or not?||Her: He will, of course. (423e7-424a1)

Once the details of a naturalist account of naming are down, the spotlight is now turned on Cratylus, who initially advanced the naturalist claim that there exists a certain natural correctness of names. Cratylus is wholehearted in his endorsement of Socrates' naturalist account; however, Socrates now claims that he is not so sure of his own arguments and suggests that they be put to the test. Much of the remainder of the dialogue (427d4-439b10) appears to be an exercise in subjecting the naturalist theory to a series of objections, with Cratylus taking up its defence. It is the observations that Socrates makes in refuting Cratylus that are of most interest to this discussion.[7]

0.1.1 The Outcome of the Debate.

The famous passage of the *Cratylus*, often referred to as the dialogue's *denouement* appears to show Socrates comprehensively rejecting Cratylus' naturalism. I take this rejection at face value; however, I believe that Socrates does not opt for any obvious notion of "conventionalism" in its place but rather returns to something very close to Hermogenes' original position.[8]

---

[7] It is not necessary to go into all of the objections here: some of Socrates' objections target naturalism specifically and point out internal inconsistencies in the position and so are of little interest here. Other arguments, however, appeal to observations on the day-to-day business of communication and are interesting to note as parallels to Davidson.

[8] The upshot of this passage for Socrates' final position is a matter of some controversy, with a wide range of formulations in the secondary literature. Numbered among those that opt for the conventionalist interpretation are: Bestor (1980), Robinson (1969), Schofield (1982), Levin (2001); for naturalism see Kahn (1973), Kretzmann (1971), Sedley (2003); for some compromise or aporetic positions see Anagnostopoulos (1971), Mackenzie (1986). Discussions of the possible positions can be found in Sedley (2003:143-51:), Schofield (1982: 72-3), Levin (2001: 90). See also Baxter (1992: p.187), who reads the dialogue as rejecting naturalism and Barney (2001: 136-7), who ascribes a "pessimism" about names to Socrates "which is reducible neither to naturalism nor to conventionalism; though it could arguably be seen as a combination of the two".

In this passage Socrates draws attention to a word ("*σκληρότης*" - "hardness") which has alternative pronunciations in different dialects, and which is inconsistent with the naturalist scheme set out in the dialogue so far. The phonetic composition of the name should render it sufficiently unlike its nominata to pick it out. Nevertheless, not only do we understand someone to mean *hardness* when they utter that word, it is clear we understand the word when pronounced as it is in the Eretreian dialect.

> Soc: What about when someone says *σκληρόν*, and pronounces it the way that we do at present? Don't we understand him? Don't you yourself know what it is that I am saying?
>
> 434b7-9

In asking this question, Socrates crystallises a shift in the focus of the investigation – from the meaning of words abstractly conceived (be that natural or conventional, or something else) to how we understand a word as actually used by people in actual instances of communication. The question now becomes one of the necessary and sufficient conditions for mutual understanding (Schofield, 1982: 78). Cratylus' response is to retreat to *ethos* as enabling understanding in this case. But Socrates points out that all *ethos* amounts to is whatever enables successful communication to take place.

> Cra: I do, but that's because of habit (*ἔθος*).
> Soc: When you say "habit", do you mean something other than *agreement* (*ξυνθήκη*)? Do you mean something by "ethos" besides this: when I utter this name, you know that that is what I have in mind? Isn't that what you're saying?
> Cra: Yes
> Soc: And if, when I utter something, and you understand, hasn't a disclosure occurred to you from me?
> Cra: Yes.
> Soc: Even though my utterance is unlike what I have in mind – since "*λ*" is unlike hardness (to revert to your example). But if that's right, surely you have entered into an agreement (*ξυνέθου*) with yourself, and the correctness of names becomes a matter of agreement (*ξυνθήκη*) for you, for isn't it the chance of usage (*ἔθος*) and agreement (*ξυνθήκη*) that makes both like and unlike letters express things? And even if usage (ethos) is

> completely different from agreement (*ξυνθήκη*), still you must say that expressing something isn't a matter of likeness but of custom, since custom, it seems, enables both like and unlike names to express things.
> 434b10-435b3

Thus regardless of whether some words do imitate their objects in the way that has been suggested, the etymological/mimetic thesis sheds little light on how one successfully communicates their thoughts to another. Using names that imitate their nominata is not a necessary condition for communicating one's thoughts about that object. The successful communication of what the speaker "had in mind" to the hearer has determined the meaning of the utterance in question.

Crucial for my purpose is what Socrates *does not* suggest here. Although the dialogue opened with what looked like two opposing views of names divided along traditional *nomos-phusis* lines, the claim that *nomos* determines the meaning of words is conspicuously absent here, with Socrates only claiming that *ethos, suntheke* and *homologia* are the determining factors in communicating successfully what one has in mind. The conspicuous absence of *nomos* is important, as this word has the strongest connotations of "rule" or "norm": by refraining from mentioning it here, Socrates seems to implicitly reject any notion that speakers will fail to communicate if they do not adhere to some norm of usage ranging over a linguistic community. This is particularly interesting since earlier, in conversation with Hermogenes, Socrates initially suggests that it is *nomos* that provides us with the names we use:

> Soc: Whose work does the teacher use when he uses a name?
> Her: I don't know.
> Soc: Don't you know this either: who provides us with the names we use?
> Her: Indeed I don't.
> Soc: Don't you think that *νόμος* is the provider of these things?
> Her: That's likely.
> Soc: So: whenever a teacher uses a name, he uses the name of the *νομοθέτης* (custom-establisher).
> 388d6-e3

At the close of the dialogue, however, it seems that if a speaker successfully gets across what she has in mind to a hearer then in some sense the meaning of the utterance is determined by this at this particular time. No mention is made of a wider linguistic community or any regularity of such a community that must be followed.

> Soc: Since we agree on these points, Cratylus – for I will put down your silence to agreement – it is necessary, surely, that both agreement (*ξυνθήκη*) and usage (*ἔθος*) contribute to the disclosure of what we have in mind when we speak. For, my dear Cratylus, if you are prepared to turn your attention to number, where do you think you will be able to bring in names resembling each one of the numbers from, if you do not allow this agreement (*ὁμολογία*) and compact (*ξυνθήκη*) to have *a* certain authority in regard to the correctness of names? I myself would prefer it (*ἀρέσκειν*) if names were like things as far as possible. But, truly, as Hermogenes says, this attraction of likeness is not worth much, and it is necessary to use in addition this vulgar thing, agreement (*ξυνθήκη*), in the correctness of names. But perhaps things would be spoken as finely as possible if their names were completely or mostly like the things spoken of, that is to say appropriate, and most poorly in the opposite case.[9]
>
> 435b3-d1

Thus the "correctness of names" turns out to be little more than what suffices for communication, and agreement and habits of usage suffice for communication. One could further suggest that there might be additional judgements about whether one has used the best or most appropriate expressions to communicate one's thoughts (Socrates says he would prefer this), but this is independent of what suffices for communication.

A few passages earlier Socrates asks Cratylus to consider the case of being greeted incorrectly by someone while travelling abroad. Taking Cratylus by the hand, the foreigner says "Hello Athenian friend, Hermogenes son of Smikrion" (429e2-5).

---

[9] Sedley (2003:147), notes that *ἀρέσκειν* is commonly used in Plato as endorsement of a doctrine and so should not be translated to imply that Socrates is presenting "little more that a pipe-dream". I agree, however, with Schofield's (1982) claim that "some authority" should be understood as dramatic litotes.

Cratylus responds, consistently with his own position, that such a man would be merely making unintelligible noises, equivalent to "someone banging on a brass pot" (430a7). But our intuitions suggest otherwise, as surely Socrates expects them to. In such a context, accompanied by the relevant extra-linguistic accoutrements (which Socrates takes care to specify), we could suppose that the foreign gentlemen has not failed to communicate what he intended to communicate.[10]

In response Socrates demonstrates further to Cratylus that one can successfully use a word even if it is used "incorrectly" according to some further criteria.

> Soc: What do you mean? How are they different? Is it not possible to approach a man and say, "this is your picture" and show him, perchance, a likeness of himself or, perchance, a likeness of a woman? And by "show" I mean to bring the likeness before his sense of sight.
> Cra: Well, yes.
> Soc: What then? Can I not again approach the man and say to him "this is your name"? I mean, a name is an imitation just as a picture is; can I not say to him "this is your name" and then bring before his sense of hearing, perhaps an imitation of himself, saying he is a man, or an imitation of the female of the human race, saying that he is a woman? Or does it not seem to you to be possible and to occasionally happen?
> Cra: I am willing to concede that you are right, Socrates.
>
> 430d8-431a6

In this passage, Socrates elicits Cratylus' agreement that it is possible to use the naturally incorrect name to successfully pick someone out. The point Socrates makes, however, appears to mark a distinction between success and correctness that shows that not only natural correctness *but correctness as determined by the*

[10] The context of Socrates' exchange with Cratylus is complicated by a complex discussion of truth and falsehood, and is not relevant to the current discussion. Smith (2008) provides a fuller analysis of the extra-linguistic factors, together with demonstratives, indexicals and grammar that Socrates introduces into his case studies to show how successful communication can be secured by non-standard words. A point I don't make there, but is worth making here is that of course, if the "natural" meaning of the words can be overridden by such factors, then the same goes for any subsitute "conventional" meaning.

conventional rules of a linguistic community is not required in the successful use of words. To say that one can use "the wrong" word to successfully express what one intends constitutes an attack on more than merely naturalism, but on any account of meaning whereby the norms of a community determine the meanings of words.

To claim that Socrates envisages an opposition between naturalism and conventionalism as we understand it is anachronistic. Socrates' concerns are the necessary and sufficient conditions for successful communication of what one "has in mind": nowhere does Socrates concern himself with the linguistic community at large. Socrates appears to be much more concerned with "getting away with it" than with any notion of correctness whereby people fail to communicate if they "get it wrong".

We might suggest that Socrates' final position comes down to a rejection of naturalistic constraints on what a word can mean, but in his alternative proposal, the determinants of meaning seem to be usage, or habit, and agreement. Socrates makes no mention of one's being responsible to *nomoi* at all: all that is required is that speaker and hearer agree on how the words are being used.

0.2 Motivation

I have above given an outline of a reading of Plato's *Cratylus*, that sees Socrates prioritise communication over any abstract notion of "what words mean". Neither nature or, for that matter, the norms of "convention" determine the meaning of a particular utterance. Nothing follows about whether there are conventions operating in a given community, or whether there are better and worse ways of constructing or using words. From what has been stated, it may even be the case that language could be revised to better reflect the world around us.

It is not my intention here, to fully justify the interpretation of the conclusions of the *Cratylus* I offer; a defence of this must take place another in arena. I have given an outline to provide my motivation for writing this thesis, which is a desire both to afford the position (I suggest) Socrates espouses in the *Cratylus* a modicum of philosophical respectability, and to investigate how such a position might contribute to, and be reinforced by, a wider discussion of Plato's preference for dialogue and dialectical exchange for conducting philosophical enquiry.[11] The question that frames this thesis (which I term the "Hermogenean question") is:

Does Plato's preference for dialogue fit in well with the view of language I have suggested Socrates endorses?

With this in mind, I turn my consideration to Davidson. Like Socrates in the *Cratylus* (as I believe), Davidson offers a model of linguistic communication whereby the meanings of expressions and utterances are governed not by convention, but rather the intentions of the speaker in the context of successful communication. Like Socrates in general, Daivdson appears to have some very clear views on the importance and role of dialogue (or, here, communication) in determining the meanings of utterances. An investigation of the latter's views, I hope, will throw light on the former's position. In the Appendix to this thesis I offer an outline of how the conclusions of this thesis on Davidson might help answer the Hermogenean question.

---

[11] The comparatively little work that has been produced on the *Cratylus*, with a few notable exceptions (e.g Barney (2001), Williams (1982), among others), has been produced by scholars of a largely philological persuasion, many of whom, it would seem, have not read beyond Russell. Many philosophers, and scholars of a philosophical bent tend to avoid the *Cratylus*, since they believe it contains little of philosophical worth (perhaps persuaded by the writings of the philologists).

0.3 General Focus of Thesis

In this thesis, I examine Davidson's anticonventionalism, and investigate what plausible account of linguistic communication can be offered on the back of claims that conventions do not determine the meanings of utterances.

A very rough and ready account of a Davidsonian rejection of conventionalism might be advanced as follows:

Davidson's conventionalist opponent supposes that conventions determine the meanings of expressions in a language and that therefore knowledge of those conventionally determined meanings is necessary (and perhaps sufficient) to interpret those expressions.

We might state this position in the following conditional form:

If conventions determine the meaning of expressions in a language, then knowledge of those conventionally determined meanings is necessary (and sufficient) to interpret sentences uttered in that "language".

Davidson's discussion in "A nice Derangement of Epithets" (henceforth, *Derangement*) aims to show that knowledge of the conventionally determined meanings of words is neither necessary nor sufficient for interpreting the meaning of a given utterance. Therefore, by *modus tollens*, conventions do not determine the meanings of that utterance.[12]

The situation is actually more complicated than this: I discuss the details at length in chapter 1 of the thesis. However, here I put the case like this in broad

[12] Please see Abbreviations, for a full list of abbreviations used to refer to Davidson's essays.

outline, because I think it helps us to focus in on the central underlying question of this thesis.

For conventionalists of most stripes, the relation between what determines the meaning of expressions in a language and how we understand the expressions of a language is, in broad outline, more or less straightforward. Conventions figure substantially both in the account of meaning and in the account of how we can come to understand those meanings. Fortunately, conventions are public by nature; they can accessed by closely attending to the behaviour of their participants, or by merely consulting them (or dictionaries etc.). We might say then, that the relation between conventionalist accounts of meaning and conventionalist accounts of understanding exhibit a certain symmetry.

Davidson's alternative model of what determines the meaning of an utterance appeals to the intentions of the speaker to be understood in a particular way (subject to certain modifications: see chapter 1). We might state Davidson's alternative thesis as follows:

> **The meaning of a given expression is determined by the (legitimate) intentions of the speaker to be interpreted in a certain way when he utters that expression.**

However, a symmetrical account of understanding cannot obviously be provided. We might suppose that this would run along the following lines.

> **Knowledge of the speaker's (legitimate) intention to be interpreted in a certain way when he uses a given expression is necessary and/or sufficient for understanding that expression, so uttered.**

We can see how the problem arises. While it may be true that if I knew what someone's relevant intention was when they uttered an expression I would thereby know what their utterance meant, this cannot be the account we offer of how we

understand utterances of others. In the first place, it puts the cart before the horse: we interpret expressions to find out what a speaker's intentions are, not vice versa. The capacity to directly access such intentions would render linguistic communication redundant. In the second place, and more importantly, unlike conventions, intentions are not by their nature directly publicly accessible. We are not telepaths: access to a person's intentions is mediated by our observation of their behavioural manifestations.

Prima facie, this account of what determines the meanings of utterances, and how we come to understand those utterances does not display the easy symmetry of the alternative conventional accounts. This is not, of course, to deny that they are deeply intertwined. Meaning, for Davidson (e.g. *CToT*: 148; *SCoT*: 245), must be publicly accessible: it is open to public determination. Thus the possibility of a hearer or interpreter being able to grasp the meaning of an utterance constrains what one can *legitimately* intend one's utterance to be interpreted as meaning. One cannot, then, quite like Humpty Dumpty in the epigram at the start of this chapter, make one's words mean whatever one wants; one is constrained by the capacities of one's audience to grasp the intended interpretation. Successful interpretation, or at least the reasonable possibility of it, underlies the account of meaning.

In this thesis I am sympathetic to Davidson's claim that conventions (and for that matter prior regularities) do not determine the meanings of utterances, but rather that the (legitimate) intentions of the speaker do. This is in part due to the fact that I find the arguments presented persuasive myself, but also because I am seeking to investigate what I label above the Hermogenean question: Davidson's

anticonventionalist claim here seems a useful starting point, because it shares relevant similarities with the view I attribute to Socrates at the end of the *Cratylus*.

My primary interest, and the primary focus of the investigation of the thesis, is the interpretation question. To elucidate the claim that one can mean by an utterance what one legitimately intends by an utterance, we must investigate the question of how an interpreter can come to grasp the meaning of an utterance. Thus the overarching question of this thesis is as follows:

How can an interpreter grasp the meaning of a non-standard utterance?[13]

My approach to answering this question can be seen in the outline of the chapters as set out below.

In chapter 1, I give an overview of the argument of *Derangement*. I then discuss Davidson's anticonventionalist claims in more detail, and suggest that Davidson is advocating more than a mere rejection of Lewisian conventions (often supposed to be his main target); Davidson's claims not only rule out these as determinants of utterance meaning, but also any pre-theoretical notion of convention, as well as past linguistic regularities specific to the utterance at hand. I then critically analyse Lepore and Ludwig's (henceforth *L&L*) (2005) analysis of Davidson's anticonventionalism and suggest that this, by and large, fails to hit the mark; nevertheless, my criticisms of *L&L* are useful, I believe, in sharpening the terms of the discussion.

In chapter 2, I examine radical interpretation. When initially proposed by Davidson, a radical interpreter, who interprets an alien language from scratch

[13] In this thesis I use the terms "non-standard" and "innovative" to describe utterances or words that an interpreter is not prepared in advance to interpret. This is distinct from a 'novel' utterance, which the interpreter is prepared to interpret in advance although he has not met that utterance previously.

(without the help of translation manuals or other linguistic aids), could have been envisaged as constructing a theory that mapped the *stable linguistic conventions* of his target speech community. Now that conventions are out of the running, the possibility of the radical interpreter's being able to interpret innovative utterances of the sort Davidson proposes in his attack on conventionalism must be examined further. My investigation points to a negative conclusion: the radical interpreter would not necessarily have the resources to interpret an innovative utterance, both in principle, and in practice.

In chapter 3, I examine a distinction between speaker meaning and utterance meaning that is relevant to the conventionalism-anticonventionalism debate. It is key to Davidson's argument that a speaker's *utterance* has a non-standard meaning, rather than that the *speaker* means something non-standardly when so uttering. After all the conventionalist may retort that *Mrs Malaprop*, in uttering "A nice derangement of epitaphs" might mean *A nice arrangement of epithets*, but her words do not. I suggest that Davidson's view of language provides him with good grounds to claim that the non-standard meanings can be attributed to a speaker's utterances and not merely the speakers themselves. However, I also suggest that in many cases of interpretation of non-standard utterances, communication is successful because an interpreter first grasps the speaker's relevant intentions, rather than the meaning of the utterance itself. Even in cases where this happens, there nevertheless remains the question of how this can be done.

In chapter 4, I examine whether Davidson's triangular externalism can provide a positive account of how we can interpret others, particularly in cases of non-standard utterances. I suggest that trying to press the account of triangulation into

service of directly explaining how such interpretation can succeed (I call this the ambitious project), results in failure; however, reflection on triangulation could be fruitful in indirectly providing an account of successful interpretation (the modest project). Although there are details to be fleshed out (beyond the scope of this thesis), triangulation may figure usefully an account of the development of a linguistic or communicative subject who consequently has the relevant linguistic and non-linguistic skills to deal with non-standard utterances. However, this requires that regularities, and possibly some notion of convention (albeit a much deflated one), are let in through the back door.

At the end of this thesis I return to the framing question of the relationship of Plato's views on language to his preference for dialogue and suggest ways in which the investigations of this thesis may help us in outlining an answer.

0.4 Scope and Limitations of the Thesis

Davidson is known for being one of the most systematic philosophers of recent years. His views on language, epistemology, mind and metaphysics are so closely connected that it is difficult to separate out one strand without taking into account much more. However, due to the limitations of word count, there must be restrictions on which of Davidson's wide-ranging discussions can be brought to bear in this thesis. I therefore limit critical discussion as much as possible to issues directly related to the questions at hand, focusing as far as possible on issues raised directly in Davidson's work that specifically deal with communicating without conventions. Naturally, other areas of Davidson's work will be relevant, but in those cases I attempt to take them at face value: they cannot be subjected to detailed critical scrutiny themselves in this arena.

There is one problem lying underneath the surface of this thesis that I confess I have not been able to resolve to my own satisfaction. In my defence, I do not think the question is an easy one, and its difficulties have led to confusion on the part of many. However, it is important to acknowledge this as a problem and explain the approach I take in dealing with it as best I can. This a very general exegetical question concerning Davidson's discussions of interpretation and their relation to questions of how we actually interpret the utterances of others. In *SAL,* Davidson protests that he has been misrepresented by Dummett (1986), who sees him as trying to attribute theories of meaning to actual speakers and interpreters.

> I do not think I have ever conflated the (empirical) question how we actually go about understanding a speaker with the (philosophical) question what is necessary and sufficient for such understanding. (*SAL*: 111)

Dummett, according to Davidson, has misunderstood him. But it is easy to see how such misunderstandings arise: Davidson appears to equivocate in many places, particularly on the *actual* possibility of radical interpretation, for example, and the question of the extent to which a speaker and interpreter "have" theories.[14]

Davidson's attacks on traditional notions of languages and the role of convention therein are born out of *actual* cases of successful interpretation of non-standard utterances, and the account of triangulation is illustrated by, and gains support from, plausibly realistic, if not real, interactions between infants and mature language users.

Rather than try to resolve these apparent inconsistencies, I take the approach that I think most suits this investigation. Intuitively, it seems that if we want to say that the meanings of *actual* utterances are not determined by conventions, but by the

[14] A significant part of Dummett's (1986) response to *Derangement* appears to be based on attributing to Davidson the view that speakers and interpreters have "theories" of each other: a position which Davidson denies.

(legitimate) intentions of speakers to be interpreted in a particular way, we must focus on the question of whether and how *actual* interpreters can come to grasp those meanings. This I believe can provide important insights into language, because it is surely in large part down to the epistemic and cognitive limitations of the human condition that human languages are the way they are. Given this, the helpfulness of the suggestion that we could learn something about language from consideration of, say, how omniscient beings might be able to communicate with each other is debatable.

That is not to say that questions of what is in principle possible might not prove fruitful (and I do address relevant theoretical issues where relevant), but I aim to relate these questions back to actual (or realistic) language use, and the question of what is necessary for actual communication.

Chapter 1: The Necessity and Sufficiency of Convention for Communication

1.0 Introduction

The first part of this introductory chapter is primarily exegetical, I give an outline of *Derangement* and draw from it some important questions and points for consideration, including:

1) What counts as a legitimate intention to be interpreted in a particular way? And,

2) How is Davidson's anticonventionalist stance best characterised?

In the second part of the chapter I investigate and Lepore and Ludwig's (2005) analysis of Davidson's claims, and suggest that it does not really hit the mark. However, an investigation of their analysis serves to sharpen the issues at stake.

1.1 "A Nice Derangement of Epithets"

In *Derangement*, Davidson draws attention to a phenomenon of communication arising from errors of deliverance such as malapropisms. Davidson argues that the successful decipherment by a hearer of what a speaker intends her utterance to be understood as meaning when she utters, for example, a malapropism, cannot be accounted for by a picture of language whereby languages consists of a shared set of rules or conventions agreed upon by a linguistic community. Significantly for Davidson, this sort of linguistic peculiarity is not a rare occurrence, but rather

happens all the time. Furthermore, he claims, "if the conditions are generalised in a natural way, the phenomenon is ubiquitous" (*Derangement*: 98).[1]

Davidson characterises what happens when a hearer interprets a speaker's malapropism as follows.

(1) The Hearer realises that the "standard" interpretation of the words spoken cannot be the intended interpretation.

(2) The absurdity of what the speaker would have meant had the words been taken on their standard interpretation alerts the hearer to trickery or error.

(3) The similarity in sound between the malapropism and the "right" word "tips off" the hearer to the appropriate word. (Davidson denies that there must be a similarity of sound, but presumably *something* must tip off the hearer if this stage is necessary in the interpretive process).

Taking as basic the belief that "nothing should be allowed to obliterate or even blur the distinction between speaker's meaning and literal meaning" (*Derangement*: 91), Davidson suggests the phenomena of malapropisms threaten a deep distinction between what on a given occasion a speaker's words mean, and a what a speaker means, since in such cases the intended meaning of the words takes over from the standard meaning *of the words*. Since Davidson is concerned that we need to account for how it is that the words can take on non-standard meanings in these contexts, (and since he has already stated his belief that the cases in question are generalisable to language as a whole), he proposes that we need to modify our view on what it is to know a language, and even our views on what a natural language is.

[1] Davidson (*SAL*: 115) provides more evidence to the wide range of phenomena that support his case in an attempt to fend off Dummett's (1986: 472) objection that such instances are the exception rather than the rule.

By way of setting the ground, Davidson makes a distinction between conventional meaning and what he calls “first meaning” (more typically called literal meaning). *First meaning* is what the speaker initially intends a hearer to understand by his utterance: it is first in the order of interpretation (regardless of what the speaker also intends the hearer to understand by uttering the expression in question). So for example, “if the occasion, the speaker, and the audience are ‘normal’ or ‘standard’ […] then the first meaning of an utterance will be what should be found by consulting a dictionary based on actual usage” (*Derangement*: 91).

Barber (2004) clarifies this by using Grice’s (1957: 380) example of a reference for an applicant for a philosophy job: “Dear sir, Mr X’s command of English is excellent, and his attendance at tutorials has been regular. Yours, etc.”.

The first meaning of the utterance, “Mr. X’s command of English is excellent” – is that Mr. X’s command of English is excellent, even if the audience is also meant to understand by the sentence that Mr X would be a poor candidate for the job. Furthermore, Davidson claims that a speaker necessarily intends that his audience grasp the first meaning, and the first meaning is grasped if communication succeeds.

Having established this account of first meaning, Davidson then suggests criteria that first meaning must satisfy if it is to count as linguistic meaning on a traditional understanding of language. (1) First meaning must be systematic: there must be systematic relations between the meanings of utterances to enable an interpreter to interpret novel utterances in a language on the basis of its parts and the rules of composition. (2) This method of interpretation of first meaning must also be

shared between speaker and hearer. (3) First meaning is prepared: it must be governed by learned (and so pre-existing) conventions or regularities.

However, Davidson argues that linguistic phenomena such as malapropisms demonstrate that first meaning does not satisfy these criteria: the phenomena 'threaten standard descriptions of linguistic competence', even those descriptions for which, Davidson confesses, he himself is responsible.

In cases where such phenomena arise, a speaker and hearer must each make use of two theories of meaning, a "passing" theory and a "prior" theory.[2] Neither of these satisfies all three criteria that Davidson sets out. To see why Davidson thinks that this is, it is first necessary to clarify what he means by "prior" and "passing" theories.

1.0.1 "Prior" and "Passing" Theories

An interpreter comes to the occasion of an exchange already armed with a theory – a *prior theory*. Importantly this is not merely a theory that covers the finite axioms and rules of composition that we might expect to find in a Davidsonian theory of meaning but also one that has already been adjusted to account for "knowledge of character, dress, role, sex, of the speaker and whatever else has been gained by observing the speaker's behaviour, linguistic or otherwise"(*Derangement*: 100). In other words, it is a theory of meaning already geared towards the speaker's idiolect.

[2] Davidson denies that the interlocutors *have* these theories; however, he often appears to talk as if they do. This has, it seems led to some confusion in interpreting his claims and led to objections that may be off the mark. See, for example, Dummett's (1986: 467) objection that the theories must be of "infinite order", and L&Ls response (p. 291). For simplicity, I may refer to the interpreter and the speaker as "possessing" or "having" the theory, or theories, in question.

The speaker also has a *prior theory* of the hearer. This theory models what the speaker believes about how the hearer is apt to interpret, given the speaker's knowledge of the hearers past use of expressions, dress, age, sex etc. Thus the speaker arrives at the conversation pre-disposed to speak in a certain way.

However, as the speaker speaks, the theory that the hearer has prior to an occasion of utterance by a speaker is yet further altered. In other words, the interpretive theory is altered to match the occasion. Some of the alterations made to a theory occur during the course of an utterance, since it is during the course of this utterance that significant evidence for how the utterance is to be interpreted is provided. The speaker may, during the conversation, also alter her theory of the hearer depending on the hearer's responses to her utterances. These theories, the one that the speaker actually uses in gearing his utterances to the hearer, and the one that the hearer actually uses to interpret the speaker, are the *passing theories.* Thus the theory of interpretation used to interpret (or produce) an utterance on any given occasion is different to the theory of interpretation that can be used to describe the knowledge or skill of the interpreter or speaker prior to the occasion of the speaker's utterance.[3]

Davidson is prepared to go some way to accepting that the knowledge and abilities that *constitute* the *prior* theory may be called conventions – they are systematic, and learned in advance of and given exchange; however, there is no reason to suppose that the prior theory *need* be a theory of the standard or conventional uses of a linguistic community of which the speaker considers herself to

[3] I will henceforth focus on interpreter's theories alone for the sake of simplicity.

be a member – the prior theory could be pre-geared to interpreting an idiolect that is not consonant with standard uses of expressions in a community.

At any rate, it is not the *prior* theory that must be shared by a pair of interlocutors for the exchange to succeed but rather the *passing* theory, since this is what is actually used to interpret an utterance, but nevertheless is not known in advance of the exchange. As such the *passing theory* cannot be said to be constituted by antecedently-known shared conventions.

As such, Davidson concludes, neither theory characterises a speaker's or interpreter's linguistic competence, nor do they satisfactorily characterise a language. He considers whether an alternative theory could do any better, first assessing a proposal that a basic framework, containing vocabulary and grammatical rules for a language could count as an account of language that would be sufficient for explaining communication. This proposal, he suggests, would be even worse off than his *prior* theory, since it is subject to the same pit-falls (it is not actually what is shared when an utterance by a speaker is interpreted), and does not have the advantages of the *prior* theory: being so general, it is not specific to the idiolect of the speaker. Such a basic framework, he acknowledges, may be an ingredient in communication, but it is not sufficient.

So: neither prior nor passing theory meets the three criteria of sharedness, systematicity and preparedness as stipulated in the traditional characterisation of a language. The prior theory is systematic and prepared, but not shared; the passing theory is shared and (presumably) systematic, but not prepared.

1.0.2 Criteria for Successful Communication and Utterance Meaning

An instance of successful communication, on Davidson's account, occurs with the convergence of the passing theories of both speaker and interpreter: this is the content we can give to the idea that two people share the same language. But this characterisation of a language, is, he suggests, very different from the standard understanding of linguists and philosophers – a passing theory is "derived by wit, luck, and wisdom from a private vocabulary and grammar, knowledge of the ways people get their point across, and rules of thumb for figuring out what deviations from the dictionary are most likely" (Derangement: 107).

Passing theories converge when the hearer interprets the speaker's utterance as she intends her utterance to be interpreted. If this happens, then the speaker has "gotten away with it": "[The] speaker expects to be, and is, interpreted as the speaker intended although the interpreter did not have the correct theory in advance" (*Derangement*: 99).

A speaker's "getting away" with having her utterance interpreted as she intended it to may, *prima facie,* appear to be sufficient for her utterance having the meaning that she intended it to. But, this would have an unpalatable consequence: a speaker could be successfully interpreted accidentally: the interpreter manages to guess at the meaning with a pure "stab in the dark", or perhaps by mishearing what was said and, by sheer coincidence, correctly interpreting the "misheard" acoustical production. Moreover it would seem a far too strong a constraint on meaning to suggest that being successfully interpreted was a necessary condition, as there may be a multitude of external factors that may prevent communicative success but that have nothing to do with what was said or how it would have been interpreted in more

congenial circumstances (a plane roars past, drowning out my voice as I start to speak to you, etc.).

Davidson heads off these worries by introducing *legitimate expectation* as a necessary constraint on utterance meaning (*Derangement*: 98). One needs to have a legitimate expectation that one's utterance can be interpreted in the way that one intends in order for that utterance's meaning to be so determined. This requirement Davidson borrows from Donnellan, for when Donnellan is accused by McKay of adopting a "Humpty Dumpty" approach to language, Donnellan (1968: 212) remarks that he cannot simply mean whatever on want to mean as Humpty suggests, because one cannot even form an intention to mean a particular thing by one's words if there is no chance of being successfully understood. One cannot wave one's arms up and down with the intention of flying. Similarly, Humpty Dumpty, according to Davidson, can't "get away with it": he has no reasonable expectation that Alice will understand what he means by the word "glory".

However, what makes an expectation to be interpreted in a particular way reasonable or legitimate? We might say that someone's belief is reasonable merely because it is consistent with other beliefs that they have – even if these beliefs are false. Take Mrs Malaprop, for example, who "gets away with it" without realising it. Mrs Malaprop's belief that her "A nice derangement of epitaphs" will be interpreted as *A nice arrangement of epithets* probably follows from her false belief that she is speaking standard English, that her audience speaks standard English, and that "A nice derangement of epitaphs" will be interpreted as *A nice arrangement of epithets* by standard English speakers. In one sense we could say that her belief is reasonable in the light of her other (false) beliefs, but this seems too weak a

constraint. Otherwise, Mrs Malaprop's belief that she could use "ooog ogoggg" to mean *A nice arrangement of epithets* would be equally reasonable if it followed from the beliefs that she and her audience were speaking standard Enlgish and that "ooog ogoggg" meant what "A nice arrangement of epithets" does in standard English and so on. A belief's being reasonable must in some sense depend on facts of the matter concerning whether Mrs Malaprop can actually be interpreted as she intends.

Davidson (*SAL*) subsequently states constraints on what counts as reasonable that suggests he counts as reasonable *what is reasonable in the light of how things are*, as opposed to *reasonable in the light of what one already believes*:

> If a speaker reasonably believes that he will be interpreted in a certain way, and speaks with the intention of being so understood, we may choose to say he means what (in the primary sense) he would have meant if he had been understood as he expected and intended. Reasonable belief is such a flexible concept that we may want to add that there must be people who would understand the speaker as he intends, and the speaker reasonably believes that he is speaking to such a person. (*SAL*: 121)

It might seem that the constraints on one successfully getting one's utterance to bear an intended meaning are stronger than the constraints forming intentions in general. With general cases of intending, Davidson suggests that, while, of course one cannot intend to do the impossible, one can be unsure as to the success of some action being intended, yet nevertheless "intention is not necessarily dulled" (*Intending*: 131). Yet, in *SAL,* reasonable expectation appears to require more than mere uncertainty about whether one will succeed in being interpreted. We should, therefore, bear in mind that there is a certain amount of latitude in how we understand "reasonableness" or "legitimacy" of intention: Davidson has not clarified this to the extent that we would like. However, what counts as legitimate in these

circumstances surely hangs, however, on the broader question of how a hearer can come to understand an utterance, rather than pre-empt it.

1.0.3 Lewis's Conventions and Davidson's "Anticonventionalism"

Since Lewis's account of convention is commonly considered to be the main target of Davidson's attack when he claims that utterance meaning is not governed by convention, I will give a brief outline of it here. However, I will also suggest that Davidson's claims amount to more than merely a rejection of the claim that linguistic conventions, as characterised by Lewis, are necessary for communication and determine the meanings of expressions in a language.

Lewis (1969) provides a game-theoretic account of convention that shows it as solving a co-ordination problem. Once established, conventions are self-perpetuating: individuals will conform because it is in their interest to conform given that others conform. Such conventions are neither explicit nor implicit agreements. Thus, conventions do not need to be convened *in language*, but occur and persist as a result of the rational activity of humans in order to get the best overall pay-off from coordinating their activities. This solves a concern advanced by the likes of Quine (1969: xi) and Russell (1921: 190) that without being already in possession of a language, a community could not engage in a process of establishing linguistic conventions in the first place (Rescorla, 2007). Like Grice, Lewis holds that the meaning of a sentence derives from regularities of speakers of a language to use that sentence for a certain end; however, Lewis also wanted to demonstrate how the notion of language in formal semantics could be integrated into this understanding. In formal semantics a language can be seen as a function which maps meanings onto strings of sounds. In order to provide a defence of our intuitions that languages are

conventional, Lewis attempts to give a robust account of what such conventions in language must be.

In sharing a language, a given population must have a convention of truthfulness and trust in that language. A given population share a language (*L)* if they observe the particular sort of regularity (*R*) in language use whereby:[4]

> (1) Everyone conforms to *R*.
> (2) Everyone believes that others conform to *R*.
> (3) The belief that others conform to *R* gives everyone a good and decisive reason to conform to *R* himself. […]
> (4) There is a general preference for conformity to *R* rather than slightly-less-than- general conformity – in particular, rather than conformity by all but any one. […]
> (5) *R* is not the only possible regularity meeting the last two conditions. […]
> (6) The various facts listed in conditions (1) to (5) are matters of *common* (or *mutual*) *knowledge*: they are known to everyone, it is known to everyone that they are known to everyone, and so on. (Lewis, 1969: 226-7)

This account provides sufficient conditions for our pre-theoretic or intuitive notion of conventionality in language, presented by Laurence as:

> 1. Linguistic properties are not intrinsic properties of the utterances which have them. The marks and sounds we use could have (or could have had) linguistic properties other than those they in fact have.
> 2. Though we might speak of any of a large number of possible languages, the members of a given linguistic community have a common interest in linguistic coordination (i.e., in speaking the same language), because they have a common interest in communication. (1996: 273)

Many objections can and have been made against Lewis's account of linguistic conventionalism. Laurence (1996: 286ff) himself objects that the complexity of the mental states Lewis requires of those observing a linguistic convention aligns linguistic ability with general reasoning, but that there is good empirical evidence to

[4] See Laurence's (1996: 275) gloss on Lewis for a more detailed account of Lewisian conventions, which integrates the convention of truth and trust in a language that is required for conventionalism in language in particular.

the contrary. Burge (1975) objects to Lewis's requirement for common knowledge, noting that, it is perfectly possible for people to be unaware that the putative conventions that they adhere to are arbitrary: i.e., that there is an alternative, yet incompatible regularity that the belief that others conformed to that regularity would give one a reason for conforming to it, and that there would be a general preference for all to conform. A linguistic community could think that the meanings of the expressions in their language were intrinsic properties of them, and that no substitutes could be used. Others (e.g. Gilbert, 1992; Davis, 2003; Milikan, 2005) have complained that the account that Lewis provides of conventions does not sufficiently capture their normative force. Lewisian conventions are not normative, although he does acknowledge that conventions can become norms if in place for a sufficiently long time.

However, Davidson's objection to Lewis is very different in tone. He claims to be willing to accept Lewis's account of linguistic conventions as a satisfactory account. Davidson, importantly, accepts that linguistic conventions exist. However, Davidson denies that an investigation of linguistic conventions casts any light on the essential nature of language, precisely because – even granting that such conventions exist and are made use of in linguistic communication – they are not an essential feature of language.

One might suppose that Davidson envisages an alternative account of conventions that he would deem essential to language. He does not.[5] He does not, I believe, do this, because it is not just a particular account of conventionalism that he wishes to attack, but rather that which would be a constitutive feature of *any account*

[5] Although Davidson appears to leave the door open to this possibility at the end of *Derangement*, he never pursues it.

of linguistic conventions that could be offered.[6] If he demonstrated that this constitutive feature of conventions is neither necessary nor sufficient for successful linguistic communication, then *a fortiori,* any conventionalist account of language will fail.

Davidson, of course, does not deny that expressions of a language are *arbitrary*, but, as he rightly points out, the arbitrary nature of words does not entail that they are conventional (*C&C*: 265). Thus he acknowledges Laurence's first condition of our pre-theoretic notion of linguistic conventionalism. It seems to me however that he would take issue with the second condition, namely that members of a linguistic community have a common interest in linguistic coordination (i.e. speaking the same language) *because* they have a common interest in communication. The problem here is that the 'because' in this condition, so stated, seems to be susceptible of alternative formulations. Among other things, one might have an interest in x because it conduces to y, or one might have an interest in x because it entails y (or is necessary and sufficient for y). Davidson would and does perfectly well accept that 'speaking the same language' (appropriately understood) *facilitates* successful language, but he vehemently denies that speaking the same language is *necessary* (*SP*: 114).

Thus, we might say on Davidson's behalf that he would accept condition 1 and a modified version of 2; however the conditions so stated do not amount to the claim that linguistic conventions on any formulation are necessary and sufficient for linguistic convention.

---

[6] As we see in chapter 1 of this thesis, I suggest that L&L's "*ahistorical* convention" fails to shed any light on the issue precisely because such an *ahistorical* convention nevertheless requires participants to antecedently know the content of the convention.

Davidson would deny that linguistic conventions even understood pre-theoretically are necessary and sufficient for linguistic communication. This is because Davidson attacks the necessity of conventions for linguistic communication via mounting an attack on the necessity of a hearer's *prior knowledge* of a speaker's dispositions to use words. Put simply, a hearer, according to Davidson, does not need to know what a speaker means by a sentence prior to an utterance in order to interpret that sentence.

While acknowledging that these conditions do normally hold in practice in linguistic interaction, Davidson (*C&C*) claims they are not necessary for linguistic communication and thus are not a constitutive feature of language. His argument can be formulated as follows.

1) Convention depends on a mutually understood practice.
2) There must be 2 or more people involved in such a practice.
3) Successful communication requires that the hearer (H) understands the speaker (S).
4) H & S need not mean the same thing by the utterances they make, since H could understand what S means by an utterance even though H uses the utterance in a different way.
5) Therefore there is no requirement that there be a convention that two people conform to the same use when they speak.
6) All that is required is that H & S assign the same meanings to S's words.
7) This requires some form of agreement.
8) However there is no reason to suppose such agreement must be reached *prior* to an exchange rather than after it.

9) Lewis's conditions for a convention require regularity over time and that speaker's and hearer's theories coincide prior to interpretation.

10) Thus Lewisian conventions are not required, since no regularity over time and no prior agreement is required for H to understand S.

We can agree with Davidson that in order for communication to succeed, speaker and hearer must be agreed on (or converge on) a meaning for the expression uttered. In principle, it is also possible to agree with Davidson up to (7) here. From the perspective of the radical interpreter, and assuming that idiolects are stable, it follows that people don't need to share practices, in terms of "speak the same language". If two radical interpreters, A and B, had satisfactory theories of each other, they could interpret each other even if language each spoke was different. (*SP*: 114).[7] The two would be following different practices but communication would still be possible. Thus Davidson does not need the additional arguments of *Derangement* to present a case against the necessity and sufficiency of conventions for communication.

But Davidson is arguing for a more radical point than this. He has already built in the possibility that this prior theory does not map conventions in operation in a particular linguistic community: "[T]he prior theory has in it all the features special to the idiolect of the speaker that the interpreter is in a position to take into account

[7] Davidson *ad loc.* also denies that the capacity to understand a language is the same as the capacity to produce utterances in that language: "it might even be that because of differences in our vocal chords we couldn't make the same sounds, and therefore couldn't speak the same language. I know of no argument that shows that under such circumstances communication could not take place."

before the utterance begins" (*Derangement*: 104). [8] Already, then, one form of conventionalism (which states that members of a linguistic community must conform to that community's conventions) is out of the running.

However, the controversial claim in this argument is (8): there is no reason to suppose that the agreement that is reached on the meanings of words must be reached *prior* to an exchange rather than after it. If the speaker and hearer have not agreed in advance on the meaning of a particular utterance, but can do this afterwards, then it follows that the hearer need not have knowledge of the meaning of the expression in the mouth of the speaker (at that time, in that context etc.) prior to its utterance.

Davidson, then, denies that the interpreter's having *prior knowledge* of the meaning of a speaker's token utterance is necessary for interpreting it (*Derangement*: 102,107); *a fortiori*, no prior knowledge of any *conventionally* determined meaning is necessary. But it seems that the denial of the necessity of prior knowledge rules out far more than just conventionalism, Lewisian or otherwise; it also renders prior knowledge of regularities in a speaker's past usage unnecessary for interpretation. Since adhering to conventions or past regularities is not necessary to have one's utterance interpreted as one intends, then neither convention, nor the precedent of past regularities can determine what one's utterance means.[9]

---

[8] Dummett (1986: 461) notes that Davidson already has, in the initial stages of the argument of *Derangement*, deployed a notion of language prioritising idiolect: a notion of language in which "no two speakers need speak the same language".

[9] We grasp that 'epitaph' in the mouth of Mrs Malaprop means *epithet despite having never encountered the word being so used before*. Moreover, Davidson's claims about how utterances are interpreted does not depend on supposing Mrs Malaprop will always make her 'mistake'; once is enough to summon up a passing theory assigning a new role to 'epitaph'" (*Derangement*: 103). Thus neither knowledge of conventions *nor* knowledge of past regularities can be said to be necessary for interpreting an utterance.

1.2 "There is No Such Thing as a Language" and the Necessity and Sufficiency of Knowledge of Conventional Meanings

Davidson concludes *Derangement* by stating that "there is no such thing as a language", and moreover that we "should give up the attempt to illuminate how we communicate by appeal to conventions" (107). Davidson's denial that languages (as traditionally conceived) exist follows from his denial that the conventions of a particular speech community determine the meaning of an utterance produced by its members. The remaining sections of this chapter will analyse Davidson's claim that convention is neither necessary nor sufficient for language (whatever that means) or linguistic communication, with particular attention being paid to and Lepore and Ludwig's analysis of Davidson's claims.

Before examining the claim that convention is neither necessary nor sufficient for linguistic communication, I will consider if it follows from the observations about malaprops and associated phenomena.

1.2.1 Malaprops, generalisability and First Meaning

On reading *Derangement* it might seem, as Dummett (1986) supposes, that Davidson is picking on a feature of communication (successful interpretation of malapropisms) that is the exception rather than the rule, and then generalising this phenomenon without justification. Dummett claims that the cases that aroused Davidson's interest are undeniable, but "are, however, in the nature of things, atypical cases: if taken as prototypes for linguistic communication, they prompt the formulation of an incoherent linguistic theory" (1986: 472). Barber (2004), however, defends Davidson by saying that those who present the situation in this way may be

“misinterpreting the strategy behind Davidson’s introduction of these “rogue” phenomena. He seems to offer them, not as counterexamples to the standard picture (in which case the above response would be legitimate), but as vehicles for the illustration of an alternative vision of communication that applies even in regular cases”. If so, we should consider why Davidson believes that a general claim about linguistic utterances and their interpretations can be inferred from his observations about the specific case of malapropisms.

Essential to the generalisation from “rogue” cases is the distinction that is made in *Derangement* between (1) speaker meaning (SM), (2) conventional meaning (CM) and (3) first meaning (1M).

- (SM) Speaker meaning should be understood as what the speaker intends his audience to have understood him to have conveyed by his utterance. Barber clarifies this by using Grice’s example of a reference for a philosophy job applicant: “Dear sir, Mr X’s command of English is excellent, and his attendance at tutorials has been regular. Yours, etc.”. The speaker meaning in this case is traditionally understood to be that the applicant in question was a poor candidate for the job, since that is what the speaker meant to communicate on this occasion by using this sentence in her reference.
- (CM) The conventional meaning of an utterance is what the utterance would standardly, or conventionally, be taken to mean when used in a particular linguistic community. So in the above case, the conventional meaning of the utterance would have been that *Mr X’s command of English was excellent and his attendance at tutorials had been regular*.

- (1M) First meaning, as Davidson calls it, is distinct from both these “types” of meaning. First meaning is “first in the order of interpretation”. First meaning is what the speaker can legitimately expect the interpreter to take his *utterance* to mean in the context in which it is so uttered (*Derangement*: 97).

Although first meaning can coincide with either conventional or speaker meaning, it need not necessarily coincide with either. *Prima facie,* the non-coextensivity of 1M and either CM or SM can be seen when considering the Gricean example mentioned above: here 1M would coincide with conventional meaning, but not the speaker meaning. The intended audience would first interpret the words literally (or according to the conventions of his linguistic community) and then make inferences from the conventional meaning of the words plus certain facts about the context and purpose of the utterance to what the speaker/author really intended to convey (the speaker-meaning), namely, that Mr. X was not a suitable candidate for the job.

The case of malapropisms demonstrates that 1M need not coincide with CM either. The conventional meaning of Mrs. Malaprop’s “ A nice derangement of epitaphs” is *A nice derangement of epitaphs*, yet, according to Davidson, its first meaning is *A nice arrangement of epithets*, since that is what an interpreter is expected, by the speaker, to initially interpret her words to mean.

1M need not coincide with either CM or SM in any given utterance. To return to the Gricean example again: suppose the referee, intending to write “Mr. X’s attendance at tutorials has been regular etc.”, had instead written “Mr. X’s attendance at tutorials has been regulated etc.”, and had written this with the intention that his audience interpret his expression as meaning *Mr. X’s attendance at tutorials has been regular etc.*, and him as meaning that Mr. X was a poor candidate for the job.

Then we would have a clear separation of SM, as described by Barber in the Gricean example, (that Mr X is a poor candidate), CM (that Mr. X‘s attendance at tutorials was regulated) and 1M (that Mr. X’s attendance at tutorials was regular).

So far Davidson’s observations seem to point to a category of meaning which is distinct from both CM and SM, although on many occasions it may coincide with one or the other, or all three types of meaning may converge.

Let us grant for the moment that the phenomenon of the malaprop successfully demonstrates these distinctions, and that, in the case of the malaprop, the first meaning of the utterance that is interpreted by a hearer does not coincide with the conventional meaning. Let us assume that in other, “standard”, cases, the first meaning of an utterance does coincide with the conventional meaning that a particular linguistic community ascribes to it. Davidson’s point seems to be that appeal to a hearer’s knowledge of conventional meanings can at most only explain the communicative success of *some* linguistic utterances. However, the hearer’s grasp of the first meaning of an utterance can explain the communicative success of *all* linguistic utterances. Thus, if we want to explain what is essential to the meaning of a linguistic utterance, we must look to the notion of first meaning, not conventional meaning: first meaning is essential to linguistic communication in the case of a particular utterance.

On this account, it seems that the focus has shifted from talk of what words and utterances in a particular language mean, to talk of the meanings of utterances in a particular linguistic exchange: “Meaning in the special sense that we are interested in when we talk of what an utterance literally means gets its life from those situations in which someone intends…that his words will be interpreted in a certain

way, and they are" (*SAL*: 120). It is the meanings of utterances understood in this way that seem to be of primary interest to Davidson. As such, conventional meaning appears to be a deflated kind of meaning: to say that "A nice arrangement of epithets" has the conventional meaning of *A nice arrangement of epithets* (in the English language community) is to say no more than that the speakers of that community typically use that expression to communicate that content. However, expressions don't have genuine meaning aside from token utterances of them, since this abstract meaning has no direct tie to how the token instances are correctly interpreted (*SP*: 108-10, 116).

Therefore, the distinction between first meaning and conventional meaning is not aimed at showing that there are occasions where first meaning, rather than conventional meaning, is grasped, but rather that first meaning is what is always grasped in a successful exchange even if it is often informed by knowledge of the conventions in operation in a particular speech community.

Armed with this picture of linguistic communication, we can now examine the claims of the necessity and sufficiency of conventions for linguistic communication in general.

1.3 *Lepore and Ludwig's Formulations of Davidson's Anticonventionalist Claims*

Lepore and Ludwig (*L&L*) formulate the main claims of Davidson's thesis as follows (I have put these claims in my own words for the sake of clarifying what is at stake in the distinctions that *L&L* make):

(1) **Knowledge of the conventional meanings (KCM) of the words that a speaker uses is insufficient for interpreting him**

(2) **Knowledge of the conventional meanings of the words that a speaker uses is not necessary for interpreting him.**

(1) and **(2)** can in turn each be interpreted in stronger and weaker senses.

(1S) **There is *always* additional knowledge that must be brought to bear in addition to knowledge of conventional meanings to correctly interpret a speaker.**

= Knowledge of conventional meanings is *always* insufficient to interpret a speaker.

(1W) **There is *on occasion* additional knowledge that must be brought to bear in addition to knowledge of conventional meanings in order to correctly interpret a speaker.**

= Knowledge of conventional meanings is *on occasion* insufficient to interpret a speaker.

(2S) ***As a matter of fact, speakers being as they are*, knowledge of conventional meanings is not necessary to interpret a speaker.**

(2W) ***In principle*, if not as a matter of fact, speakers being the way that they are, knowledge of conventional meanings is not necessary to interpret another speaker.**

L&L proceed to state which combination of the claims (1W) or (1S), and (2W) or (2S) they believe Davidson did, in fact, mean to suggest.

In short, they conclude that Davidson suggests (1S), from which (1W) follows *a fortiori,* and they believe that (2W) follows from what Davidson says; however, they are somewhat more circumspect about (2S): "knowledge of conventions for word use seems, for us at least, to be essential for communicative success, even if it is not a

matter of necessity that we always interpret words in accordance with public conventions" (*L&L*: 280). However, as far as (2S) is concerned, *L&L* suggest that Davidson should be attacking a particular historical notion of convention, rather than convention in general; indeed, they claim, this "historical" notion of convention is not necessary for communication. However, some notion of convention is required for communication and they suggest that an *ahistorical* construal of convention will fit the bill, in so far as linguistic communication is concerned. According to *L&L*, we should still view public, shared language as having a role in our philosophical investigation of human linguistic understanding. Nonetheless, its role "will not be one it has in virtue of our being linguistic beings, but in view of being the cognitively and epistemically limited linguistic beings we are" (*L&L*: 296).

1.3.1 Weak and Strong Sufficiency Claims

Before discussing whether Davidson means to suggest 1W or 1S, it is important to note that 1W follows *a fortiori* from 1S, but not vice versa. 1W and 1S can be rephrased simply:

(1W): KCM is, *on occasion*, not sufficient for communication.

(1S): KCM is, *always* (necessarily), not sufficient for communication.

(1W) Can be relatively straightforwardly established from the conclusions that Davidson draws from his observations concerning the phenomenon of malapropisms and there ilk. (1S) Can be established from the conjunction of these same conclusions with an additional sceptical argument concerning how one is to know

that the default (conventional) interpretation of an utterance should be chosen on any given occasion (*L&L*: 273).

As we have seen, 1M is distinct from conventional meaning. Davidson claims that successful interpretation of non-standard first meanings demonstrate that the words of one speaker can on occasion be interpreted without appeal to a prior convention shared by speaker and hearer. (1W) Follows from this: a hearer can interpret a speaker's words, at least on occasion, when an interpretation of those words in line with the interpreter's knowledge of CM would otherwise fail.

The occasional phenomenon of malapropisms could support (1W), but it is not immediately clear how they could support (1S). If the speaker utters a sentence with the intention that it have the meaning that it standardly or conventionally does in the community, and the interpreter knows those standard meanings, it would seem that in these (default) cases, no additional knowledge must be brought to bear in the course of the interpretation. At this point the defender of (1S) can retort with a "sceptical argument" against default interpretation: since an interpreter is *never* in a position to know whether the case in point is going to be a default case (where knowledge of CM would suffice for successful interpretation), then additional knowledge must *always* be brought to bear in an interpretation, at the very least to establish whether the case will be a default interpretation or not (*L&L*: 274).

One may object that, as a rule, when the default meanings are intended, no interpretation takes place: the speaker does not consciously or unconsciously use additional information to establish that the default (conventional) meaning is in play, the hearer simply and automatically understands the utterance (Dummett, 1986:

471). It is only if a speaker's utterance appears false or incoherent on the default interpretation that the hearer makes use of additional evidence in interpreting the utterance in question. Thus (1S), the claim that knowledge of CM is always insufficient for understanding the utterance of a speaker, may be rejected.[10] *L&L* conclude that whether we accept (1S) or not depends on what we mean by saying knowledge of CM is insufficient for the task of interpreting:

> ..[T]he question of whether [(1S)] is justified boils down to what we intended by saying that knowledge of conventional meanings is sufficient for interpretive success. If we meant that sometimes, even often, we are not called upon to revise our view that the speaker speaks with the majority, then [(1S)] should be rejected. If we meant that knowledge of convention meanings, all by itself sometimes suffices for interpreting another as speaking in accord with conventional norms, then since this is not so, we should accept not just [(1W)], but [(1S)] as well. Indeed, in this case, [(1S)] follows simply from the observation that human beings are fallible. (*L&L*: 271)

L&L certainly seem to be tending towards (1S), and I agree that whether one adopts (1S) and (1W) may just depend on whether one accepts that in addition to the knowledge of the conventions, one also needs accompanying knowledge concerning whether the speaker is adhering to those conventions or not. However, the question of whether one adopts (1W) or (1S) seems to be of little relevance to Davidson's overall claim. On the account of the distinction between first meaning and conventional meaning that I set out above, (1W) is all that is required to show that first meaning and conventional meaning are not identical; if knowledge of conventional meanings is not sufficient to interpret a linguistic utterance even on one occasion, but the meaning of the utterance is nevertheless interpretable, then the meaning of the utterance is not identical with the conventional meaning ascribed to it by a particular linguistic community.

[10] Cf. Dummett (1986: 467) on the distinction Wittgenstein makes between interpreting (*Deutung*) and understanding in general (*Auffassung*)

To put it another way, if it could be shown that conventions were sufficient for communication in all instances of communication, then it would be difficult to show that conventions were *not necessary* for linguistic communication. If it were proven that prior knowledge of convention was sufficient for *all* cases of linguistic communication, then it would not follow that it is a *fundamental feature* of language or linguistic communication, since there may conceivably be a different linguistic feature that could do the job equally well. Nevertheless, one could not then find empirical evidence for opening up the divide between conventional meaning and first meaning in the first place.

Yet in order to state the problem, it seems, all that Davidson needs to show is the non-identity between first and conventional meaning. The claim that conventions are, on occasion, insufficient for linguistic understanding achieves this end.

In order to show that conventions are not a fundamental feature of language or linguistic exchange, the important claim to focus on is that conventions are not *necessary.* If communication can proceed without conventions, then it is likely that there is very little we can say about the fundamental features of language (or "language") by appealing to them.

1.3.2 Weak and Strong Necessity Claims

As I have mentioned above, *L&L* give the claim that conventions are not necessary for communication, a weak and a strong formulation: either (2S) as a matter of fact, speakers being as they are, knowledge of CM is not necessary for communication; or (2W) *in principle*, if not as a matter of fact, speakers being the way that they are, knowledge of CM is not necessary for successful interpretation. *L&L*,

as we shall see, consider that (2S) cannot be justified from considerations derived from the phenomena of malaprops and their ilk, but they do believe that (2W) holds water. I detail their argument below; however it may be worth noting that Davidson's own claims imply the weaker version of the theory:

> Knowledge of the conventions of language is [...] a practical crutch to interpretation, a crutch we cannot in practice afford to do without – but a crutch which, under optimum conditions for communication, we can in the end throw away, and could in theory have done without from the start. (*C&C*: 279)

1.3.3 Argument for 2W

L&L (278) attempt to clarify (2W) with the following question: "is there knowledge an interpreter could in principle have, leaving aside natural limitations of knowledge and perspicacity, which would enable him to interpret a speaker of whom he had no prior knowledge at some given point?", and say that this is equivalent to "asking whether there are facts independent of linguistic conventions that determine (or could determine) what a speaker means by his words".

They suggest that if an interpreter knows a speaker's dispositions to use words in his environment, then that is all that is required for the interpretation. Thus they endorse (2W): *in principle* an interpreter who knows the speaker's dispositions to use words etc. need not appeal to those words' conventional meanings in order to interpret successfully. They seem happy with the idea that even though such knowledge would in practice be impossible, the theoretical possibility of such knowledge existing is enough to show that conventions are not necessary for communication.

The claim that it is possible for a condition to be unnecessary *in principle,* but necessary in fact, or in practice, requires some unpacking, as does L&L's explication.

How much sense can be made of a claim that some condition x is *in fact* necessary but *in principle* not necessary for some result y? Typically, to say that x is *in principle* necessary for y would amount to saying that, in general, x is necessary (in a loose sense) for y (although there are cases where y occurs without x). However, they do not mean by "in principle" *what is generally the case*; rather, they suggest that knowledge of CM is only contingently necessary for successful interpretation, since we can conceive of an interpreter equipped to interpret a speaker who does not adhere to conventions.[11] We could, one supposes, envisage a world where users of language were sufficiently different in psychological and cognitive make-up from users in our world so as to successfully interpret without reliance on their knowledge of CM (*L&L*: 279).

However, if this is what is meant in this context by "in principle" then *L&L's* (and Davidson's) claim that knowledge of CM is not *in principle* necessary for communication seems to rest on an assumption that we can characterise language, or define what language is, independently of real language users. This point becomes more apparent when we consider *L&L's* illustration of how prior knowledge CM of may *in principle* be unnecessary for successful interpretation:

> Indeed, if knowledge of a speaker's dispositions plus knowledge that he is of the same psychological type as oneself is sufficient for correctly determining what his words mean, it will be in principle possible to interpret another speaker without relying on any prior knowledge of any conventions or regularities for the use of his words. Two gods could speak to each other, each relying on knowledge that the other knew all of his dispositions, without any need to appeal to knowledge of how either had used or understood words in the past. (*L&L*: 279)

[11] If we compare the alternative formulations of the necessity conditions to the alternative formulations of the sufficiency conditions, we can see that the distinction between the weak and strong formulations are not parallel. In the sufficiency formulations the distinction seems to be between existential quantification in the weak case and universal quantification in the strong case; in the formulation of the necessity conditions the distinction appears to be modal.

The appeal to "gods" here is not obviously useful in making the case. First, which dispositions do *L&L* have in mind here? If *L&L* mean to suggest that a god has knowledge of all of another's dispositions, including their dispositions to behave in any particular situation, a person's speech behaviour would be a subset of this. The god would then know what someone will say in a particular environment at a particular time given a particular stimulus: in other words, the god would be able to predict the content of the speaker's acoustical product regardless of what sounds emanate from the speakers mouth.[12] But then why must the speaker communicate by something that is recognisable as a linguistic expression rather than a grunt, a squeak, or a whistle? If a god had enough information about a speaker's dispositions and the environment in which he has them, the utterance of the speaker would merely flag up a desire on his part to communicate: something which could, presumably, be done non-linguistically, by a stamping of a foot, say. An omniscient being would presumably know enough about the speaker to know *that* he intended to communicate and what he intended to communicate without even this signal. But then why would gods bother talking at all? After all, it is because of the epistemically and cognitively limited conditions that we mortals find ourselves in that we have language at all: we do not have access to the complexities of the dispositions or intentions of another except though language. To speak of omniscient beings as communicating linguistically, then, seems unhelpful.

Perhaps *L&L* intend a weaker claim: the significant knowledge that the god has is a particular subset of dispositions. These are a speaker's dispositions to use

[12] Davidson (*M&M*: 44) denies a god's knowledge could extend this far: the "omniscient interpreter is not supposed to know everything" for then he would "know what the speaker believed without having to go through the process of interpretation, the nature of which is supposed to guarantee substantial agreement between interpreter and speaker."

words in particular ways in particular conditions. Thus the god would not be predicting the content of an utterance at a given time, but rather, since the god knew how the speaker was disposed to use words in any environment, he would be correctly prepared to interpret any utterance, but the interpretation would nevertheless be necessary in order to assign content to that utterance and thoughts to the individual who so utters.

We might suppose, then, that, some of the things an omniscient interpreter might know about a speaker would include information such as the following.

1) S is disposed to use the expression “dumde dum dum” on a Tuesday, when it is cloudy, to mean that it is cloudy.
2) S is disposed to use the expression “dumdidi dum dum” on a Thursday afternoon, when it is cloudy, to mean that it is cloudy.

And thus when S does utter “dumde dum dum” on a Tuesday, when it is cloudy, the omniscient hearer correctly interprets the utterance as meaning that it is cloudy. Importantly he can do this if he has had no prior experience of a regularity in S’s speech behaviour to this effect, and even if S has never uttered “dumde dum dum” on a Tuesday before.

If this is conceivable, then, the possibility of linguistic communication without convention is conceivable; indeed, the possibility of a hearer correctly interpreting a speaker without having documented relevant past regularities in her speech behaviour, would be conceivable. However, if the hearer knew the speaker’s dispositions to use these words, the hearer would still have *prior knowledge* of what the words “dumde dum dum” meant in the context it was uttered, and be thus prepared to interpret them. But it is Davidson’s denial that such *prior knowledge* is

necessary for communication (whether it be of conventions, regularities or even dispositions to use words) that is at the heart of his anti-conventionalism. It would seem that the whole point of the Mrs Malaprop example was to demonstrate that we can interpret her *without* knowing what she would mean by "a nice derangement of epitaphs" in advance of her saying it – it is this feature of our abilities in the linguistic arena that has piqued his interest.

In summary, the idea that prior knowledge of conventional meanings is in principle unnecessary but in practice necessary for communication is open to objection. In the first place, where the omniscient interpreter has access to the entirety of a speaker's dispositions, language and linguistic communication would be redundant. If we are merely speaking of the omniscient interpreter's knowledge of all of a speaker's dispositions to use words in any context, then *L&L* appear to be misrepresenting a fundamental part of Davidson's discussion: he is interested in our capacity to interpret someone's utterance without prior knowledge their relevant prior dispositions – the epistemic vantage point of a god can throw no light on this feature of linguistic communication.

There is clear evidence that (2W as presented seems to be out of line with Davidson's (*SP*: 109) interests in considering the fundamental features of actual, rather than theoretical linguistic communication: "[r]esearch [in to an abstract language] pays dividends of various sorts, but it is unrelated out our normal concern with understanding the speech of others, or learning to make ourselves understood by them. Our practical, as opposed to purely theoretical interest in linguistic phenomena is this: we want to understand the actual utterances of others, and we want our utterances to be understood".

1.3.4 Argument for *(2S)*

L&L accept (2W), which I have suggested offers little insight into Davidson's concerns. They do not however, endorse the strong formulation of the necessity claim, (2S).

(2S), we recall, is the claim that, *as a matter of fact*, knowledge of CM is not necessary for interpreting the utterances of another. Against this, L&L remark that:

> We clearly cannot know what someone's dispositions to use words are without either having observed him over a period of time, or located him within a linguistic community whose regularities in word use we have antecedently learned. Even with his complete physical description and a correct theory of physics the computational problem would be intractable. We simply are not gods. There is no prospect for knowing what someone means by his words at a time with no grounds whatsoever to think that, prior to the communicative interchange, he uses them one way rather than another. (*L&L*: 279)

L&L's argument that: (1) knowledge of a speaker's dispositions to use words (relativised to context) is necessary for interpreting her utterances. (2) We could acquire such knowledge of his dispositions either from (a) knowledge of past regularities in his speech behaviour, or (b) knowledge of the regularities (or conventions) of the speech community in which she is located. Thus if we have not acquired prior knowledge of a speaker's dispositions to use words from observing past behaviour, we will not be able to interpret her on a particular occasion.

Davidson, I suggest, differs from *L&L* here in that he would not accept step (1), at least as it stands. As I have said, Davidson's own words suggest that prior knowledge of a speaker's dispositions to use words is unnecessary for successful interpretation: the theories coincide *after* the utterance is made – they must do, Davidson says, or communication is impaired. However, the theories do not need to coincide *prior* to the exchange. Thus, just as prior knowledge of convention is

required, prior knowledge of a speaker's dispositions to use words is not required either.

These concerns temporarily aside, *L&L*, however, attempt to modify (2S) to render it more acceptable. For *L&L* the plausibility of (2S) depends on what we take convention to be. The notion of "convention" that Davidson has set his sights on has been taken to be that of Lewis (*L&L*: 280; Barber, 2004). For Lewis, a convention is in part constituted by an actual regularity existing in a community. L&L, however, express doubt over whether an *actual* regularity is essential to convention. Can a convention not be established, they ask, that is yet to be followed? There are, after all, cases where conventions are established before they become regularities, as, one might think are the principles stated in the Geneva convention, among others: "If we agree to the conventions, and dispose ourselves to follow them, then they in effect exist, even before anyone's behaviour is governed by them" (*L&L*: p.281). Thus they modify Lewis's definition of convention, which, they claim, focuses on showing how actual regularities in behaviour constitute conventions, to include the dispositions of a given member of a community to behave in accordance with a convention in a given situation, even if that situation has never in fact arisen, and will never arise in future. This notion of convention can be construed as an *ahistorical* convention.[13]

Conventions at least so understood, are required, *L&L* suggest, even if they are unstable, since, for speaker and hearer to communicate successfully their passing theories must converge, and for their passing theories to converge they must

[13] An additional problem with these conventions would seem to be that they would have a strong normative force; after all, without the game-theoretic explanation for how they arise and persist, what other reason would there be for following them, other than there being some sanction in place to enforce their use? This seems contrary to the spirit of Davidson.

converge on common conventions governing the speaker's use of her words. In so far as conventions of this sort may be required in the interpretive process, Davidson's claim that prior knowledge of the conventions is not required for interpretation, understood as (2S), is incorrect. However, this, we are told does not constitute an objection to Davidson's claim interpreted as (2S) since it is the Lewisian notion of convention that Davidson had as his target when he claimed (and if he claimed) that prior knowledge of shared conventions is not necessary for interpretation and communication: "convention is essential to language. But prior, shared conventions are not" (*L&L*: 282).

In summary, *L&L's* defence of (2S) construes the target conception of convention for Davidson as "historical", whereby conventions are understood to be regularities actually occurring in a given community. However, conventions need not be understood as regularities in this way: our notion of convention can be modified to an *ahistorical* one, covering expected norms of behaviour that have never actually, or might never actually occur. (2S) Is incorrect if *ahistorical* convention is in play; however, convention, traditionally conceived, is not necessary in practice for communication and/or interpretation.

While L&L seem to have grasped something about our intuitive understanding of convention that Lewis's account does not cover, it is difficult to see how they can relate it to the case in hand, namely interpretation in a communicative exchange. As such, it is not clear how the necessity of at least *ahistorical* convention so characterised rescues Davidson's claim that *prior knowledge* of shared conventions are unnecessary for successful interpretation. The interpreter is still expected by *L&L* to have *prior* knowledge of such *ahistorical* conventions, in order to interpret, even if

he had not acquired that knowledge through experience of regularities in the speaker's past behaviour. Their convention may be *ahistorical* in the sense of never previously actualised or manifested, but it does not follow from this that *prior knowledge* of it would not be required.

So it is not clear what the *ahistorical* conception of convention actually adds to the discussion. It seems that convention so modified would still require the prior, shared knowledge between speaker and hearer that Davidson denies is necessary.

Thus I take L&L to be mistaken in their conclusion that "the observation that language is not necessarily conventional, because the historical requirement may not be met in possible communication, appears, then, to be less damaging to traditional views about the relation of convention to linguistic communication" (*L&L*: 284) because they miss the central point. The nature of the convention is not at issue: whether successful communication requires *prior knowledge* of it *is*.

1.4 Summary of Discussion of Lepore and Ludwig's Defence of Davidson

L&L agree with Davidson that knowledge of prior shared conventions is not *sufficient* for linguistic communication, and their defence of Davidson in this regard is plausible enough, although I have suggested that Davidson needs only the weak version of the sufficiency claim in order to show that first meaning and conventional meaning are not identical. However, their characterisation and analysis of the Davidsonian argument concerning the necessity of convention for linguistic communication is weak. In the case (2W), the claim is made that omniscient beings who knew all of the dispositions of a speaker could communicate without prior knowledge of shared conventions. However it seemed to follow that not only was

knowledge of CM unnecessary for communication, but recognisably linguistic utterances did not seem necessary for communication either (one could communicate to a god with a grunt): hence one must wonder what light this throws on our understanding of *language and linguistic communication.*

In the case of (2S), again little light is shed on the issue since *L&L* and Davidson seem to be at cross purposes: regardless of the nature of the conventions, it is the non-necessity of the *prior knowledge* of them that the malaprop example is meant to demonstrate.

More thought is required, then, on what Davidson means by his claim that knowledge of conventional meanings is not necessary for linguistic communication, and whether his position, correctly interpreted, is defensible.

L&L (285) see Davidson's anti-conventionalist claims as naturally evolving from his taking the stance of the radical interpreter as fundamental in understanding linguistic behaviour. If speakers' dispositions to use words fix the meanings of those words, then it is not the conventions of a linguistic community that determine the meanings of utterances but facts about speakers themselves (*L&L*: 295). However, they hold that in practice, prior knowledge of a speaker's dispositions to use words is necessary for correct interpretation of her utterances. This would also seem to follow from their view that Davidson's views on radical interpretation lie at the heart of his anticonventionalism: given the epistemic constraints on the radical interpreter, and his procedure in constructing a compositional theory of meaning for his subject(s) on the evidence of their speech behaviour and the environment, it would seem that no theory could get off the ground without a speaker's behaviour exhibiting regularities.[14]

[14] Radical interpretation is discussed in chapter 2.

However, Davidson's example of Mrs Malaprop shows that in practice we need not have prior knowledge of someone's dispositions to use a word on a particular occasion in order to interpret that utterance. The prior theory models a hearer's dispositions to interpret a speaker based on his beliefs about how she is apt to use words. In the passing theory we see the hearer interpreting the speaker even though this could not be done on the basis of his prior theory – and, in the limit, the passing and prior theories could have very little or no overlap at all. Thus prior knowledge of a speaker's dispositions to use words, *pace L&L*, would not seem, for Davidson, to be required.

1.5 Conclusions: Sharpening the Terms of the Debate

My discussion of *L&L*'s analysis of the necessity and sufficiency of convention was primarily to highlight the important issues that must be considered when evaluating Davidson's claims that conventions are not essential to language.

I have suggested that the interesting claim that Davidson makes is that convention, and regularity, is not *necessary* for linguistic communication. The sufficiency claim in itself is less interesting, because it sheds little light on the fundamentals of language or linguistic communication. However, if conventions (and regularities) are shown to be unnecessary for linguistic communication, then something fundamental will have been said about the nature of language. Consequently, I will primarily be considering the implications of the necessity claim in the remainder of this thesis.

This chapter focussed on the interpretation of a single utterance in a particular context and outlined what Davidson concludes from his observations that a successful interpretation of that utterance can take place without the hearer's appeal

to antecedently known conventions of his linguistic community, or knowledge of prior regularities of an individual speaker's speech behaviour.

L&L agree with this conclusion to the extent that they claim that it is a speaker's dispositions to use words that fix their meanings, rather than conventions or past regularities in use. They also suggest that Davidson's anticonventionalism is a natural offshoot of his views on radical interpretation. Davidson however, claims that the expectations and intentions of hearer and speaker, and the fact of a successful interpretation of an utterance, are essential in fixing the meaning of an utterance; these claims cannot be straightforwardly reduced to talk of dispositions, nor to what can be discovered by the radical interpreter. This will be discussed in more detail in chapters 2 and 3.

Chapter 2: Radical Interpretation and Innovative Utterances

2.0. Introduction:

It has been suggested that Davidson's anticonventionalist claims result from his taking the stance of the radical interpreter as fundamental in understanding linguistic behaviour (e.g. *L&L*: 295; Bar-On and Risjord, 1992). In this chapter I will investigate whether the resources and procedure of radical interpretation enables an adequate account of the interpretation of linguistic innovations such as the malaprop[1]. After giving a brief overview of radical interpretation I will suggest that interpretation of a malaprop in practice is not radical and, moreover the resources and procedure of the radical interpreter do not sufficiently constrain interpretation. Thus while radical interpretation may cast light on the possibility of successful communication in general, it does not explain why conventions do not determine the meanings of utterances.

2.1. Radical Interpretation

To radically interpret is to interpret the language of another speaker from scratch, that is, using evidence that 'must be of a sort that would be available to someone who does not already know how to interpret utterances the theory is designed to cover' and without making any essential use of "unexplained linguistic concepts" (*RI*: 128; *RTF*: 176). Radical interpretation was advanced by Davidson, in

[1] I reserve the term 'innovation/innovative utterance' for utterances that cannot be accommodated or generated by an interpreter's prior theory. This is in contrast to 'novel' expressions, which can be generated by an interpreter's theory but which have not previously been encountered.

part, to counteract various legitimate criticisms of his claim that a Tarski-style truth theory could model linguistic competence for a natural language. The principle aim of radical interpretation is to identify the additional constraints required on a truth theory to enable it to be interpretive.

A Tarski-style truth theory alone will not suffice as a meaning theory for a natural language. Whereas Tarski assumes meaning in providing a theory of truth, Davidson assumes truth to arrive a theory of meaning. However this leads to a theory that is merely extensionally adequate, not interpretive.[2]

In order for the Tarski-style truth theory to serve as a meaning theory for natural languages, then, it must be subject to additional constraints. Inspired by Quine (1960: ch. 2), Davidson proposes that the perspective of the radical interpreter (an interpreter interpreting another language from scratch – without recourse to his own prior linguistic knowledge), once identified will provide constraints on the Tarski-style truth theory that will render it adequate to the task of providing an adequate translation of an alien tongue.

2.1.1 Radical Interpretation and the Procedure of Radically Interpreting a Language

We might think about the radical interpreter in two ways: as choosing between a variety of already-completed meaning theories for a particular speaker or group of speakers; or as constructing one of these theories. In both cases the interpreter's access to the evidence is the same; however, in the former case where the radical

[2] To state that the right hand side of a biconditional statement is true iff the left hand side is will not demonstrate that the right hand side interprets the left hand side: ""Schnei ist weiss" is true iff grass is green" is true, but it does not specify the meaning of "Schnei is Weiss". The fact that natural languages contain demonstratives may mitigate this kind of objection, but can do nothing for T-form sentences such as ""Schnei ist weiss" iff snow is white and grass is green".

interpreter has to choose from between a variety of competing complete theories, the difficulties are fewer. In such a case we might imagine the radical interpreter faced with his subjects and in possession of a number of "translation machines" which when fed an arbitrary utterance of an alien tongue (together with information about context etc.), spits out an utterance in the interpreter's language (*Derangement*: 95). The "machine" which the radical interpreter will pick will be that which maximises the intelligibility of the subjects and this is done by making the subjects hold true the maximum number of sentences that are true according to the interpreter's view of the facts (a version of the principle of Charity, discussed below).[3]

In Davidson's early works, it would seem that he is more concerned with how the evidence available to the radical interpreter could constrain meaning, given a truth theory for a language, while appearing to take the construction of the truth theory for granted (*RI*: 136). Radical interpretation viewed in this way avoids questions that I am interested in here, for example, how can one get the procedure of interpreting an alien off the ground? And, how does the radical interpreter compile evidence for his theory? I therefore focus on the radical interpreter as a constructor of theories (a context of discovery) rather than as an adjudicator between already constructed competing theories (a context of justification).

---

[3] In the "complete theory case" Davidson (*Inscrutability*: 230) suggests that there is no reason to suppose that the interpreter will be able to narrow down the possibilities to just one "machine" – the evidence available to the interpreter underdetermines the theories available to him: two competing theories may render the subjects as holding true overall the same number of beliefs as are true in the interpreter's eyes. However these competing theories may assign different truth conditions to individual sentences. Hacking (1986: 454) comments that "indeterminacy implies that there are two acceptable systems of interpretation, J and K, such that, for some utterances, made by X speaking language L, J implies that S is true L if and only if p, while K implies the contrary". A discussion of indeterminacy of interpretation is, however, outside of the scope of this thesis.

2.1.2 The Principle of Charity and the Procedure of the Radical Interpreter

When attempting to break into an alien language, the radical interpreter can only avail himself of very limited information. Without knowing the language of his target speaker (or speech community) he cannot access their beliefs, and without knowing anything of their beliefs he would seemingly have no way of working out what an utterance produced by that speaker means. Yet to be capable of interpreting the alien language, the interpreter must arrive at a theory of meaning (interpretation theory) for that language, knowledge of which would suffice in enabling him to interpret the utterances of that language.

The radical interpreter cannot appeal to knowledge of the meanings of the alien language expressions, nor can he appeal to knowledge of "the complex and delicately discriminated intentions with which the sentence is typically uttered" (*RI*: 127). He *can*, however, help himself to as much knowledge of the physical world he likes, knowledge of the speaker's hold true attitudes (his attitudes to the truth of sentences in his language), and knowledge gained from observation of the brute facts of speaker's interactions with his environment.[4]

The radical interpreter can, it is claimed, through systematic observation of the behaviour of his subjects, determine which sentences they hold true under which environmental conditions. It will be necessary at this stage to limit the evidence to occasion sentences – those sentences which are held true at one time and false at

[4] By brute facts here I mean descriptive facts about such interactions that do not presuppose anything about the speaker's intentional states.

another ("It is raining", for example).[5] Having successfully identified such evidence the radical interpreter will be in possession of inductively confirmed laws of hold true attitudes such as:

(L) For all speakers *S*, times *t, ceteris paribus, S* holds true *s* at *t* iff p.

The *ceteris paribus* constraint is introduced because speakers are fallible and so apt to make mistakes (*L&L*: 184). The hold true attitudes of the speakers for occasion sentences will be the primary evidence for the interpreter; however, this will neither provide the interpreter with the meaning or truth of that sentence: all that can be inferred from a speaker's holding a sentence true under a particular set of circumstances is that the speaker holds that sentence true under that particular set of circumstances. Once the interpreter has assembled his evidence, he applies the principle of Charity (PoC). In radical interpretation, PoC permits the interpreter to infer from a speaker's 'hold true' attitudes to a meaning theory for that speaker's (or group of speakers') language, which has a truth theory at its core: it is a "principle which warrants inferring from L-sentences corresponding T-form sentences […] that can be used to project an interpretive truth theory for the entire language" (*L&L*: 184). The addition of PoC is intended to constrain the Tarski-style truth theory to make it adequate, that is to say, interpretive.

To see how PoC constrains interpretation, it is necessary to determine what it states: however, the principle can be formulated in different ways, and there is evidence for different formulations in Davidson's own work, as *L&L* note. *L&L* limit their discussion of PoC to those aspects of it relevant to moving from the inductively

[5] Vermazen (1982: 117) expresses concern as to whether radical interpretation can ever move beyond occasion sentences to sentences "couched in a vocabulary untapped by occasion sentences".

derived and defeasible sentences such as that specified above (L-sentences) to target theorems that the meaning theory aims to prove (T-form sentences).[6] *L&L* (185-90) find three formulations of the PoC in the literature; these appear in both primary and secondary sources, although, they claim, the failure to recognise these formulations as distinct has led commentators into confusion concerning Davidson's project.

> ***The Principle of Veracity*: *ceteris paribus,* a speaker's hold true attitudes directed toward occasion sentences are true;**
>
> **"For all speakers *S*, times *t*, sentences *s*, *ceteris paribus*: S holds true *s* at *t* iff *s* is true (*S*, *t*)"**

> ***The Principle of Charity*: *ceteris paribus,* a speaker's beliefs about his environment are in fact true.**
>
> **"For any speaker *S*, time *t*, belief *b*, *ceteris paribus*: *b* is a belief of *S*'s at *t* about and prompted by *S*'s environment iff *b* is true"**

> ***The Principle of Agreement*: *ceteris paribus,* the speaker and interpreter agree about the speaker's environment.**
>
> **"*Ceteris paribus*: the interpreter believes *p* iff the speaker believes *p*"**

L&L argue that the second of these must be taken as basic, since the first and third "transparently fail to achieve all that Davidson wants from the principle of Charity (restricted to it's application to the transition from L-Sentences to TF-sentences)". PoC needs to be characterised in such a way as to be sufficient to entail interpretiveness, which L&L state as follows:

---

[6] Verheggen (2006: 215-6) notes that the *point* of the principle of Charity is not to maximise true beliefs on the part of the speaker, but to maximise intelligibility; interpretation aims at understanding and understanding requires the attribution of true beliefs to the subject.

Interpretiveness:

For any sentences s_1, s_2, by and large, if s_1 is an inductively confirmed L-sentence and s_2 is its corresponding TF-sentence, then s_2 is interpretive.

Veracity and Agreement, they claim, are not sufficient. In the first place, applying the principle of Veracity to L-sentences will yield extensionally adequate biconditional TF sentences, but nothing guarantees that these sentences are interpretive. Veracity does not relate the speaker's beliefs about his environment to the meanings of the sentences he expresses. [7]

Moreover, the constraint of Agreement alone will not ensure that the resulting theory will be interpretive since it merely ensures that the speaker's and interpreter's view of the facts by and large maximally agree – not that the speaker's beliefs are true (since the interpreter's views may not be true either). As such, Agreement will provide an insufficient constraint: it will allow a move from a (range of) sentences specifying hold true attitudes to incorrect truth-form sentences which are then pressed into the role of target canonical theorems for the theory to prove. One can,

---

[7] The initial (relevant) objection to the use of a truth theory as a meaning theory for a language is that, since the theorems of a truth theory are extensional, they will not necessarily be interpretive. L&L (103) illustrate the objection as follows:
"Suppose in *L*, 'dog' means 'canine with a heart', and *L* has one proper name, 'Fido'. Suppose, as a matter of fact, something is a canine with a heart iff it is a canine with a liver (these predicates being tenseless). Then axioms a1 and a2 provide an extensionally adequate truth theory $\mathscr{T}$ for *L:*
a1. 'Fido is a dog' is T in *L* iff Ref('Fido') is a canine with a liver.
a2. Ref('Fido') = Fido
From $\mathscr{T}$ we can prove *T*- form sentence [1], but not *T*-sentence [2].
[1] 'Fido is a dog' is *T* in *L* iff Fido is a canine with a liver.
[2] 'Fido is a dog' is *T* in *L* iff Fido is a canine with a heart.
Since the *T*-sentence is [2], and not the mere *T*-form sentence [1], locating a predicate in the metalanguage to devise an extensionally adequate axiom for a predicate in the object language *L* fails to guarantee that the axiom is interpretive. Thus nothing guarantees that an extensionally adequate truth theory is *ipso facto* interpretive."

they suggest "*squeeze* Charity out of Agreement only by assuming that interpreter's beliefs are mostly true" (*L&L*: 191).[8]

The principle of Charity, as L&L formulate it is stronger than the principle of Veracity:

> **The Principle of Charity: *ceteris paribus,* a speaker's beliefs about his environment are in fact true.**
>
> **"For any speaker *S*, time *t*, belief *b*, *ceteris paribus*: *b* is a belief of *S*'s at *t* about and prompted by *S*'s environment iff *b* is true"**

What advantages does this have over veracity? L&L explain by appeal to the aetiology of a speaker's beliefs and consequent hold true attitudes. Hold true attitudes are a result of what a speaker believes and what his sentences mean. To this add the assumption that the speaker knows, "by and large", what his sentences mean and infers from this and knowledge of his beliefs that the sentences that express his beliefs are true. In the case of occasion sentences, a condition in the environment causes the speaker to have a belief, and this causes him to hold true a particular sentence that expresses that belief. L&L continue, "the suggestion is that we can arrive at the interpretation of a speaker's sentences in the light of correlations between his hold true attitudes and conditions in his environment, by assuming that, *ceteris paribus*, those beliefs about his environment prompted by it (his beliefs whose contents are expressed by occasion sentences) are true" (*L&L*: 188-9). On this formulation of Charity, the only sentences that the radical interpreter could, apparently, apply Charity to in moving from statements of hold true attitudes to T-form sentences, are occasion sentences.

---

[8] Davidson (CToK: 317) explains why interpreter and speaker cannot share false beliefs. See Dalmiya (1990) for a discussion of argument.

L&L object, however, that the principle of Charity, so stated, is insufficient to the task of rendering the resulting truth theory interpretive. The radical interpreter would not be able to sufficiently individuate the content of the speaker's utterance: Charity "requires *only* that a speaker's beliefs about his environment are true, and consequently, it does not secure his true beliefs being correlated with the conditions that prompt them in a way that ensures any statement of such conditions specifies the contents of his beliefs" (*L&L*: 193). This would need to be the case if the T-form sentences derived are to be interpretive. L&L illustrate this abstractly. Suppose a speaker, S, holds true a sentence *s* at time t iff condition p holds. If S holds true *s* at t he does so because he has a belief that q corresponding to his holding *s* true (i.e. he holds *s* true because he believes that q and he believes that *s* means q, and that if *s* means that q, then s is true iff q). However, Charity does not guarantee that p is identical with q, since a belief prompted by a set of conditions does not have to be about those conditions. In order for speakers to be interpretable, L&L suggest, they have to be in a "state of Grace", and to interpret them the interpreter needs the additional assumption that speakers' beliefs prompted by their environment are about those prompting conditions.

Although L&L do not illustrate their discussion here with a concrete example, *prima facie* it would seem that Grace would only come into play when the environmental stimulus was both necessary and sufficient for a belief, where that belief was not itself about that condition. Let us call any such belief that is caused by, but not about, an environmental stimulus, "belief$_p$". If the stimulus were necessary and sufficient for a "belief$_p$" then beleifs$_p$ would accompany all and only occasions where the relevant environmental stimulus was present. If the environmental stimulus

was merely sufficient, then there are potentially occasions where some other stimulus would elicit the same belief in the speaker. In such cases we might imagine that the radical interpreter, over a period of time, would be able to ascertain that the belief being expressed was indeed a belief$_{p,}$ as opposed to a belief about the corresponding environmental conditions; he could do this by employing a broad base of evidence, or by manufacturing situations against which he can test his L-sentences. From an empirical perspective, then, Grace might not be as integral a constraint as L&L suppose. Nonetheless, in cases where the stimulus is both necessary and sufficient for a beleif$_p$, Grace would be required.

Aside from the difficulty in identifying such a case in practice, what justification is there for stating that speakers are typically such a state of Grace? If radical interpretation is merely a theoretical device aimed at elucidating certain features of language, the fact that speakers do not, in practice, exist in such a state if Grace may not be a problem. However, if the claim is that people do in fact radically interpret their interlocutors then appealing to Grace would be to no avail. Furthermore Grace (along with the other formulations outlined here) appears only to be applicable to occasion sentences, or sentences that depend on observation of a speaker's interaction with some element of the environment. This *may* be enough to get a meaning theory for a language going. However, the application of Grace would make little sense at the latter stages of theory formation, where the radical interpreter applies the theory he has devised wholesale to the target language and adjusts it to render the subjects maximally intelligible. Here a version of PoC could be sought that merely maximises true statements on the part of the speaker. At any rate it is not clear that a principle such as Grace can be used at the latter stages of theory

construction, or uniformly appealed to in order to restore communication if the sentence uttered is not about a speaker's environment.

2.2 Radical Interpretation, Conventions, and Mrs Malaprop

Having provided a brief outline of the procedure and resources of radical interpretation, I proceed to address the question of whether radical interpretation casts any light on the interpretation of Mrs Malaprop. The question can be broken down into two: (1) Do we in fact interpret innovations, such as Mrs. Malaprop's utterance, 'radically'? and (2) Do the resources and procedure of radical interpretation provide sufficient constraints on interpretation to enable the hearer to assign the correct meaning to her utterance?

It is worth rehearsing a point I made in my discussion of Davidson's anti-conventionalism at the beginning of chapter 1. I suggested that a hearer's prior theory of a speaker need not map the conventions of a linguistic community that is putatively shared between hearer and speaker. There could be a case where practices between hearer and speaker are not shared (section 1.0.3), and thus there would be no conventions to be mapped by a hearer's prior theory of a speaker. Nevertheless the hearer might have a correct theory of the speaker that has succeeded thus far, but which cannot accommodate some further innovative utterance that the speaker makes.

Thus, in this section I focus on whether there is a role for radical interpretation in moving from a prior theory to a passing theory that successfully interprets an innovative utterance, and consider whether the resources available to the radical interpreter are sufficient for the creation of a passing theory of a speaker.

I posit two types of interpretation, radical interpretation (RI) and malaprop interpretation (MI), which assign different epistemic positions and different resources to the interpreter who is interpreting an innovative utterance.[9] For both types, the prior theory a hearer has for a speaker will not generate the meaning of the speaker's present utterance. For this reason both types of interpretation may loosely be called "radical" and we can say, in this minimal sense, that the MI also radically interprets the innovative utterance, but the interpreter is otherwise in an epistemically superior position. We might thus suggest that MI satisfies a weak construal of radical interpretation. As was noted in section 2.1.1, the true radical interpreter can only avail himself of very limited information: information concerning his subject's behaviour, interactions with his environment and knowledge of the physical world.

The basis for the distinction between the two types of interpreter/interpretation is rooted in certain equivocations on the part of Davidson in his work, as to what he means by radical interpretation on various occasions. It is worth, therefore, noting how "radical interpretation" is used by him.

2.2.1 Equivocations on Radical Interpretation

There is evidence in the literature that Davidson himself did not use the term "radical interpretation" entirely consistently. This inconsistency may relate to the dual role of radical interpretation as it appears over time in Davidson's work: the focus of radical interpretation appears to change from being a mere theoretical construct aimed at elucidating certain concepts and features of natural languages to something which language users as a matter of fact engage in.

[9] MI and RI will also abbreviate "malaprop" and radical interpreters.

In the first place, Davidson appears to equivocate on the question of whether radical interpretation of a language is possible. Davidson (*RI*: 125) makes the bold claim that “all understanding of the speech of another involves radical interpretation”. Davidson later appears to retreat from this position:

> I do not think I have ever argued for the claim that radical interpretability is a condition of interpretability. Not only have I never argued that every language is radically interpretable; I have not even argued that every language can be understood by someone other than its employer, since it would be possible to have a private code no one else could break. I do not think, and have not argued, that radical interpretation of natural languages must be possible; I have argued only that it is possible. The point of the "epistemic position" of the radical interpreter is not that it exhausts the evidence available to an actual interpreter, but that it arguably provides sufficient evidence for interpretation. (*RII*: 121)

If there exists a language that is not radically interpretable then it follows that not all language users engage in radical interpretation.

Yet Davidson later appears to return to his earlier position, suggesting that not only could language users radically interpret any language, but that they actually do:

> The fact that radical interpretation is so commonplace – the fact, that is, that we use our standard method of interpretation only as a useful starting point in understanding a speaker […].(*C&C*: 279)

However, there is a world of difference in using a standard method of interpretation as a starting point, and interpreting another speaker from scratch: this “radical interpretation” appears to be an attenuated version.

We might, sympathetically, suggest that Davidson has different things in mind when using “radical interpretation” in these different contexts. There is evidence that he is conscious of his different uses of radical interpretation:

> Reading over the papers I have written on interpretation over the last 27 years, it is clear to me that neither my terminology nor my views have held absolutely steady. I have sometimes used the term "radical interpretation" to refer to any interpretation from scratch, that is, without the aid of bilingual speakers or dictionaries, and sometimes to refer to the special enterprise of interpreting on the basis of a limited and specified data base.(*RII*: 128, n.2)

Thus if Davidson has in mind stronger and weaker sense of radical interpretation, we might suggest on his behalf that all people in practice radically interpret (weakly construed) whereas radical interpretation strongly construed is merely part of a theoretical device aimed at enabling us to understand more clearly the constitutive features of language. However, Davidson himself has not been as explicit as he could have been in making such a distinction and this has led to criticism.

The problem lies in accommodating these dual roles of radical interpretation: can reflection on radical interpretation elucidate both the nature of language in the abstract and the question of how language users interpret in practice.

Gustafsson (1999: 435) claims that Davidson is "vacillating between treating meaning-theories as mere representations of the linguistic abilities of a speaker and seeing them as playing a more substantial role in communication". He (437-9) suggests that Davidson attempts to combine seemingly incompatible approaches of attempting to systematically explain an idealized notion of language scientifically (which would require dispensing with anomalies), *a la Quine,* while, at the same time, taking the accommodation of linguistic anomalies as fundamental to the task of giving an account of language. The simplest response to the observation of anomalous linguistic phenomena would be to throw out such systematisation, (439) but Davidson prefers instead to replace his original single meaning theory with many theories that are constantly revised and modified (443). Moreover, Davidson's commitment to holism requires not only that we infinitely multiply our theories to account for all the quirks and regularities of actual linguistic practice, but also that "when a word

changes its meaning, or takes over the role of some other word or phrase, or when a new word is invented, this has repercussions on the theory as a whole" (445).

Ultimately, the sheer complexities involved in Davidson's revised account of meaning theories of language renders it deeply implausible, according to Gustafsson. On the early Davidsonian model, while the field linguist would be faced with a mammoth task, probably taking a number of years to complete, this task would not be Sisyphean (445). However, on the new picture, language would lack the required stability for the empirical procedure to get off the ground: the theory has to be static enough to be confirmed by the data, and if it is changing at a rapid rate then the data collected would be so sparse as to be very little use at all. Moreover, there are problems in trying to account for how a theory has been adjusted in the interpretation of an innovative utterance when one has no access to a prior theory to base the adjustment on: "[H]ow do we find out about the theories that a speaker has *without* assuming such previous knowledge? To this question, Davidson provides no answer" (447). The new model, then, according to Gustafsson is a methodological cul-de-sac, leaving no room for "honest empirical work". The old Davidsonian picture at least had the virtue of being useable as a framework for scientific enquiry. But if language users are successfully interpreting each other by using non-systematizable "wit, luck, and wisdom" (*Derangement*: 107), what is the use of the theory?

For Gustafsson at least, reflection on the procedure of the radical interpreter as a constructor of theories (collecting data, formulating t-sentences, constructing a theory that models a speaker's dispositions to use words etc.) can offer little insight into how speakers actually interpret one another, once the considerations of *Derangement* are taken into account. Davidson's view of the radical interpreter as

fundamental in our understanding of the “language” we are left with at the end of *Derangement* should not, then, be taken for granted.

2.2.2 The Resources and Strategies of the Malaprop Interpreter

In this section I make some suggestions as to what the MI can avail himself of when interpreting and illustrate some ways in which the MI might arrive at a successful interpretation of a linguistic innovation. This is done to highlight the differences between the MI and the RI traditionally construed, and to suggest RI is not helpful in explaining how we do interpret innovations on a day-to-day basis. After this, I will investigate whether the true RI could in fact interpret Mrs Malaprop, since, if he could, this might nevertheless elucidate certain features of language in general.

There is a very plausible way in which the PoC, or some version of it, will enter into our interpretation of, say, a malaprop. Interpreted along standard lines, the speaker of the malaprop appears to be unintelligible, we therefore charitably dismiss this interpretation. However, it is not as simple to attribute an alternative meaning to the utterance by the same method. There is a possible infinite number of replacement interpretations one might attribute, which could preserve the speaker’s intelligibility, general pattern of true beliefs and rationality: the trick in forming a passing theory is to hit upon the right one.

It should go without saying that our epistemic position when interpreting Mrs Malaprop is far different to that of the radical interpreter as mentioned above. The information that we do, in fact, avail ourselves of in interpreting non-standard and innovative utterances – utterances which cannot be generated by or assimilated into our prior theories of an individual as they stand, are subject to such weaker constraints as to make one wonder whether the process we engage can be called

radical interpretation at all. Davidson, however, seems to imply that we not only can avail ourselves of such extra information, but indeed must do so. Moreover, his admission that the interpreter of innovations can make use of his awareness of "interests and attitudes" (not just hold true attitudes) leaves the MI in a far superior epistemic position to the radical interpreter (*SAL*: 110).

Davidson (*C&C*: 278) further acknowledges that in actual instances of communication an important source of our awareness of our subject's attitudes and intentions (and potential future dispositions to use words non-conventionally) comes from our previous linguistic interactions with them. This occurs, for example, when we are primed by Donnellan to interpret "glory" as meaning *A nice knock-down argument*.

> McKay says that you cannot change what words mean (and so their reference if that is relevant) merely by intending to; the answer is that is true, but you can change the meaning provided you believe (and perhaps are justified in believing) that the interpreter has adequate clues for the new interpretation. You may deliberately provide those clues, as Donnellan did for his final 'There's glory for you'.(*Derangement*: 98)

It does not follow from this, however, that prior knowledge of the meanings of Donnellan's words, *in general*, are not necessary: the clues Donnellan provided for the interpretation of "There's glory for you" would have been no use to someone who did not well enough understand what he had said previously in the normal way.

The MI has a great wealth of information he can appeal in interpreting an innovative utterance: he can appeal to the attitudes and intentions, beliefs and desires of his interlocutor to varying degrees, he can perhaps even appeal to psychological facts he believes to be common to someone in his culture, or perhaps which he believes common to humanity. He can even appeal to knowledge of the speaker acquired through previous linguistic exchanges.

It may be worth considering some potential cases of interpreting innovations to see what resources the MI might appeal to in constructing a passing theory, and whether a true radical interpreter can appeal to those same resources.

a) "Detective work" and appeal to common linguistic practices through dialogue

There are many ways in which the successful passing theory may be derived from the prior theory in such a way that means that the advance linguistic knowledge a hearer has of a speaker can be employed in the successful construction of a passing theory. In the first place, an interpreter might appeal to linguistic practices he believes that he and the speaker have in common to work out the meaning of the words he does not know. Consider the following conversation with Mrs Malaprop.

> Mrs M: If I reprehend any thing in this world, it is the use of my oracular tongue, and a nice derangement of epitaphs
> I: I'm not sure I quite get you. Do you mean to say by that that if you understand anything in this world, it is how to speak as the common people do and how to attach appropriately descriptive adjectives to proper names?
> Mrs M: Yes. Pretty much. But by "appropriately descriptive adjectives" (by which I think you mean in my language "appropriately ascriptive conjunctives") do you mean words which characterise a person rather well and instinctly.
> I: Yes. Pretty much – if by "instinctly" (which in my language is "succinctly") you mean something like briefly….

We might say that the interpreter has "a good idea" what Mrs Malaprop is intending to get across as is testified to by his more-or-less accurate re-description of the offending phrases in his gloss. But in order to confirm his interpretation the interpreter would intuitively appeal to what he thought would be shared vocabulary – or practices of using that vocabulary in a common way.

b) Inference from prior knowledge of syntax and grammar.

Many of Mrs Malaprop's practices are in line with the interpreter's prior theory for her: a good number of the words, it transpires, are common to both parties and, moreover, syntax and grammar are common too. One must surmise in this case that the syntax together with the remainder of shared vocabulary gives the interpreter plenty of cues and hints as to how to interpret the expression as a whole. The fact that she is talking about her tongue, for example, prepares the interpreter to expect that what is to follow is some claim about her word use.[10] Thus an interpreter may be able to "fill in the blanks" of an utterance purely by deducing what would most likely fit in place of the unknown word, given the grammatical structure and other known vocabulary of the utterance (Bar-On and Risjord: 186).

Davidson's discussion of malaprops and their implications for conventionalist accounts of language focuses on non-standard vocabulary, but he makes little mention of syntax, although he does imply that syntax is less prone to deviation or innovation (C&C: 279). [11] The failure to consider the role of syntax, grammar and other properties of language in linguistic interpretation is not Davidson's alone.

---

[10] One might object that, given the anomalous words, the interpreter has no warrant to assume that *any* of the other vocabulary is being used by Mrs Malaprop in line with the interpreter's prior theory of her. However, one could respond that the interpreter is not consciously making an assumption about that at all – the default interpretation of vocabulary that the prior theory accounts for is automatic, and it is not until the elements of an expression are shared and so automatically understood are taken into account that the interpreter is in a position to re-adjust the theory to accommodate the non-standard uses of the other words. Thus without background of shared practices to appeal to, it is not clear that communication could, succeed in this instance.

[11] Davidson does suggest that we can *perceive* a grammatically garbled sentence as well formed (*Derangement,* p.95)– but perception and interpretation are not the same – there might well be an argument for this that stems from considerations of innate universal grammars. Alternatively we might as easily suggest that an interpreter can make an inference from what he knows of the vocabulary to the correct syntax and grammar, much as a student who lacks the relevant grammatical knowledge may be able to construct a reasonable pass at a Greek sentence from his knowledge of the vocabulary.

Laurence (1996: 280) attacks Lewis, suggesting that, while his account might provide a reasonable story about the conventionality of vocabulary, it cannot be generalised to cover other aspects of language such as syntax, grammar, phonology or morphology, and, moreover, that nobody believes that such an account could be provided. Laurence's comments here are pertinent because they seem to constitute as much an attack on Davidson as Lewis. In the assimilation of linguistic communication to general reasoning and communication in general, Davidson would surely agree with Lewis in so far as both see language processing as merely a part of "how we find our way around the world generally" (*Derangement*: 107). But just as it is impossible to give a Lewisian account of syntax, because "no one seriously believes that speakers and hearers have the requisite attitudes or pattern of reasoning concerning the phonological, morphological or syntactic properties of utterances as is supposed in the case of semantic properties" (1996: 278). It is not plausible that a hearer can work out the syntax, grammar etc. of an innovative utterance through reasoning about the attitudes and desires of the speaker alone. It is much easier to see how we can fill in a few blank words than manufacture a passing theory with an entirely new grammar, say. But if we recognise the grammar and syntax then it is often possible to make sense of an anomalous sentence by supplying new meanings for individual words.

c) Hearer's background knowledge of the speaker.

There are cases, then, where our ability to construct passing theories to accommodate innovations is parasitic on our prior theories. However, in the cases mentioned so far there is significant overlap between prior and passing theories: demonstrating that one can incorporate new vocabulary etc. into one's passing

theory does not entail that *some* prior knowledge of the hearer's linguistic habits is not required on the part of the speaker.

Let us grant, however, that there exist cases of successful interpretation where no such resources are appealed to. At the limit, a passing theory may not overlap with a prior theory at all. But even if no *linguistic* clues from prior linguistic exchanges were available to help interpret a particular future exchange, then there still might exist a considerable problem for the idea that one can, in general, communicate linguistically without advance knowledge of the idiolect of the speaker. Let us suppose that H's prior theory cannot generate the expression uttered by S and, moreover, the prior theory and the theory required to successfully interpret the utterance are entirely distinct. Suppose that S, who has hitherto been speaking English receives a knock on the head and as a result he can only produce utterances in Inuit.[12] It may be, that H, having known S well prior to the accident, knows a great deal about the intentions, interests and general attitudes of S. As such, without any prior knowledge of Inuit, but given the context, and through the exercise of his "imagination and awareness etc." and familiarity with the speaker, H may be able to work out what the Inuit expressions uttered mean as spoken by S on that occasion. In such a case, the interpretation would not be radical: the interpreter would assign meaning to the speaker's utterance in virtue of a *prior* assignation of intentional states to her. Because the interpreter knows the speaker well, he stands in a good

[12] Davidson requires that a speaker have a legitimate expectation that her utterance be interpretable by their intended audience (*Derangement*: 98), so let us assume, for argument's sake that the speaker believes for some reason that she will be interpretable – that for example, the hearer knows her well enough to interpret him. Or perhaps the requirement for having legitimate expectations that one be interpreted be relative to one's beliefs at the time, so for argument's sake, we might suggest that she either believes that she is still speaking standard English (as Mrs Malaprop does), or that everyone else around him is competent in Inuit.

stead to grasp what she intended him to understand by her utterance. In strict radical interpretation, the beliefs of the speaker and the meanings of her words must be applied simultaneously, the resources available to the radical interpreter too poor for him to access the one without the other.

In this case, the interpreter has interpreted a novel utterance without appealing to any general knowledge of the speaker's prior use of utterances; we could in such a case suggest that it is therefore right to say that past regularities in the speaker's use of words, and indeed an interpreter's knowledge of those past regularities does not determine the meaning of the expression being uttered. Thus unlike, cases (a) and (b) no use of the prior theory is made in interpreting the speaker.

The point is not that this is likely, but it does suggest that it is conceivable that an interpreter could know a speaker well enough to grasp what he intends to convey without having to appeal *in any way* to an understanding of the "language" that the speaker has used in the past. One may object that in this case the interpretation is not genuinely linguistic – one divines what the speaker intends to convey, not what the utterance means - but it does not follow that the speaker's utterance does not have this meaning (see discussion in section 3.4). Our ability to "get at" what a speaker is saying (*Derangement*: 92), even if the utterance's meaning is opaque to us, would surely one tool for day-to-day interpretation; however, it would not seem to be a tool available to the radical interpreter.

2.2.3 Limitations of the Radical Interpreter

True radical Interpretation has been described earlier in this chapter, however it is worth highlighting again the epistemic position of the radical interpreter: in order to consider which of the above strategies he may be able to employ.

We saw earlier, that the RI interprets an alien tongue from scratch. However to be sympathetic to the interpreter here we will allow that he has a prior theory for the speaker in question, however the prior theories and the correct passing theories are different. The RI must construct a passing theory that is different from the prior theory, as does the MI.

Nevertheless, given the limitations available to him, the RI cannot employ all of the above strategies. As we saw, the radical interpreter has limited non-linguistic information concerning the speaker. He can have as much knowledge of the physical world he likes, knowledge of the speaker's hold true attitudes (his attitudes to the truth of sentences in his language), and knowledge gained from observation of the brute facts of the speaker's interactions of with her environment. The interpreter *does not have* any knowledge of the meanings of the expressions in an alien language, nor any knowledge concerning the complex of beliefs, desires, and intentions of the speaker.

It is debatable whether RI could employ strategy (a). One's view of this may depend on how stringent the constraints on the RI's interaction with his subject must be. The conversation depicting (a) might appear to place the RI inside the intentional circle ("Do you mean by "x" that *y*?"), since he is assuming and testing the meanings of words by reference to other words rather than just tracking hold true attitudes. However, perhaps there is a sufficient way of redescribing this conversation in such a

way that the RI could be seen to be tracking hold true attitudes. So, for example the RI just utters the "redescription" of the sentence with appropriate changes in personal pronouns and gauges Mrs Malaprop assent. However, there would then have to be an additional assumption that Mrs Malaprop would not only be assenting to the sentence because it is true, but *also* because it is an accurate redescription of what she said. Thus there would still seem to be problems in allowing the RI the moves in (a).

We could grant the RI (b): given that he has access to a prior theory, he could on the basis of what he knows of the axioms and rules of composition stated by that theory hypothesise a limited range of possible meanings for the utterance. However, in order to confirm the hypothesis, much more work would have to be done (see following sections). The MI, perhaps by using this strategy in combination with others, would be better placed to have confidence in their interpretation.

The RI could not employ strategy (c), which requires attributing attitudinal states to the speaker prior to attributing meanings to his words. The radical interpreter cannot avail himself of any psychological facts about his subject at all, and he is not permitted to make use of his "empathy" (empirical assumptions concerning how like a subject is to him in virtue of having similar desires, interests, expectations, attitudes to our environment etc.).[13] L&L suggest that this would undermine the original project of radical interpretation. First, assumptions of this kind must be based

[13] Quine (1980, cited in Picardi (1999) introduces a notion of "empathy" in the course of the rejection of *shared* stimulus meaning in order explain how a translator can attribute content to a speaker's utterance give that, on his account, stimulus meanings are not *shared* and thus may not be similar for translator and subject. "The witnessing must be construed as essentially involving "empathy", i.e. projecting oneself into the position of the witness and issuing one's verdict from that position" (1999: 173). The notion of "projection" fits well with what I suggest here, although as far as the discussion it results from is concerned, Davidson himself rejects the notion of private stimulus meaning (e.g. MTE: 51-2)

on the behavioural evidence available to the radical interpreter and so cannot themselves form basic constraints. Secondly, taking on board assumptions about the psychology of individual speakers "would undermine the generality of results we would hope to obtain from reflection on the nature of radical interpretation" (L&L, 218-9). A simpler reason for refusing the radical interpreter "empathy" would be that it is not sufficiently radical: the RI would be attributing attitudes to the speaker that he is not entitled to (i.e. attitudes that are not merely hold true attitudes, but predictions about the interests of the speaker, what the speaker finds salient in the environment and so on). In practice, as I suggest in the next chapter, this is probably one of the most common ways in which we go about interpreting the meaning of innovative utterances.

The MI can, then, avail himself of a number of "short-cuts" that the RI is not entitled to. In day-to-day communication interlocutors are less subject to stringent constraints on "theory" confirmation: if the conversation continues to plan well, this they take to be confirmation enough of their interpretation. They may, in the light of further conversation "decide that they could have done better" in their initial interpretation (*Derangement*: 101); however, this is strictly limited to what serves their practical purposes best in the limited sphere of their communication with the speaker, not whether their overall theories, applied wholesale, maximise the speaker's true beliefs.

2.3 Applying the Principle of Charity

In further investigating whether the RI could interpret a malaprop, it is worth considering the stages of radical interpretation of an utterance in simplified form to investigate what help the principle of Charity would be in securing a correct

interpretation.[14] Imagine the field radical interpreter embarking on a new project of interpreting a speaker of an alien language. As yet he has no evidence to go on, but encounters his first utterance. At the time it is raining in the vicinity. His subject utters some sentence "@s@". What is the interpreter entitled to draw out of this first interaction?

> **Stage 1a)** The interpreter, suppose, writes in his notebook: "At a particular time, t, when it is raining in the immediate vicinity, speaker 1 utters "@s@". I repeat the sounds he made back to him and he indicates assent".

Little can be made of this so far, as there may be differences as to what the speaker and hearer find salient in the environment: perhaps the speaker, unbeknownst to the hearer, also notices a dog. Thus:

> **Stage 1b)** The field interpreter would observe the speaker over a number of occasions where similar environmental conditions prevailed. Only then (if the speaker uttered "@s@" on all of these occasions) would he have warrant to postulate the inductively confirmed law-like statement of a hold true attitude:
>
> - (L) For speaker s, times t, context c, 1 holds true "@s@" iff it is raining.

Presumably this process would be repeated to assemble a whole host of statements of hold true attitudes which would be similarly justified on the basis of behavioural evidence (the regular coincidence of the hold true attitude with the relevant environmental conditions).

> **Stage 2)** At this point the PoC is introduced. As we have seen, the application of the principle of Grace (the strengthened principle of Charity that L&L provide) would yield a T-form sentence. The interpreter might note down:

[14] I follow L&L (196-7) here in the characterisation of the procedure.

- (TF) Ceteris paribus, for speaker 1, t, c, “@s@” is true iff it is raining.

Stage 3) The interpreter would repeat this process for as many observation sentences as required, and then construct a truth theory which yields the TF sentences specified in stage 2.

Stage 4) The interpreter can then derive further novel TF sentences and test them against the behavioural evidence of the subject language speaker or group.

Stage 5) The principle of Charity may be applied again at the end stages of the interpretive process, whereby any alternative theories that the interpreter has produced may be examined to see which generates theorems that make more of the beliefs purportedly expressed by the speaker’s utterances true “according to the interpreter’s view of the facts” (*RI*: 136).

From this outline, we do not see the PoC applied at the initial stage of the interpretive process, but rather in the secondary and final stages. However, the position of the radical interpreter of Mrs Malaprop looks to be at stage 1a. Since the interpreter at 1a is not yet at the stage appropriate to applying PoC, it would seem that it cannot be the application of the principle of Charity that enables him to successfully interpret her utterance. But if PoC (in some formulation) is necessary to ensure that the T-form sentence is interpretive, then the interpreter, in this case, has not succeeded.

Mrs Malaprop makes her statement only once: there are no grounds as yet to consider a theory which accounts for it to be confirmed or interpretive. It would seem that the RI only has grounds to re-adjust the theory after assembling a new (L) sentence stating her hold true attitude (requiring her behaviour to be understood over time and a regularity in her hold true attitudes to emerge). It is only once this (L) sentence is discovered that the principle of Charity can be applied.

Moreover, L&L suggested that the initial stages of theory construction were limited to occasion sentences: as we saw, their formulation of *Charity*, as opposed to *Agreement* and *Veracity* stated that "For any speaker *S*, time *t*, belief *b*, *ceteris paribus*: *b* is a belief of *S*'s at *t* about and prompted by *S*'s environment iff *b* is true". But the examples that Davidson provides in *Derangement* and elsewhere are not occasion sentences. It would seem that the radical interpreter is restricted to occasion sentences until the theory is sufficiently formed to give a substantial picture of the speaker's idiolect such that non-occasion sentences can be generated. This is all the more true if L&L are right to suppose that the stronger principle of Grace is required to render the theory constructed as interpretive.

The point is that Davidson's initial aim with radical interpretation is to show how a truth theory, suitably empirically constrained can serve as an interpretive meaning theory for a language. Even supposing that the procedure of radical interpretation does, in fact, provide those constraints for a stable language (what Davidson (*SAL*) confesses he is guilty of calling a language), and thus the theory produced is guaranteed to be interpretive in that case, the empirical constraints on the theory *are not met* with regard to the Mrs Malaprop case. The radical interpreter of Mrs Malaprop, merely by postulating a new theory in the light of a one-off statement, cannot guarantee that the theorems his theory yields, theorems such as "For Mrs Malaprop, times t, context c, "There's nice derangement of epitaphs" is true iff …. A nice arrangement of epithets ect….", are interpretive.

Thus we must address two further related questions concerning the criteria for success in successfully understanding or interpreting a speaker:

(1) In virtue of what is the interpreter said to have *understood* Mrs Malaprop?

and

(2) When is understanding said to have taken place?

Let us suppose for argument's sake that, as a matter of fact, it is true that for speaker 1,t,c, "@s@" is true iff it is raining.

At what point are we entitled to say that the radical interpreter has understood, or grasped the meaning of the particular utterance "@s@"? At the one extreme, we should hesitate to say that an interpreter has "understood" the speaker if he applies the correct interpretation of the words purely by chance.[15] At the other extreme, the hearer is required to have a complete theory of the speaker's idiolect in order to arrive at a correct interpretation of the utterance. This position appears to be that of Davidson: "the interpreter must be able to understand any of the infinity of sentences the speaker might utter" in order to render interpretation possible (*RI*: 127), and that "we can interpret a particular sentence provided we know a correct theory of truth that deals with the language of that sentence" (138).

The problem, as has already been discussed, is of ascertaining that the right side of the biconditional in a T-sentence is not only true iff the left is, but also that it is *interpretive*. In Davidson's initial formulation of radical interpretation, one cannot be taken to have interpreted an individual utterance of a sentence in a language without

[15] The speaker, say, utters a series of sounds "s@s" meaning to express that it is raining, and background noises in the environment conspire in such a way as to distort what the interpreter hears such that she seems to herself to hear the English "It is raining". It would seem that through the bizarre set of circumstances successful communication has occurred, since the speaker meant to express that it is raining, and the interpreter understood him as so expressing. But it would be hardly accurate to say that the interpreter interpreted, understood or grasped what the speaker's *words meant*. Not least because if she incorporated a T-form sentence containing "It is raining" on the left side of the biconditional into her theory of the speaker the theory would be incorrect.

also having the entire theory for that language. [16] The resources of the radical interpreter might, therefore, appear too poor for the job of constructing a passing theory.

2.4 Further Difficulties for Radical Interpretation

We might respond on Davidson's behalf that there are "degrees of correctness" in interpretation, as Kemmerling (1990: 93) does:

> An act of linguistic communication between S and H is *successful* if the hearer eventually arrives at a correct interpretation of the speaker's utterance. The hearer's interpretation is *completely correct* only if the following holds: the interpretation yields (1) what the speaker's words mean in his mouth on the occasion of the utterance, and (2) this result is arrived at by the hearer's application of exactly the theory which the speaker intends him to use to interpret this particular utterance. (The hearer's interpretation may be *correct* even if it is not completely correct. Correctness of interpretation allows for degrees.)

We might allow the radical interpreter in this case a less stringent constraint on correctness. However, Kemmerling's response raises further worries.

2.4.1 Timing issues: Humpty and Alice

First, Kemmerling's use of "eventually" here is not obviously in line with what Davidson has in mind in *Derangement*. Davidson's claim that the hearer and the speaker's theories must coincide only after the utterance has occurred might appear to be in line with this; however, Davidson denies that Humpty Dumpty has said anything meaningful because he had no legitimate expectation that he would be successfully interpreted *by Alice* as meaning what he intended his words to mean

[16] See also Davidson (BM: 148): "[T]o interpret a particular utterance it is necessary to construct a comprehensive theory for the interpretation of a potential infinity of utterances. The evidence for the interpretation of a particular utterance will therefore have to be evidence for the interpretation of all utterances of a speaker or community."

when he so uttered. Yet, Humpty and Alice's *passing theories* do in some sense later coincide: only a few breaths later Humpty informed Alice what he intended "glory" to mean. Of course, Alice didn't know what "glory" meant until Humpty told her; but upon being told, Alice did indeed know what Humpty had intended to mean when he uttered it. Their passing theories coincide, but this cannot legitimise Humpty's expectation. There must be some reasonably close proximity between when an utterance is encountered and when an utterance is understood, if an account of our understanding of utterances that our prior theories will not generate is to shed light on the day-to-day business of linguistic communication.

This is all the more apparent when we consider the role of the *speaker* in securing successful communication. Davidson claims that the speaker must have a legitimate expectation that his words will be interpreted in the way that he intends them to be interpreted for the meaning of his utterance to be so fixed. These legitimate expectations are derived from the situation (environmental conditions, knowledge of hearer etc.) *as it stands at the time of the utterance.* It would not seem in the spirit of things to suggest that someone can legitimately expect their words to have a particular literal meaning, if their intention is such that the hearer can only *eventually* (after 5 years? 10 years?) grasp what they intended to communicate.

2.4.2 The Observer Effect: Radical Interpreters and Speaker's Intentions

One might respond to my claim that the notion of a field radical interpreter coming to understand an utterance over an extended period of time is implausible when considering how a malaprop speaker comes to be interpreted, is as follows: we should rather consider the case from the perspective of an ideal radical interpreter. He, perhaps, comes equipped with all the relevant evidence and the processing

power required to interpret Mrs Malaprop. While it might take a “human” radical interpreter a number of years working in the field to confirm his interpretations, the ideal interpreter can do this in a number of seconds, let us say. However, faced with this theoretical response, on may counter-object by considering the role of the speaker and the speaker’s passing theory in the model of linguistic understanding.

Recall that Humpty is out of it – he can’t mean by “glory” what he intends to mean by it - because Alice does not have the relevant background knowledge to interpret his utterance. However, if we were to bring in the ideal radical interpreter here it might be that the ideal radical interpreter could determine from the evidence available to *him* what Humpty meant. But if the ideal radical interpreter could interpret him (as one might suspect he could), then it would seem that “glory” could mean a nice knock-down argument – which Davidson denies. So Davidson would not, if he were to be consistent, consider the perspective of the radical interpreter to be of any help here.

Moreover, reflection on the perspective of the radical interpreter could not explain how *Alice* could understand Humpty, and as such is no help in an account of what Humpty meant when *talking to Alice*. Humpty cannot base his intention to mean something by his words to Alice on the basis of what a radical interpreter could work out over time, or what an ideal radical interpreter could work out more or less instantaneously. Now that meaning is considered to be fixed in relation to a particular speaker and hearer, the radical interpreter seems to be out in the cold.

Again, the nature of one’s audience on any given occasion determines what one can legitimately intend to mean by one’s words. Thus, in order for a speaker’s utterances to mean something when speaking to a radical interpreter, she must

speak in accordance with her legitimate expectations of what the *radical interpreter* would be able to interpret. But then, in so far as the speaker is understood by the radical interpreter, all the radical interpreter can understand is the speaker's language (passing theory) *as geared to the radical interpreter*. We might assume that she has different passing theories for her family, friends, business associates etc. The radical interpreter then, can only ever have a very partial, and *artificial*, understanding of what the speaker's language or passing theory is, since the radical interpreter can only ever interpret the speaker's passing theory that is geared to him.

Finally, since there is no such thing as a radical interpreter: it would seem that there is no such thing as being able to form a legitimate intention to be interpreted by one, just as there is no way to form an intention to speak in such a way as to be interpreted by a unicorn. A radical interpreter would only be able to interpret an equally mythical radical speaker.

In short, Daivdson's focus on the speaker's and hearer's awareness of each other, and the necessity of speaker's and hearer's passing theories to converge for communication to take place requires an account of communication and meaning. Such an account would seem to be situated firmly in the realm of the actual. It is not clear how the radical interpreter can help us with this.

2.5 Conclusions

To suggest that much light can be thrown on how hearers come to understand speakers by referring to the task of the radical interpreter does not seem plausible: in actuality the time elapsed between initially encountering an utterance and formulating a theory might be very large indeed. On the old, stable, conception of language, where the radical interpreter is viewed as an observer of the speech behaviour of a

community, with (perhaps) a conception of language as determined by the conventions of that community, radical interpretation may have been informative. However, given the fluxy nature of the meaning theories that follow from Davidson's claims in *Derangement*, the task for the interpreter would be positively Sisyphean.[17].

This seems consistent with Davidson's own claims that imagination and awareness of the attitudes and intentions of our target speaker are the resources we appeal to when interpreting innovations. These resources are not available to the radical interpreter, who must restrict his knowledge of a speaker's attitudes to hold true attitudes. If a speaker's dispositions to use words were unstable and changed from one minute to the next, that problems for a radical interpreter mapping them would be intractable. With a constantly changing evidence base and no guarantee of sufficient evidence to confirm a statement of a hold true attitude, the radical interpreter would not be in a position to adjust his theory, let alone confirm it.

However, Davidson appears to also use "radical interpretation" in a loose sense when discussing actual interpretation of innovative sentences. I have suggested that the actual interpreter of malaprops is in a superior epistemic position to the true radical interpreter and has access to resources that the radical interpreter does not. I have further suggested that these resources enable the malaprop interpreter to interpret an utterance as the speaker intends. This is enough to support the claim, made by Davidson, that past regularities in a speaker's speech behaviour

[17] Gustaffson (2003: 445): "The theory which the data are supposed to confirm or falsify cannot be assumed to be in constant flux, but must be held constant long enough to guarantee that a sufficiently rich database can be collected. As was indicated by the previous, `static' scenario, collecting data enough to construct just one theory of meaning would probably require several years of intense field research. But now, if meaning theories are supposed to be very short-lived, then data will have to be collected much faster. Consequently, the available database will inevitably be very sparse, and so its usefulness for theory-construction will be practically nil."

do not determine what a speaker's utterance means on a given occasion, and thus, *a fortiori*, the conventions of a linguistic community do not determine the meaning of any given utterance of its putative members either.

One deflationary conclusion of the chapter would seem to be that it becomes increasingly difficult to tell what Davidson has in mind by radical interpretation – particularly in his later works. I have suggested that the waters are muddied because Davidson has moved from using radical interpretation as merely a constraint upon a Tarski-style truth theory aimed at modelling an interpretive theory of meaning for a language in general to something akin to a skill used in the actual day-to-day business of communication.

2.5.1 Limitations: What Determines Meaning and What is Required for Interpretation?

Above I have suggested the MI can, on occasion, interpret utterances that are not in line with past regularities. This by itself should be enough to suggest that the meaning of *that* utterance is not determined by past regularities. This is a crucial point: if reflection on the procedure of the MI results in the claim that utterance meaning is not determined by past regularities, then reflection on the procedure of the RI would, *prima facie*, not seem to be necessary to our account.

However, the case has not been made that *in general* communication does not rely on past, shared regularities. In section 2.2.2 we saw three cases of interpretation that could not appeal to prior theory alone. In case (a) and (b) cues were taken from the prior theory. Even in case (c) the interpreter could not have gained access to the beliefs and attitudes of the speaker without prior successful linguistic exchanges with him. An infinite regress would seem to threaten here, for if

the previous linguistic exchange succeeded for the same reasons as stated above, then *that exchange* could have only succeeded through the interpreter's knowledge of the speaker acquired through linguistic exchanges prior to *that.* At some point along the chain, it seems, the interpreter would have had to grasp the meaning of the speaker's utterance simply in virtue of knowing what that utterance meant.

So while, so far, we may accept that *on occasion* an interpreter may interpret a speaker successfully with a passing theory that is entirely distinct from a prior theory, it does not follow that *in general* prior knowledge of how a speaker is disposed to use words is not a necessary feature of linguistic communication.

An appropriate analogy might be that, although on any given day a human may go without food or water, it does not follow *that in principle*, *or in general*, a human may go without water. Importantly to this analogy the human can go without food and water on any given day because *in general* she eats and drinks: As Bar-On and Risjord say, "the interpretive flexibility manifested in our everyday treatment of malapropisms may be, for all that has been shown, parasitic on the expectation of general adherence to linguistic conventions shared across speakers or at least over time" (Bar-On and Risjord, 1992: 185).

Chapter 3: Speaker Meaning and Utterance Meaning

3.0 Nice Arrangements and Nice Knock Down Arguments: Speaker Meaning and Word Meaning, First Meaning and Literal Meaning

This chapter examines the distinction between speaker meaning and first meaning (which Davidson ascribes to a speaker's words rather than the speaker herself), and offers an explanation as to why Davidson is committed to the view that Mrs Malaprop's *words*, as well as Mrs Malaprop herself, mean *a nice arrangement of epithets*.[1] I will begin by considering Davidson's comments on Donnellan's referential use of proper names and definite descriptions in *Derangement* and suggesting that the discussion there accounts for why Davidson believes Mrs Malaprop's words mean what she (and Davidson) believes they mean *in her idiolect*. I will suggest, however, that considerations to do with the distinction between first meaning and speaker meaning do not impact on the procedure of the interpreter: even granting that Mrs Malaprop's words mean what she intends them to mean, it does not follow that the interpreter interprets what her *words* mean rather than what she means by those words (the content she intends to convey by using those words); nor does it follow that what a speaker intends to be understood by her utterance necessarily coincides with what that utterances means when she so utters (Talmage: 1994).

[1] Davidson takes first meaning and literal meaning to be the same thing. However, recent attempts to prize apart first meaning and literal meaning may lead to some terminological confusion (Talmage: 1994). Until I discuss Talmage's argument in section 3.3, I will be using the terminology first meaning/literal meaning to describe meaning that is ascribed to words or utterances and distinguish this from speaker meaning which is the primary information that the speaker intends to convey.

3.1. "Whose Meaning is it Anyway?"

It cannot be taken for granted that "A nice derangement of epitaphs" so uttered by Mrs Malaprop means *A nice arrangement of epithets*. To do so would be to beg the question against the defenders of conventionalism. They would, broadly speaking, consider the common language to be primary and to consist of a set of rules or conventions to which a speaker of that language must hold themselves responsible. If those conventions have so determined that "A nice derangement of epitaphs" means *a nice derangement of epitaphs*, then when Mrs Malaprop utters the expression her words will mean *A nice derangement of epitaphs* even if *she* meant *A nice arrangement of epithets* by so uttering. One cannot, therefore, begin to refute the common language view by assuming that Mrs Malaprop's utterance has a non-standard meaning. Rather an independent argument must be formulated from which the possibility of non-standard meanings results.

George (1990) suggests that, even on a conception of language that sees the idiolect as primary, what a speaker means and the meaning of the expression uttered can come apart. He asks us to consider the case of Smuth, "who believes that "ingenuous" means just what "ingenious" does, assuming that, perhaps "the first is a variant pronunciation of the second, or that it is a less common form" (279). However, George believes that in such a case, even though Smuth believes that "ingenuous" means *ingenious*, we are still in a position to say that Smuth is wrong about the meanings of words in his idiolect. Even in his own idiolect, despite Smuth's beliefs, "ingenuous" does not mean *ingenious*: we are able to make a distinction between what Smuth believes about what words in his idiolect mean, and what they actually

mean in his idiolect. Thus we can make a distinction between what *he* means when he so uses "ingenuous" and what the word "ingenuous" means (280).

One may object however, that the word "ingenuous" simply means in Smuth's idiolect what "ingenious" means in ours – Smuth is not wrong (George calls this the "no-error" theory); however, George responds that this is unconvincing because it fails to capture the "phenomenology of error".

> Consider the case of word meaning. When I came to believe that 'livid' does not mean *red* but *bluish-grey*, I then took myself to have been in error. When I thought back to the times when I used the word, I felt like I was discovering what I had actually said, as opposed to what I thought I had. On the no-error view, however, such reactions are out of place. According to it, there are really two words that are pronounced the same, one in the communal language with the meaning *bluish-grey,* the other in my (then) language meaning *red*. […] I take myself to have always used the very 'livid' that I now employ – except that in the past I was mistaken about its meaning. […] At [no] point of self-correction did I take myself (or did others take me) to have started using a different word. Rather,[…] I took myself to be employing the same word as always but now correctly.(1990: 288-9)

To George we may simply object that the phenomenology of error has little to do with it: his intuition derived from his own first person perspective on the meanings of his words could be mistaken.[2]

On the conventionalist view, if speaker meaning and conventional meaning diverge, then the speaker has got it wrong. We might respond on Davidson's behalf that he is not concerned with "getting it right" but rather with "getting away with it". But nothing so far has been produced to show that a given speaker "got away with it" (accepting that he does) because a hearer interpreted his *utterance* in the way it was

[2] Given that a radical interpreter would assign different meanings to the words in George's "then" and "now" idiolect, it is probably fair to say that Davidson would dismiss George's first-person perspective on what has happened here. Cf. Smith (2006a), on Davidson's prioritisation of the third-person perspective over the first-person perspective.

intended, rather than because a hearer interpreted what the speaker intended to convey *in spite of* that utterance.

Yet, it is unfair to take Davidson's assertion that Mrs Malaprop's utterance bears a non-standard meaning as merely an assumption. We should rather see it as a natural implication of his views on language and interpretation more generally.

3.2 Davidson on Donnellan

Donnellan (1966) made a famous distinction between different uses of definite descriptions that might help us in clarifying why Davidson attributes non-standard meanings to utterances. The referential and attributive uses of a definite description may be briefly characterised as follows: in an assertion containing a definite description used *attributively*, the sentence states something about whatever fits the definite description, and this seems to fit with a traditional view of the nature and role of a definite description. Donnellan's referential use, however, designates the use of a definite description in a sentence whereby the definite description is used to identify a particular object which the speaker wishes to communicate to his audience.

Davidson makes use of Donnellan's distinction, in part as a foil to clarify his own view. Noting that Donnellan has made a plausible distinction between different uses of definite descriptions (and proper names), Davidson goes on to remark that on being attacked by McKay (1968) for espousing a "Humpty Dumpty" theory of meaning:

> Donnellan, in answer, explains that intentions are connected with expectations and that you cannot intend to accomplish something by a certain means unless you believe or expect that the means will, or at least could, lead to a desired outcome. A speaker cannot, therefore, intend to mean something by what he says unless he believes his audience will interpret his words as he intends (the Grician circle) (*Derangement*: 97).

Thus, like Mrs Malaprop, Donnellan "gets away with it" when he says, on the basis of his exchange with McKay, *he* can use "glory" to mean a nice knockdown argument.

However, Davidson rejects any connection between Donnellan's response here to McKay and Donnellan's own distinction between referential and attributive use of definite description: some "firm sense of the difference" between speaker meaning and word meaning can solve the problems that arise in Donnellan's initial discussion, he claims (*Derangement*: 98).

It is not *prima facie* obvious why Davidson should dismiss Donnellan's earlier (1966) remarks as irrelevant to the case in hand. Why should the Jones case be a case of speaker meaning, but "glory" as used by Davidson *mean glory*? As such it is worth considering in more detail what Donnellan says:

> Suppose I think I see at some distance a man walking and ask "Is the man carrying a walking stick the professor of history?" We should perhaps begin by distinguishing four cases at this point. (a) There is a man carrying a walking stick; I have then referred to a person and asked a question about him that can be answered if my audience has the information. (b) The man over there is not carrying a walking stick but an umbrella; I have still referred to someone and asked a question that can be answered, though if my audience sees that it is an umbrella and not a walking stick, they may also correct my apparently mistaken impression. (c) It is not a man at all but a rock that looks like one; in this case, I think I have still referred to something, to the thing over there that happens to be a rock but that I took to be a man. But in this case it is not clear that my question can be answered correctly. This, I think, is not because I have failed to refer, but rather because, given the true nature of what I referred to, my question is not appropriate […]. (1996: 295-6)

For present purposes I am interested in case (c). Donnellan envisages a situation where a speaker has a mistaken belief about some feature of his environment and utters a sentence that expresses that belief. Let us simplify the sentence for current purposes to "The man is tall" keeping the intended referent as a

rock.[3] Donnellan would wish to assert that the speaker of this sentence has successfully referred to something (in this case a rock) and that *he said something true of it*.

In *Derangement,* Davidson endorses Donnellan's idea that one cannot use any word to mean anything, as Humpty Dumpty does, because one cannot intend use a word in a particular way if one does not have a reasonable expectation of succeeding. For Donnellan (1968), the response to Wittgenstein's (1950: §510) challenge to "say it's cold here" and mean *it's hot here* is to "ask someone to flap his arms with the intention of flying. In response he can certainly wave his arms up and down, just as one can easily on command say the words "It's cold here." But this is not to do it *with the intention of flying*" (252). The formulation of an intention requires a reasonable expectation of success. For Davidson, Mrs Malaprop has such a reasonable expectation: she can intend to mean *a nice arrangement of epithets* by "a nice derangement of epitaphs".

Yet Donnellan would also appear to have a legitimate expectation that "the man is tall" will be understood as picking out the intended object by his audience and saying of it that it is tall. Donnellan believes that the object he is referring to is a man, and so necessarily believes that his belief (that the object over there is a man, and that it is tall) is true, and also believes that, other things being equal, his audience will share that belief. Mrs Malaprop believes that "derangement" means what "arrangement" does in standard English, that she is speaking standard English, and that her audience understands standard English: put more simply, she believes that her audience uses the word in the same way as she does. Given their beliefs, the

[3] This removes any complicating factors such as indexical pronouns.

expectation of being interpreted on the part of both speakers ought to be legitimate in the way in which Donnellan suggests, and Davidson endorses.

However, when viewed against the background of Davidson's commitments to the compositionality of language, and holism in attitude content, we can suggest why Davidson distinguishes Donnellan's use of "glory" in his response to McKay, from what Donellan says about the referential use of definite descriptions. It also elucidates on Davidson's distinction between speaker meaning and utterance meaning.

For Davidson, the compositionality of language explains a competent speaker's ability to understand an infinite number of sentences based on knowledge of only a finite set of axioms and rules; it also shows that complex expressions can be understood on the basis of the meanings and arrangement of their parts. Any meaning theory must capture languages as demonstrating this compositionality. The primitives of a language can only be understood in term of the contribution that they make to the whole, as Davidson notes:

> If sentences depend for their meaning on their structure, and we understand the meaning of each item in the structure only as an abstraction from the totality of sentences in which it features, then we can give the meaning of any sentence (and word) only by giving the meaning of every sentence (and word) in the language. Frege said that only in the context of a sentence does a word have meaning; in the same vein he might have added that only in the context of a language does a sentence (and therefore a word) have meaning. (*T&M:* 22)[4]

In his commitment to holism in attitude content, Davidson endorses the idea that "in attributing any given attitude to a speaker, one is thereby committed to a

[4] Davidson (RII:24) later admits to some exaggeration in the passage quoted here. In the light of Davidson's admission, *L&L* (213) gloss the passage as follows: "Its point [is] only to emphasise that to understand a single sentence is to be in a position to understand an indefinitely large range of other sentences". Even this weaker interpretation suits my purpose here, however, which is to claim that what "man" means in Donnellanese must be understood in the context of a large and related portion of Donnellanese (what it means when it appears in Donnellan's (potential) sentences in general, and not merely on this particular occasion).

perhaps endless range of related attitudes, though, importantly, not any specific set of additional attitudes" (*L&L*: 211-2). Attitudes can therefore not be attributed to a speaker individually but interconnected patterns must be attributed instead: "before some object in, or aspect of, the world can become part of the subject matter of a belief (true or false) there must be endless true beliefs about the subject matter" (*T&T*: 168).

> To believe, or speculate, or wonder whether a gun is loaded, one needs to believe such things as a gun is a weapon, that it is a more or less enduring physical object and so on (*T&T*: 157).

This impacts on the present discussion when we take into account both Donnellan's background patterns of beliefs concerning the object standardly referred to by the use of "man" in his language, and his background patterns of beliefs concerning when it is appropriate to use "man" in the various sentences he constructs.[5] Thus Donnellan would presumably believe that a man is a rational animal, a man is mortal and so on: beliefs that he would not hold true of rocks. However, the background patterns of beliefs that relate to Mrs Malaprop's use of "derangement" will, we might expect largely match the range of beliefs we have which underpin our use (as speakers of "standard English") of "arrangement".

Accordingly, we would, therefore, expect to see Mrs Malaprop holding true the sentence "The flowers are badly deranged" in circumstances where the flowers are badly arranged, but not expect to see Donnellan holding true the sentence "I am caught between a man and a hard-place" in circumstances similarly relevant to his

[5] I do not mean by this that Donnellan holds himself responsible to any conventional standard. "Appropriate expression" in this case means whatever is most likely to enable Donnellan to communicate what he intends to. Davidson (e.g. *SP*:113-5) would acknowledge that this sort of "appropriateness" is integral to one's words meaning what one intends them to mean – what is appropriate is what enables Donnellan to make himself interpretable to another.

utterance. From this we can see why “man” does not mean *rock* in Donnellanese, but “derangement” will mean *arrangement* in Malapropese.

So Davidson’s suggestion that, just because Donnellan is mistaken about the nature of the object he is referring to, this does not change the meaning of “man” in Donnellanese is consistent with commitments he voices in earlier work, the referential/attributive distinction has “nothing to do with words changing their meaning or reference”.

> Similarly for Donellan’s claims that Jones has said something true when he says ‘Smith’s murder is insane’, provided the man he believes (erroneously) to have murdered Smith is insane. This is done intentionally all the time, in irony or metaphor. A coherent theory could not allow that under the circumstances Jones’ sentence was true; nor would Jones think so if he new the facts. Jones’ belief about who murdered Smith cannot change the truth of the sentence he uses (and for the same reason cannot change the reference of the words in the sentence). *(Derangement: 98)*

Davidson’s view is largely in line with Kripke (1977: 264-5) here: that what the speaker refers to is what the speaker believes to “fulfil the conditions required for being the semantic referent of the designator”, if Jones has a specific intention to talk about a man who he thinks fulfils the conditions required for fitting the description “Smith’s murder” but is mistaken about who that is, he has managed to speaker-refer to the man in question, but nevertheless, we may say his disposition to use the word “Smith’s murderer” is to use it to refer to someone just in case that person murdered Smith.

3.2.1.Does “Glory” Mean A Nice Knock Down Argument?

Is Davidson’s endorsement of Donnellan’s claim that “glory” *means* a nice knock-down argument incongruous with his commitments to compositionality and holism?

We could reasonably claim that Mrs Malaprop's use of, say, "derangement", is in line with her dispositions to so use it. However, given Donnellan's associated beliefs about "glory" and *glory*, the other contexts in which he is apt to use the words threatens to give the lie to the idea that "glory" means anything other than *glory - even in Donnellanese*. This is because he doesn't have the relevant analogous pattern of background beliefs, and is unlikely to use "glory" in utterances expressing his beliefs about nice knock-down arguments in general. Donnellan's beliefs about the meaning of "glory" have not changed and are in line with standard English. Mrs Malaprop's beliefs about the meaning of "derangement" are not in line with standard English, but they are interlocked with many other beliefs she has about the world around her and how she uses words in general.

However, Davidson may respond that Donnellan has primed his audience enough to have a legitimate expectation that upon using "glory" with the intention of being interpreted in such a way. He thus has a disposition to use "glory" to mean a nice knock-down argument in *that particular context,* however ephemeral. Furthermore, an interpreter who knew he had such a disposition would understand his utterance.

This raises an interesting question. As we saw in chapter 1, *L&L* view the radical interpreter as modelling a speaker's dispositions to use words, and claimed that a speaker's dispositions to use words fixed those word's meanings. They also expressed concerns that, without prior knowledge of dispositions, interpretation would not be possible. Davidson, on the other hand, emphasises the role of a speaker's intention to be interpreted in a particular way (and being so interpreted). What, therefore, is the relation between one's intentions that one's words be

interpreted in a particular way and one's dispositions to use words? Do such intentions necessarily track dispositions to use words (or vice versa), or can intentions and dispositions be prized apart. If they can be prized apart, is it dispositions or intentions that determine an utterance's meaning? In the next section I investigate the possibility of a bifurcation between intention to be interpreted and disposition to use words and consider the implications for Davidson's model of communication.

3.3 Do a Speaker's Dispositions Fix the Meanings of his Words?

As we saw in the last chapter, one of *L&L's* main concerns with the practical possibility of linguistic communication without conventions or regularities appeared to be motivated by a combination of beliefs:

(1) The dispositions of a speaker fix the meanings of their words.

(2) An interpreter must have antecedent knowledge of a speaker's dispositions to use an expression in order to interpret it when the speaker so utters.

(3) Such prior knowledge can only be acquired through observation of past regularities in speech behaviour.

I suggested there that Davidson, in contrast appeared to deny (2) and thus would not be committed to (3). It is also not clear whether Davidson really endorses (1) given the outline of communication he presents in *Derangement* and related papers.

To say that someone has a disposition to φ is to say that she would φ if subjected to certain stimuli. The theory of meaning that the radical interpreter produces is a theory of a speaker's dispositions to use words, since the theory would generate many utterances or sentences that may never be uttered by a

particular speaker. We might suppose from this that a speaker's dispositions to use words in a particular context (relativised to time, environment, audience) will fix the meanings of those words when they are so uttered. Accordingly if we knew how someone was disposed to use a word (in a particular context) we would know what a particular utterance would mean when so uttered.

However, Davidson (e.g. *Derangement, SP*) claims rather that it is the speaker's intention to be interpreted in a certain way *and her being so interpreted* that determines the meaning of her utterance. Utterance meaning is therefore "constituted by the match between communicative intention and successful interpretation" (*SAL*: 121). But it is not clear that an intention to be interpreted in a particular way and one's dispositions to use words necessarily coincide.

Talmage (1994) suggests that, contrary to Davidson, the literal meaning and the first meaning of an utterance can be prized apart. She takes literal meaning to be what is understood at the point of successful communication, and first meaning as the meaning that one intends one's hearer to assign to an utterance. The literal meaning would be what the radical interpreter would assign to the utterances as generated by his theory, in virtue of what the speaker holds true in what circumstances. First meaning, may not coincide with this, since, for example, a speaker may have an intention that his utterance be *misinterpreted*. Talmage's example imagines a situation where someone who knows standard English utters "There's a nice derangement of epitaphs" while admiring a nice arrangement of epithets:

> [S]uppose that, unlike Mrs Malaprop, I intend my hearer (whom I believe to be less adept than most in noticing similarities in sound) to assign to the utterance ["There's a nice derangement of epitaphs"] its conventional meaning, that I intend him, that is, to interpret it as being true if and only if this is a nice

> derangement of epitaphs. Given this, it is clear that the literal meaning that would be arrived at through radical interpretation does *not* coincide with the literal meaning I intend my hearer to assign to my utterance. (1994: 219-20) [6]

In such a circumstance, Talmage asks, which of the two meanings should count as the literal meaning (the meaning that is grasped in the course a successful communicative exchange)? Talmage answers that it must be the meaning generated by radical interpretation, not, on this occasion, that intended by the speaker. This is because radical interpretation captures a speaker's linguistic standards – it reveals his beliefs about the meanings of the words in his own language, and (principle of Charity applied) the utterance's truth conditions. Moreover, although Talmage is not explicit in saying that first meaning is a species of speaker meaning, in claiming that she "assume[s] that literal meaning just is that which is grasped in *linguistic understanding*," (221) she strongly implies that the first meaning (since it is distinct from the meaning that is grasped linguistically) is speaker meaning, and, moreover, can be interpreted non-linguistically.

Talmage concludes that given that first meaning and literal meaning can diverge in the way that she suggested, Davidson ought to abandon the claim that the speaker's intention to be interpreted in a particular way (and subsequently being so interpreted) determines the literal meaning of the utterance, and revert to his claim that the meaning of a particular utterance, as spoken by a speaker, be determined by the speaker's linguistic standards, captured by radical interpretation.

---

[6] Kotatko (1998: 235), draws on similar of examples of intentions to be misinterpreted in his attack on Davidson's notion of utterance meaning: "Imagine a colonel who gives an order which he wants to be misinterpreted, in order to gain an opportunity to punish a soldier; or imagine somebody who warns somebody else that it would be dangerous to do p, since it is his duty, but wants to be misinterpreted because he would like his audience precisely to do p, etc. In all such cases, the speakers count on a notion of utterance meaning as something independent of how the utterance is intended to be interpreted and of how it is interpreted."

Gustafsson (unpublished: 9-10) asks us to imagine an English man, Jack, with no knowledge of Swedish being asked for a ticket on a Swedish train. A man dressed in attire indicating a conductor comes up to Jack and utters “Kan jag få se på biljetten tack”. Jack, from the contextually available clues, correctly supposes that the man wants to see his ticket and so offers it to him. For that matter, Gustafsson points out, it wouldn’t have mattered if the conductor had said “Tycker du om köttbullar?” (“Do you like meatballs”). It would be quite plausible to say, Gustafsson suggests, that in neither case is the communication that has taken place *genuinely linguistic*. The conductor could even have conveyed the message by simply staring at Jack. Jack does not interpret the utterance, but interprets what the speaker intended him to understand by it.

However, it would not follow from the *communication* not being genuinely linguistic that the *utterance* has no meaning: as far as what a radical interpreter could model of the conductor, since he is disposed to mean “Can I see your ticket please”, by “Kan jag få se på biljetten tack”, then that would be the literal meaning of his words. Jack, on the other hand, having only contextual clues to go on, successfully interprets the utterance as intended, but not from any knowledge of the speaker’s dispositions to use words in a particular way.

I reserve evaluation of Talmage’s conclusions here; however I will reiterate some key points that seem relevant to my current discussion. Let us assume:

(1) A complete theory of meaning for a speaker constructed through radical interpretation models a speaker’s dispositions to use words.

(2) A complete theory of meaning is not a sufficient condition for capturing a speaker’s intention to be interpreted in a particular way on a particular occasion (from Talmage)

(2) would follow from (1) with the additional assumption that a speaker's intentions to be interpreted in a certain way do not track her dispositions to use words.[7] Let us also assume:

(3) utterance meaning is "constituted by the match between communicative intention and successful interpretation" (*SAL*: 121).

(2) and (3) together would seem to lead to the rather odd conclusion that a complete theory of meaning for a speaker is not sufficient to interpret a speaker's utterances, which is contrary to what Davidson would himself claim (*RII*: 121).

One way to save radical interpretation from the claim that a theory of meaning derived through it is not sufficient to interpret a speaker's utterance is to suggest that what is being interpreted when first meaning (as opposed to literal meaning) is grasped is not the meaning of the *utterance* at all, but rather something about the speaker's intentions to be interpreted – a variety of *speaker meaning*.

This preserves *autonomy* of meaning: that is to say, that it preserves the distinction Davidson wishes to make between what an utterance means and the non-linguistic purpose that it is put to. The same utterance, he remarks, once understood can be used to serve almost any extra-linguistic purpose (*C&C*: 274-5; *T&T*: 164-5). Moreover, it seems consistent with Davidson's belief that "What a fully informed Interpreter could know about what a speaker means is all there is to learn" (*CToT*: 315).

[7] If a speaker's dispositions to use words are determined by what sentences a speaker holds true.

Another way of approaching this problem is through making a distinction between our discussions of what constitutes utterance meaning, and talk of what it is an interpreter interprets.

3.4 The Perspective of the Interpreter

If an interpreter, whose prior theory cannot accommodate Mrs Malaprop's utterance, yet, through exercise of his creativity and imagination, grasps what Mrs Malaprop intends to communicate, what does the interpreter interpret first – the utterance meaning or the speaker meaning? This question is all the more apparent when the speaker's intention for an utterance to be interpreted in a particular way is a one-off. That is to say that, while we might say that "derangement" means *arrangement* in Malapropese because a radical interpreter could ultimately over time find evidence to correlate her use of "derangement" with "arrangement" in his own idiolect – we, upon hearing Mrs Malaprop and grasping what she intends to convey do not engage in any such elaborate process. So what is it we grasp? Davidson himself acknowledges that "it often happens that we can descry the literal meaning of a word or phrase by first appreciating what the speaker was getting at" (*Derangement*: 92); we can, then, interpret an utterance by first grasping what the speaker intended to convey.

Consider Donnellan, who has a false belief that there is a tall man in a given environment, and Mr. Malaprop, in whose idiolect "man" means *rock*. If both were to utter "the man is tall" in the presence of a salient rock we might infer that Mr. Malaprop's utterance meant that the rock is tall in Malapropese, but that Donnellan's utterance did not. However, from the perspective of someone interpreting both for the first time it would seem that the interpreter could surmise that both *speakers* were

referring to a rock and intending to say of it that it was tall, but if the interpreter inferred from this that "man" meant *rock* in *both* of the speaker's idiolects he would, of course, be wrong. Alternatively, if the speaker assumed both were speaking "standard English" and concluded that both had mistaken beliefs about their environment, he would also be mistaken. At the time, without further information, the interpreter cannot make any inference that would enable him to distinguish between the two men, and it would seem to follow that the interpreter is not warranted in assigning a meaning to the utterance that results in the attribution of false beliefs to both men over 'non-standard' language or vice versa. Nevertheless, both speakers *get away with it;* in both cases the interpreter somehow "works out" that both men are asserting of the rock that it is tall. But this is by correctly determining their speaker meaning.[8]

Intuitively, when encountering a malapropic utterance, we are apt to believe that there is a speaker meaning/utterance meaning mismatch: the speaker *meant* something other than what they said. If we can divine what the speaker intended to communicate; we would probably consider ourselves to have grasped the speaker's meaning *in spite of* the sounds uttered not *via* those sounds. Even if we believe this as a result of a mistaken non-Davidsonian folk conception of language, if *we* - as actually engaged in the interpretation - do not consciously assign what the speaker

[8]In both cases, when the interpreter is faced with a seemingly anomalous communicative situation he does not seek to interpret what the speaker's words mean (he does not think to himself: "Am I to assume that Mr. M. consistently uses "man" to mean *rock*? Or does his use of "man" result from a background nexus of beliefs that are akin to the nexus of beliefs that I have that underpins my use of rock, implying that he is mistaken about what he sees..?). Rather the interpreter tries to work out, in both cases, what *the speakers mean* by their words, regardless of whether that is consistent with what those words mean in their respective language – and it seems it is by correctly interpreting what the speakers mean that he determines what they wish to talk about in their language.

intended to communicate (and successfully did) to her *utterance* then we don't accommodate that information in our "theory" of her language (perhaps she was tipsy and couldn't get her words out); it would seem then that we did not understand her in virtue of grasping any linguistic meaning of her utterance.

My own account of the divergence between what we can categorise as the literal meaning of an utterance and what the interpreter successfully interprets fits in well with this move.

(1) In the intended misinterpretation case, let us suppose with Talmage that what is being interpreted is not the literal meaning of the utterance but a form of SM. It just so happens that in the intended misinterpretation case the literal meaning of the utterance doesn't coincide with SM.

(2) In an honest case, the speaker intends the hearer to interpret the words as the speaker is disposed to use them (the literal meaning). If that utterance is not in line with the hearer's prior theory it is still the speaker meaning that is interpreted, but in this case it does coincide with the literal meaning of the utterance (even though this is not the conventional meaning).

(3) This does not mean that a hearer cannot ultimately legitimately assign a meaning to the speaker's *utterance*, but this can only be done after at least some evidence of a speaker's relevant dispositions to use words has been documented.

We could say that the speaker's intended interpretation of the utterance (1) is not the literal meaning of the utterance because once a radical interpreter has had enough access to the speaker he will be able to dismiss the deception case – much in the same way that he would dismiss a subject's mistaken belief while maximising

consistency over the total range of utterances and beliefs. The theorem tentatively derived from the “intended misinterpretation” exchange would be dispensed with as inconsistent with the theory as a whole. However, the theorem derived from (2) would be preserved and the theory would model the speaker’s idiolect in line with this.

But both theorems are initially arrived at *and understood as intended* via a correct assignment of *speaker meaning*. Regardless of whether one endorses the claim that “A nice derangement of epitaphs” has that meaning in Mrs Malaprop’s language – it is not the utterance that initially is interpreted but the speaker meaning.

Talmage’s analysis poses problems for Davidson. The key point is that one’s intentions to have one’s utterance interpreted in a particular way and one’s dispositions to use an utterance can differ. If so, then Davidson is faced with a dilemma: if it is one’s intentions to be interpreted, and one’s being so interpreted, that determine the meanings of one’s utterances, then radical interpretation, which models a speaker’s dispositions to use words, does not seem sufficient to the task of interpreting a speaker. If we wish to preserve the sufficiency of radical interpretation for interpretation, then it would seem that intentions to be interpreted and so being interpreted do not determine utterance meaning.

3.5 Conclusion

In this chapter, I have argued that Davidson’s claim that Mrs Malaprop’s *words* “A nice derangement of epitaphs” means *A nice arrangement of epithets* is consistent with his earlier claims. (More accurately “A nice derangement of epitaphs” in Malapropese means the same thing that “A nice arrangement of epithets” means in my idiolect).

I have also argued that we can separate considerations of what we could say a word means in a highly circumscribed particular idiolect (relativised to time, context, audience etc.) from considerations to do with the process of interpretation. The interpreter, on encountering a single anomalous use (a use that is contrary to the expectations generated by his prior theory), is not in a position to decide whether the word has a non-standard meaning in the speaker's language, the speaker is giving the word a nonce use, or even if the speaker is expressing a false belief. Interpreting what a speaker's *words* mean can only happen in the light of the further "field work" that would be required to produce a meaning theory for that speaker. At the level of the single encounter, all the interpreter can do is to try to work out *what the speaker intends to convey* on making such an utterance. Here the distinction between speaker meaning and word meaning returns and it would seem, as in the case of Jack and the Swedish conductor, the interpreter must focus on speaker meaning in forming an opinion on what proposition the speaker means to express. Moreover, the *communication* in question here is not obviously genuinely linguistic, although the utterance's meaning may have been fixed by a speaker's intentions to be interpreted in a particular way.

An alternative way of stating the point may be as follows: suppose I endorse a "theory" that states that "all swans are white". I encounter what turns out to be a black swan. I have to recognise the object encountered *as a black swan* in order to alter the theory so that it accommodates the new evidence. Some form of extra-theoretical grasping of the datum would seem to be in order *prior* to adjusting the theory. I don't recognise the black swan in virtue of adjusting the theory – how can I? That theory tells me that all swans are white. We might say that something analogous goes for

the interpretations of anomalies. One must in some sense "grasp" an utterance *before* one accommodates it in one's theory (if it cannot be already accommodated). However, if an *utterance* meaning can only be interpreted once it is located within a complete theory of a speaker, then it would seem that it is not the *utterance meaning* that is at first grasped.

Nothing said here commits one to claiming that all cases of successful linguistic understanding are merely cases of interpretation of *speaker meaning*. There are genuine cases of linguistic understanding, but my tentative suggestion here is that genuine cases of linguistic understanding occur when an utterance fits in with the prior shared practices of interpreter and interpretee. Perhaps, however, in these cases, as Dummett suggests, what is going on cannot be said to be genuine case of "interpretation":

> Similarly there is a way of understanding a sentence or an utterance that does not consist in putting an interpretation on it […] one who grasps a sentence need not be able to *say* how he understands it. […] When the hearer does not have to search for the speaker's meaning, but takes for granted that he is using a word in just the way in which he is familiar, there is […] no process of interpretation going on (1986: 454).

Smith (2006b: 950, n. 18) also notes that our understanding of utterances in languages with which we are familiar is an automatic process more like perception than interpretation. In "standard exchanges" it is quite possible that one grasps the words directly or, in Smith's words "perceives" their meaning (951); one then goes on to largely read off the dispositions of the utterer from the utterance. It is only in the aberrant situations that *interpretation* is required; in those cases, since on first pass one finds one cannot "grasp" the meaning of the words in question, one turns to trying to divine the intentions of the speaker and then maps the results of that back on to the words (perhaps after more evidence has been accrued). But in a single

communicative exchange containing an innovative utterance, it is not the necessarily the meaning of an utterance that one is identifying, but what a speaker means to communicate when he produces an utterance.

Chapter 4: Triangulation and Interpretation

4.0 Introduction

In one sense, we frequently encounter words, phrases and utterances that we have not met before and that our "theories" (of standard English or of an individual) cannot accommodate. Davidson suggests that in such cases we adjust our theories to accommodate those innovative expressions when communication is successful. However, what is the significant difference between a mature speaker coming to know that, in Malapropese, "A nice derangement of epitaphs" means *A nice arrangement of epithets*, and a linguistically immature child coming to know (learning) that, in standard English, "The cat sat on the mat" means *The cat sat on the mat*? We might suppose that similar mechanisms are in operation. Davidson, however, initially appears to avoid the comparison in Derangement.

> The contrast between acquiring a new concept or meaning along with a new word and merely acquiring a new word for an old concept would be salient if I were interested in the infinitely difficult problem of how a first language is learned. By comparison, my problem is simple. I want to know how people who already have a language (whatever that means) manage to apply their skill or knowledge to actual cases of interpretation. All the things I assume an interpreter knows can or do depend on him having a mature set of concepts and being at home with the business of linguistic communication. (*Derangement*: 100)

For Davidson learning and interpretation of innovations appear to differ. Presumably this is because, unlike the mature user (who's theory of the speaker maps the target utterance onto expressions in his own language), the child does not yet have the organised categories of concepts of a native idiolect on to which she can to map the expressions encountered.

But this response seems to be question-begging: only by assuming that the interpreter *has a theory* of his own language on to which he can map the alien expression, does it follow that a child (who has no theory, or only a very partial one) must do something else when he learns a native tongue. Davidson offers no reason to suppose that, in actual language practice, the skills involved in the interpretation of a non-standard utterance are not parasitic upon, or a subset of, the skills involved in learning a language in general.

An infant's learning the meaning of a new expression *prima facie* does not require the stringent constraints to which the radical interpreter is subject. The radical interpreter must ultimately apply his theory to a language wholesale to confirm it is interpretive. Immature language users by definition can be said to have learnt the meaning of an expression without having a "complete theory" of a language: their understanding of a language is, at best, partial.[1]

How first languages are learnt is no doubt an "infinitely difficult problem". However without attending to this problem, how can one rule out the possibility that learning new words for already-known concepts is not merely a branch of learning, using the same skills and following the same procedures as a child?

In other works, Davidson takes a more sympathetic approach to the relation between language learning and interpretation by a competent language user. In this chapter I investigate Davidson's theory of triangulation, which he presents as, among other things, a partial account of native language acquisition, and consider whether

[1] In contrast to Fodor and Lepore (1992), who suppose it is possible to have a concept in isolation, Davidson claims that to have a concept depends on, among other things, a network of related beliefs. It is therefore questionable whether very young infants can be said to have concepts or beliefs. (*HIT*: 32-3). Nevertheless, to have a network of other beliefs is consistent with having a partial understanding of a language.

triangulation sheds any light on the mechanisms involved in the interpretation of malaprops.

I consider two projects for triangulation: the discussion of the ambitious project examines how triangulation might be directly brought to bear on the interpretation of malaprops. So, for example, Smith (2006a: 404) claims that triangulation takes place both in language learning and in aiding interpretation when communication breaks down. I suggest, however, that this project does not succeed, since triangulation proper appears at best to only directly bear on the interpretation of occasion sentences, and any modification of it (such as Davidson's account of the role of our understanding of "tradition" in reading literature) is vague and, moreover, leaves us in a hermeneutic regress.

The discussion of the modest project examines how triangulation might be indirectly brought to bear on our understanding of how malaprop interpretation is possible through its role in the initial constitution of the linguistic subject. I suggest that this project is more promising (although in need of refinements beyond the scope of this thesis); however, it provides a route back for appealing to regularity of use in explaining communication.

4.1 Triangulation: Preliminary Remarks

Triangulation accounts for how the meaning of one's thoughts and utterances are fixed via interpersonal communication. It appears in Davidson's later works and has far reaching consequences for his accounts of metaphysics, epistemology and communication. Importantly, Davidson uses the notion of triangulation to support various claims throughout his later philosophy: 1) that triangulation is key to thought about an objective reality: it is necessary for the emergence of propositional thought;

2) it is a necessary condition for determining the accounts of perceptual beliefs; 3) it also is pressed into service in accounting for the ostensive learning of a language; 4) it is advanced as an explanation of how radical interpretation is possible (as a justification of the principle of Charity) (Pagin, 2001).[2]

Triangulation supposes that a triangular relation exists between two rational beings and objects in the world. Davidson explains by analogy. In this case the analogy is used in explaining why interpersonal communication gives rise to the concept of objective truth:

> If I were bolted to the earth, I would have no way of determining the distance from me of many objects. I would only know that they were on some line drawn from me toward them. I might interact successfully with objects, but I could have no way of giving content to the question of where they were. Not being bolted down, I am free to triangulate. Our sense of objectivity is the consequence of another sort of triangulation, one that requires two creatures. Each interacts with the object, but what gives each the concept of the way things are objectively is the base line formed by the creatures by language. The fact that they share a concept of truth also makes sense of that claim that they have beliefs, that they are able to assign objects a place in the public world. (*RA*: 105)

The need for interpersonal communication to give objective content to words and utterances is explained in similar fashion, by accounting for how an immature language learner endows sounds with a particular meaning (*Externalisms*: 14). Imagine a child uttering the sound "table" in the presence of a table and being rewarded for it. Why is it that the child can be said to be responding to a table, an object in the world, as opposed to a proximal stimulus, as those on her nerve-endings?

> Just as […] the child responds in similar ways to similar stimuli, so do we […] we find the child's mouthings of "table" similar, and the natural (to us) class we find in the world that accompanies those mouthings is a class of tables. […]The child finds tables similar; we find tables similar; and we find the child's responses in the presence of tables similar. It now makes sense for us to call the responses of the child responses to tables. […] Where the lines from child to table and us

[2] Pagin (2001), however, notes that Davidson does not distinguish these tasks clearly, and that he is not clear in how these tasks relate to each other.

> converge "the" stimulus is located. It is the common cause of our response and the child's response. (*SP*: 262-3)

This triangulation is essential to communication: it is necessary for words and utterances to be invested with meaning.

> Without this sharing of reactions to common stimuli, thought and speech would have no particular content – that is, no content at all. It takes two points of view to give a location to a cause of a thought, and thus to define its content. (*TVK*: 212-213).

As I have already stated, Davidson pressed his idea of triangulation into service of a wide range of claims. In most discussions of triangulation, Davidson is not considering how to get interpretation "off the ground", or how to interpret when communication breakdown threatens. Triangulation is employed in explaining our essential relation to each other and a common world as a necessary condition of communication, objectivity, truth, the possibility of error and so on. Furthermore, rather than a supplement or alternative to the principle of Charity, it constitutes an attempt to justify it (*CToT*: 151-4; *SCoT*: 245). However, triangulation (perhaps a rough and ready version) is also used to explain our investing words with common meaning – something which happens during the course of verbal exchange, rather than prior to it. Thus it is worth considering how triangulation might bridge the gap between the knowledge available to the interpreter prior to a verbal exchange, and his ability to interpret an innovative sentence.

4.2 The Ambitious Project

4.2.1 Triangulation and Getting Interpretation "Off the Ground"

Triangulation, as interests us currently, is described by Brinck (2004: 182) as constituting "another way of breaking into the intentional circle than by using the

principle of Charity", consisting in "finding non-semantic evidence for what words mean" and depending "on knowledge concerning how perception yields the contents of belief". It relies on the following conditions:

1) The speaker makes herself interpretable.
2) The hearer understands that the speaker intends the hearer to interpret.
3) The speaker must knowingly and intentionally respond to a specific stimulus, described by Davidson as 'an aspect of a cause' (and so must have the concept of the stimulus in question).
4) This 'cause' must be common to the speaker and hearer in the sense that, in virtue of their inherent similarities, they respond in a similar way to it.[3]

In a primitive situation, the three-way relation between speaker, hearer and object can be illustrated quite simply. One person, S, responds in a similar way to a class of objects, the second person, H, notes the similarity of the members of the class that the one person is responding to, and the similarity of the responses. H then ascribes a concept (or possession of a concept) to S on the basis of correlating S's similarity responses to her own.

The basic case, already mentioned, of the infant learning a new word seems illustrative enough of a primitive case of triangulation. Through a process of repetition, with a teacher who "rewards" correct responses to objects that are relevantly similar, we can thus tell a plausible story about how individual concepts are

[3] Davidson (*TVK*: 212) suggests that facts about evolution and the learning process together with the principle of correspondence will ensure that these similarities inhere in the interlocutors.

acquired in a creature and how that creature invests a sound with a particular meaning.[4]

Whereas with PoC, the interpreter is limited to the assumption that the speaker's beliefs are by and large true, triangulation allows the interpreter to assume that the speaker is likely to respond to a particular object in a common environment in similar ways to the interpreter, enabling the interpreter to (relatively firmly) attribute a belief to a speaker and a meaning to his utterance before the construction of a complete theory. Thus, for example, as Davidson suggests:

> Suppose I am trying to interpret a speaker I do not understand. I notice that she is caused to utter the words "Sta korg" when and only when a red object is in plain sight, well illuminated, and she is facing it. If further experiments bear this out, I tentatively translate her sentence "That is red". It is obvious that this is just the sort of evidence on which we depend for learning a language […]. (*SCoT*: 245)

It is difficult, however, to see how triangulation is to be applied in practice. First, how can an interpreter individuate content of sentences or sub-sentential parts without pre-ascribing intentional states to the speaker? Secondly, even if Triangulation can explain the individuation of content in the case of occasion sentences, it is not clear that it provides sufficient resources for the interpretation of non-occasion malaprop sentences. I will consider these concerns in the next sections.

[4] It may seem that my talk of concepts here is rather un-Davidsonian. However, Davidson's discussion of triangulation is littered with concept-talk: (e.g. *RA*: 98; *SP*: 108,119,121; *EE*: 196; *EoT*123-4: "A creature does not have a concept of cat because it distinguishes cats from other things, it has a concept of cat because it can make sense of misapplying the concept […] unless you have a lot of beliefs about what a cat is, you don't have the concept of a cat". By using concept talk, I do not thereby imply that meanings of utterances or their sub-sentential parts can be reduced to concepts. I use concept in only in the sense that Davidson does here.

4.2.2 Individuation of Content

The interpreter attributes a concept to a speaker on the basis of grouping together the speaker's similarity responses. Brinck, however, suggests that this cannot be mere behavioural ascription, but rather the responses are "intentional actions that are performed in response to a particular feature of the context". In order to be aware that one's partner is conceiving of the stimulus in a similar fashion to oneself, one must ascribe beliefs and intentions to her, which cannot be divined from mere observation of behavioural data: "the same gesture, say, a repeated waving of a hand, may in one context be a way of saying "hello" to somebody else, while in another context the gesture simply serves to cool oneself. The gesture has the same structure, but different functions in the two contexts" (2004: 190).

Repeated observations of similarity responses will not serve to separate the wheat from the chaff either, since the interpreter of the mere behavioural evidence will not be able to discern that the responses are of a similar kind of behaviour to hold them constant in the first place. We have, to all intents and purposes, a chicken and egg situation: "concept attribution depends on determining common causes. However, isolating and keeping constant the triggering feature across the contexts of utterance precedes the conceptual identification of a common cause. This identification is only achieved in the long run". However, triangulation, Brinck points out, presupposes which feature of an environment a speaker reacts to in advance of interpretation (2004: 191).

Even if we are to assume, in general, that our similarity responses (as humans) tend to be in accordance, and so we are, by and large, justified in our judgements of others' similarity responses, one cannot guarantee this to be the case

in the context of an individual utterance. And, as Brinck notes, “the individual situations are necessary to get radical interpretation off the ground” (2004: 191).[5]

Even Davidson’s (*HIT*: 43) suggestion that environmental saliency provides the interpreter with grounds for the interpreter to ascribe similarity responses does not remove the indeterminacy that arises from triangulation. Often what is salient, Brinck (2004: 192) suggests, is what goes against expectations (what we find surprising in a situation), but individuals may have differing expectations. The interpreter must again appeal to the intentional states of a speaker in order to gauge what is salient or relevant in the context of a particular utterance – something that the interpreter is not permitted to do at the initial stage of interpretation.

Brinck (2004: 195), however, thinks that this dilemma can be solved by appeal to what he calls “joint attention”, which characterises the ability of speaker and hearer to focus their perception simultaneously “as a consequence to attending to each other, on a shared object of attention”. Brinck sees the advantages of joint attention as not requiring the participants as having “intentional states about intentional states”, because “the attitudes that states of attention express are not directed at sentences or other linguistic entities” (203).

Brinck’s necessary and sufficient conditions for joint attention are as follows:

(i) S and H attend to each other’s states of attention as states that are directed at objects in the environment and by being so directed are

[5] It may be suggested that the radical interpreter could tentatively hypothesise intentions to ascribe to his subject which can later be confirmed by further investigation. This will not do: to at all be able to make such repeated observations, the interpreter must hold something constant across contexts of observation. “There must be some way of telling that the observations are of a similar kind of behaviour”. (Brinck 2004:191)

about these objects, and, moreover, attend to these states also when they are directed at non-salient entities.

(ii) S and H make attention contact (attend to each other usually visually).

(iii) S and H alternate gaze between each other and the object.

Joint attention requires that S and H attend to each other in such a way as to grasp that these attentional states are directed at objects and, as such, are about them. However S and H attend only indirectly to the *intention* behind the observed behaviour by grasping that the behaviour of the other is intentional, directed at an object and purposeful (not what those intentions actually are). This does not require that they ascribe particular beliefs to the other (at this stage, at any rate). Since it is H's recognition of S's *attentional* rather than *intentional* state that motivates him to attend to the object that the speaker herself attends to, joint attention constitutes a way of breaking into the intentional circle.

Brinck claims that joint attention requires more than mutual perceptual accessibility. Conditions (ii) and (iii), he suggests, explain how S and H can perceptually identify and subsequently focus in on a common cause, through their interaction with each other. Joint attention allows speakers to coordinate perception through guiding each other via processes such as directing gaze. Communication at this level at least requires coordination and cooperation between speaker and hearer.[6]

[6] This suggestion that it is the speaker's intention to produce an attentional state in the hearer contrasts with a Gricean analysis of non-natural meaning whereby the "underlying intention of the utterance is to produce a certain intentional state in the hearer" (Brinck, 2004: 197).

The account advanced by Brinck offers one way of getting interpretation off the ground by individuating content at the initial stage of radical interpretation without ascribing intentions to the speaker. It also seems somewhat plausible in the context of infant language acquisition. If we consider the infant learning the word table as described above, it, of course, would beg the question to suppose that the infant leaned to apply the word “table” via a process which involved ascribing intentions to the teacher, for without being already in possession of a language, it would seem, on Davidson’s view, the infant would not be in a position to do this.

Brinck’s discussion gives an important account of the deficiencies of triangulation. If triangulation, as Davidson presents it, is brought in to take up the slack for radical interpretation, it fails. However, the details of his solution are less important to this discussion: he is attempting to save radical interpretation by showing how it can get off the ground through the process of joint attention. I have been suggesting that, in actual cases where communication breakdown threatens, the perspective of the radical interpreter offers little help.

Rather in such cases there is no reason to suppose that H does not combine observation of environment with the assumption that he groups sensory information as the speaker does, and other assumptions suggested by empathy, shared cultural background, history with S and so on. Moreover, in chapter 3, I suggested that assigning the relevant linguistic intentions to the speaker prior to interpretation of an utterance, while not permitted to the radical interpreter, seems to be possible in actual practice, and does happen. The interest lies not in *whether* we can do this – we can – but *what* enables us to do it. I shall return to these points below.

4.2.3. Triangulating Non-occasion Sentences

The discussion above suggests that triangulation, perhaps suitably modified, may suffice for interpretation at the level of simple occasion sentences. However, can triangulation help in cases where communication breakdown threatens? Consider two speakers threatened with communication breakdown at some point in the conversation: S produces some utterance “S1” that H’s prior theory of S does not accommodate. Let us say that in response to “S1”, H utters a request for clarification, “S2”, that S’s prior theory cannot accommodate. Both look a little puzzled. How can the situation be resolved? Let us suppose that S’s utterance concerned something in their shared environment: they are at a farm and S’s utterance could be translated into H’s language by “These cows have black and white coats”. Communicating this via triangulation seems more or less achievable, using joint attention and ostension (including repeatedly pointing to cows, black things, white things, uttering the words one is accustomed to use when talking about them etc.).

With an utterance of the sort provided by Mrs Malaprop, it does not seem possible that triangulation of this sort will easily yield a correct response: one cannot direct another’s attention to an epithet in the environment in order to have them assign one’s intended meaning to “epitaph”. Even if we were to grant that, over a period of time, it is *conceivable* that an ideally placed radical interpreter, who engaged with triangulation with Mrs Malaprop, could, from starting with occasion sentences, eventually construct a truth theory that would enable him to correctly interpret her, this would be of no help – both for the reasons specified in chapter 2 (Mrs Malaprop can’t form a legitimate expectation on the basis that her audience is a

radical interpreter) and, moreover because it does not explain how we do, *in fact*, manage to interpret her.

Triangulation as a method of aiding interpretation when communication breaks down seems limited to occasion sentences where the relevant common causes are present in the environment. When it comes to non-occasion and more complex sentences it is difficult to see what role triangulation plays: the difficulties become even more acute when considering the interpretation of written productions.[7] We are still in need of some additional resource that the interpreter can avail himself of in order to aid interpretation where passing theories and prior theories diverge.

4.2.4 A Supplementary Account of Triangulation

One possible account of triangulation that may be of use in malaprop interpretation would be one that specified the relevant relation as that between hearer, speaker and something akin to mutual knowledge acquired from experiences of previous exchanges. Davidson himself gives a description of a form of triangulation involved in reading and writing: at the apexes of the triangle in this exchange stand reader, author and what he calls *tradition.*

> In writing, the deitic and demonstrative machinery so readily available in speech cannot in the same way complete the triangle that relates writer, reader, and object or event or time. Ostension, which often serves to relate names and faces, or help to introduce a new colour or sound or category, has no immediate

[7] Picardi (1999: 180) remarks that "[a]gainst the background of Davidson's theory of communication, it becomes very puzzling how to account for the sort of understanding that we seem to enjoy of Plato's writings or of the *Iliad*… Faced with this sort of question, philosophers normally evoke the continuity of our cultural (western) tradition: there is nothing mysterious in our ability to understand the *Iliad* or *Symposium*, for we are the heirs of an ongoing cultural tradition. However, the mention of traditions, shared practices and conventions does not chime in with Davidson's approach. The answer which is available to him as regards questions concerning what makes the remote and foreign intelligible will have to resort to constitutive principles of rationality on the one hand and to similarity of responses to outer stimuli on the other."

> analogue in writing. […] Literature itself provides an important part of the background an author is apt to assume she shares with her audience. Other books help constitute the world which completes the triangle of author and reader, just as prior conversations provide much of what speaker and hearer depend on for good communication.(*LLL*: 177-80)

This qualified formulation of triangulation is further spelled out in a discussion of the way in which Joyce makes himself interpretable to his readers, through drawing "on every resource his readers command (or that he hopes they command), every linguistic resource, knowledge of history, geography, past writers and styles" (*JJHD*: 147).

In a "live" conversation, rather than a hearer (or reader) attending to a mute text we might legitimately replace 'tradition' with mutual knowledge of shared cultural heritage, experiences acquired through prior exchange etc.. Davidson would allow these experiences and shared backgrounds (perhaps supplemented with mutual awareness of objects in the environment) to complete the triangle in the case of general communication. If we add to this the interactive nature of communication – the requirement for mutual cooperation - which is brought to the fore in the account of triangulating subjects, we might have the start of an account of how an interpreter arrives at a correct interpretation of a speaker when communication breakdown would otherwise threaten.

Even if such an account provided a solution, such cases of interpretation would not be radical because just as with Joyce (where one might use their knowledge of tradition to ascribe intentions to Joyce in order to interpret him), an interpreter uses their shared background to assign intentions and beliefs to a speaker in order to interpret him – this is something that the radical interpreter cannot do: he must break into the intentional circle, not begin from within it.

At any rate, appeal to "tradition" has its own problems. There is an important asymmetry between the speaker-hearer-distal stimuli and writer-reader-tradition triangles; namely, because of the way that the world is, and because of our innate or evolved shared interests in engaging in the world, we can expect S and H's similarity responses to the distal stimuli to be similar to each other – the relevant apex of this triangle, we can say, is fixed. However, the same cannot be said of the "tradition" apex of the second triangle. The problem here is one of what we might call a "hermeneutic regress".[8] One must interpret Joyce in the light of tradition, but tradition itself is composed of other written productions that must themselves be interpreted in the light of tradition. There is no guarantee that each author/reader, given their own circumstances and conditioning will interpret the written productions of tradition in the same way as Joyce: the problems become more acute a time goes on. To understand Joyce, because one sees he is channelling Trollop, and Trollop as channelling Shakespeare, would seem to imply that one has to understand Trollop and Joyce did, and Shakespeare as Trollop *and* Joyce's Trollop did and so on *ad infinitum*. It is not clear then that reader and writer can locate (the same?) tradition and have similarity responses to it that are relevantly similar to each other. This kind of objection is familiar to the reception theorist (although not necessarily *qua* objection); it might indeed suggest that there are intractable problems for the "successful" interpretation of written texts. And, of course, the written word is mute, its author dead, or powerless to mould his production to a particular interpreter - his role at his apex of the triangle would be substantially reduced. For these reasons

[8] The idea expressed by this is familiar to literary criticism of a continental bent (not unrelated to Gadamer's notion of the hermeneutic circle). However, I am (in contrast to much of the discussion in this area) not making a metaphysical (relativist) point, but merely a sceptical one.

one can't simply transplant one apex of triangulation from a distal stimulus onto something as undefined and unstable as "tradition".

In day-to-day communication, while speaker and hearer can interact to grasp each other's similarity responses, the idea of mutual knowledge, past experience of each other, cannot directly transplant onto the distal stimuli either. Davidson would have us believe, for evolutionary reasons, that nature is pretty much how we think it is (*HIT*: 38). But S and H's beliefs about the common ground they share could be wide of the mark, and it is difficult to see how, at least in a short space of time in which an interpretation is to take place, one could detect the relevant similarity responses to confirm that they are not. Moreover, these could not be detected directly. A similar hermeneutic regress to that suggested above would seem to follow. Retreating to PoC to justify the assumption that one's interpreter has the same background as oneself, would not seem to help either, for while we might assume it guarantees that speaker and hearer largely agree in their beliefs about the world this does not constrain interpretation enough; moreover, since triangulation is intended to justify the principle of Charity, we fall into circularity. We need some other means of justifying such triangulation.

In conclusion, the ambitious project of using triangulation to explain interpretation as directly involved with securing successful interpretation in the day-to-day business of communication when breakdown threatens does not look promising. At the stage of learning a language it does look more promising, however, perhaps subject to the refinements of, say, Brinck.

4.3. The Modest Project

Recall that Davidson (*Derangement*: 98) suggests that a passing theory is "derived by wit, luck, and wisdom from a private vocabulary and grammar, knowledge of the ways people get their point across, and rules of thumb for figuring out what deviations from the dictionary are most likely". We need a more specific account of what these skills or knacks are, or, at least some plausible account of how communicative subjects come to have them.

Although, as we have seen, Davidson attempts to bring triangulation to bear in explaining cases of complex and sophisticated interpretation (passing theory construction), triangulation may play a more indirect role in ensuring successful communication in the complex exchanges of mature language users: namely, in the constitution and conditioning of a communicative subject (what we might, *pace* Davidson, call "learning" a "language"). It may be that the story we can tell about how engagement in triangulation matures a pre-linguistic creature into communicative subject can also account for the skills required in a passing theory (Simpson, 2003: 17).[9]

Defending the essential publicity of a language (or it's essentially social nature), Davidson (*SP*: 115-6) suggests that one must not merely be interpretable, but actually *interpreted* in order to mean something by their utterance. Without an interpreter, there can be no substance to the claim that a speaker has gone wrong. The support Davidson offers for the essential publicity of a language, is that of triangulation: sketching out a primitive learning situation of the sort quoted in section

[9] Simpson (2003:17) hopes that an "account that fleshes out the triangulation thesis would reflect the account of a passing theory-maker, and the story of how someone becomes a communicative subject would amount to the story of how one becomes a passing theory-maker."

4.1, Davidson suggests that the innate similarity responses of speaker and hearer enable both to locate the relevant stimuli as distal, rather than, say, on the surface of the skin. Creatures must interact, then, for their words to mean something; moreover, they must be aware of their interactions (*SP*: 120-1). Triangulation, then, shows us that the stimulus is objective, and allows us to attribute thoughts to a creature; in correlating its own similarity responses with the responses of another in the presence of a shared stimulus, there is a basis for identifying error (when correlation breaks down). To do this the creatures must be in communication; thus communication and triangulation are interdependent.

However, Simpson (2003: 13) objects that the "invocation of the triangulation thesis does not provide the required non-linguistic basis for communication because it presumes that communicative subjects appear on the scene fully fledged, for it offers no room for an account of their constitution as communicative subjects".

Simpson's solution is to accept shared practices as necessary but to show that sharing practices does not involve following rules. Our innate propensity to divide up the world as others do is to be taken as a pre-requisite for communication, but should not be deemed sufficient: what is needed, Simpson suggests, is a distinction between pre-linguistic responses to objects as "occurrent" (a range of, say, similarly-shaped objects) and "available" (objects that play a range of roles in a number of practices). Someone who has a concept must grasp the relevant object as *available*. This need not mean that we grasp a set of rules but "in order to work with a concept we require a skillful and flexible grasp of a shared practice". Moreover, we need an account of how such skills are acquired. Such an account would explain how "correction, direction and modeling would engender in a creature "a background of

habits and skills" and inculcate an "intentionality towards a world that is, we learn to presuppose, shared with similar others". It is plausible that a subject so trained will be equipped with the necessary non-linguistic communicative skills appropriate to theory formation (2003: 15-16). It is difficult, however, to assess these claims without having more detail than Simpson provides.

Davidson (*HIT*) attempts to account for the mismatch between the ease in which communicators do in fact manage to interpret each other, and the difficulties in theoretically accounting for those interpretive practices. A full answer to how we do manage to interpret each other would take "a lot of science", he tells us, although some aspects are accessible and philosophically interesting. In the relevant arguments of this paper, Davidson focuses on "conceptualization". The initial stage of conceptualization occurs only against a background of concepts "such as the contrast between true and false belief, the concept of an object, of causality, of a public space and time, and of other creatures similarly endowed with language" (*HIT*: 42). In order for a subject to be in full-blown possession of a concept, their beliefs about the object must interconnect with a number of beliefs (*HIT*: 32).

No doubt primitive triangulation with various isolated objects, together with a subject's interaction with those objects, provides much material for the development of concepts (Simpson, 2003: 16). There are many ways in which one could suggest that the move from mere classification of features of the environment to having a network of beliefs might take place. But does this help us with how people acquire the skills necessary for developing passing theories?

Perhaps: but only in very broad strokes. Davidson (*HIT*: 36) suggests that, since language and thought develop together, we learn about the objects in the world

as we learn how to talk about them; as we progress we learn the characteristic traits of objects that enable us to anticipate and predict behaviour. Much human understanding of their characteristic traits comes from interacting with them linguistically. But vague claims such as this stand in need of substantial empirical bolstering to be of much help; the move from a subject's relatively robust conceptualization to her being able to interpret malaprops stands in need of more explication.

However, let us suppose that, by utilising innate capacities of discrimination triangulation has a role to play in the development of a creature as a communicative subject. This development is in part constituted by the engendering of linguistic or communicative skills that enable the subject to interpret non-standard utterances through conditioning (being taught and modeled), just as they allow her to interpret complex and abstract words on the approach to linguistic maturity. Without such conditioning, a subject would neither have knowledge of a language (or be able to follow a practice of her own), nor the interpretive know-how to deal with innovations. This would afford regularities a substantive role in our account of linguistic competence. The learning, or conditioning, required in the development of communicative abilities, does not seem to be possible without regularities in the speech habits of others; it is only through the detection of the *similarity* of similarity responses that triangulation can occur. The teacher must repeat the same word, "table", for the infant to *invest* that sound with meaning, if the teacher does not, the teacher will fail.

Smith (2006a: 396), attempts to seek an account of language that pays due attention to the first-person conscious experience of meaning something by what one

says. He acknowledges that triangulation is a necessary condition of linguistic communication, but also suggests that it provides a starting point for an account of language that reconciles “what is immediately available to speakers in their inner acts of comprehension with what is outwardly accessible in public practice and observable speech”. Davidson’s view of an infant *endowing* a sound with meaning, rather than recognizing a meaning that is already present within a given sound is central to Smith’s account. A communicative subject develops, Smith suggests, not by coming to assume that others invest particular sounds with the meanings that they themselves do; rather, meaning is fixed together under conditions of joint attention. The meaning fixing process that a child undergoes when she learns a language provides an insight into how people understand what their own words mean: this can be traced back to the conditions under which a meaning was originally fixed for them (396-400).

Moreover, Smith claims that the first-person’s consciousness of the meanings of one’s words should be taken as basic in the understanding of another. A speaker knows what her own words mean, and takes the words of another to mean what she means by them. A speaker is entitled to this move, via the assumption that anyone who uses a given word, means what the speaker would by it. This assumption, however needs no further justification, it is automatic and the default assumption without which learning could not get off the ground. In other words, this is just what children must assume in order to learn a language in the first place (Smith, 2006a: 401-2).

This account of what it takes to become a communicating subject involves an account of regularities, or habits of speech: without these, the development of

communicative abilities cannot get off the ground. It is only through observing regularities that one can come to have the capacity to communicate.

This point is particularly interesting when considered in the light of Davidson's (*C&C*: 279; *SAL*: 114-5) remarks that we tend to speak in the same way as others as a matter of mere practical expedience. Viewed from the triangulation-based account of learning, Davidson's claim here does not ring true. This is because there seems to be, in virtue of our development into communicative subjects, more constraints on what we can mean by our utterances than merely the legitimacy of our expectations to "get away with it" when we do not speak as others do. Rather, conditioning through triangulation potentially restricts *how people can* intend an utterance to be interpreted.

In section 1.0.2 we saw the claim that one cannot intend to do what one believes to be impossible. What of Wittgenstein's challenge to us to say "It's cold in here" and mean *It's warm here*? (Wittgenstein, 1950: §510). Davidson could respond that we need not think that we cannot do this because it flouts some or other linguistic rule; rather, we find the point of Wittgenstein's challenge intuitively persuasive because we find it difficult to imagine ourselves in a context where we could utter "It's cold in here" and mean *It's warm in here*. There nevertheless may be such a context and audience where one would have a legitimate expectation that one could so utter: the challenge could, under the right circumstances, be met.

However, I also noted an equivocation on what constituted a "legitimate expectation". I suggested that there were two ways of understanding what counts as legitimate; legitimate in respect of other beliefs held to be true, and legitimate in the light of the facts. Perhaps conjoined with the PoC there would be little real distinction

between these, since, by and large, most of a speaker's beliefs are true and so other beliefs held in the light of them would largely accord with the facts. Nevertheless, it would seem that, where a belief about how one can be interpreted is not reasonable in light of the facts, this would prevent one's utterance being interpreted as one intended.

On the flip side, however, what one believes about how one can use an expression does constrain what one can mean from another, first-person, perspective. Regardless of whether something is, as a matter of fact, possible, if a subject does not believe it is possible, or perhaps if a subject has never entertained the possibility of it, then she could not intend it. In other words, if one believed that it was impossible under any circumstances that "It's cold in here" could mean *It's warm in here*, or if one had never entertained the thought that it was possible, one *could not* intend the utterance to be interpreted in this way, even this was *in fact* possible.[10] Subjects are likely to have this kind of belief, given how they come to "lean a language" in the first place.

4.4 Conclusions

In this chapter I have investigated what help triangulation might offer our account of how one can interpret the utterances of another. I first investigated the ambitious project of triangulation as providing an account of how understanding between two parties can be achieved when communication breaks down, and suggested that it had only limited application in this regard and so was not adequate to the task.

[10] Perhaps Burge's (1975) "naturalist" natives would find it impossible to do this: see section 1.0.3.

I then considered triangulation loosely construed as a relationship between speaker, hearer and some loosely specified common background knowledge (or "tradition"). This may provide a starting point for explaining successful interpretation of innovative sentences (and consequently such utterances bearing their intended meaning), but such interpretations would not be radical, since the interpretation of any innovative utterance may require the hearer to assign intentions the speaker prior to assigning content (even tentatively) to the utterance. The hearer attempts to determine what the words mean on the basis of the intentions he ascribes to the speaker, which he, in turn, infers from his prior knowledge of her and their common background knowledge. This kind of triangulation is, moreover, open to the objection of vagueness and lack of detail, and faces a potential hermeneutical regress.

I suggested, however, that the modest project, whereby triangulation constitutes an ingredient in an account of how a communicative subject develops might prove more fruitful, although there is much to be worked out in the detail.

However, should such an account work, a potentially unwelcome consequence threatens for Davidson. In section 0.3, I suggested that there was a symmetry in the conventionalist account of language and interpretation; namely, that conventions determine the meanings of expressions in a language, and thus to interpret those expressions one must have knowledge of the relevant conventions. This can be represented by the radical interpreter as the propositional knowledge required to interpret the target language.

On Davidson's alternative account of language, where conventions do not determine the meanings of utterances, I suggested, the same symmetry did not hold. Since it is (legitimate) intentions that determine the meaning of an utterance and

intentions are not directly publicly available, a different sort of account has to be given to explain how utterances are interpreted. The tentative suggestion I advance here is that we need to give an account of what it is about a linguistic subject that enables her to interpret utterances. Such an account cannot be of a sort that represents what it is to be a communicative subject in terms of propositional knowledge. It needs to explain how we have and come to have the linguistic skills and abilities to interpret when prior knowledge of meanings is to no avail. Triangulation may play a role in this account; however, to what extent is a matter for linguists.

However, viewed in this way, triangulation appears to create a substantial role for regularity (and possibly shared regularity) in an account of linguistic understanding.

Smith's emphasis on the first-person perspective suggested a way in which regularity and shared practices return to the account of linguistic communication. The subject invests sounds with meaning through triangulation with another, and in the initial stages of learning, "the meaning it has for the child will not be distinguished from the meaning it has for anyone else" (Smith, 2006a: 403). The child neither realizes that the sound might be used otherwise, or that other sounds might be used in the same way. Such a subject could not easily intend to speak other than in the same way as her teacher(s). As she develops further, she may have experiences that lead her to believe that there is more flexibility in "language" then she initially thought, and thus can form intentions accordingly. However, given the historical constitution of the linguistic subject, this is a far cry from suggesting that a subject could go as far as following a practice all of her own.

There are, then, grounds for thinking that regularities and shared practices have a *substantive* role to play in our understanding of communication. Significant parts of our prior theories are not “up to us”, on account of the conditioning that is required to form us into beings capable of linguistic communication in the first place. The very processes that provide us with the resources to be flexible in our interpretation of others, and to interpret innovative utterances, are the same processes that may by and large restrict what we can mean (because they restrict what we can intend to mean), by our words in general communication.

Chapter 5 Conclusions

5.1. Davidsonian Communication

In this thesis, I have been primarily concerned with how successful interpretation of utterances can be explained, given the assumption that conventions do not determine the meanings of words. If one accepts this assumption, then one cannot appeal to knowledge of conventions of a language to explain how, in essence, linguistic communication and understanding are possible. However, since intentions are not directly publicly available, one needs to account for some other way that an interpreter can determine the meanings of the utterances in question.

5.1.1 Radical Interpretation

With this in mind, I turned to an investigation of radical interpretation and its role in interpreting innovative utterances. Despite claims that Davidson's anticonventionalism is a natural offshoot of his earlier work on radical interpretation, I suggested that it is not clear how radical interpretation can explain how successful interpretation occurs in the case of innovations.

In chapter 2 of this thesis, I noted a certain inconsistency in Davidson's use of "radical interpretation". The claim that radical interpretation is a theoretical procedure which is not intended to explain how interpreters *actually* interpret, but rather to provide an account of what knowledge would be sufficient for interpretation appeared incongruent with other claims suggesting that linguistic communicators frequently do engage in radical interpretation. Focussing on actual interpretation, I suggested that the epistemic position of the radical interpreter was insufficient to interpret many

anomalies that we might come across in day-to-day instances of potential communication breakdown: in particular, the radical interpreter cannot appeal to what I called empathy, the assumption that one's interlocutor has the same desires, interests etc. as the interpreter. Moreover, as a point of principle, the constraints placed on what a speaker could mean by a given utterance are determined by their own legitimate expectations of what their audience could understand by the utterance in question. It seemed mistaken to suggest that a speaker's utterance could mean whatever a *radical interpreter* could determine it to mean (and so be determined by this), since a speaker's interlocutor at any given time may have far less or far more knowledge than we can afford the radical interpreter.

In chapter 3, I investigated the distinction between speaker meaning and utterance meaning. I argued that there are good Davidsonian grounds for suggesting that in most cases we can attribute a non-standard meaning to a particular *utterance*, rather than to the speaker of that utterance. However, there are also good grounds for supposing that an interpreter may not interpret the utterance itself but rather interpret the speaker's intentions to convey by so uttering. As such, although radical interpretation may capture a speaker's standing dispositions to use a word or expression in a particular way, this is not necessarily consistent with how they intend to be interpreted.

In short, since radical interpretation seems neither necessary nor sufficient to interpret anomalous utterances, I suggest that it sheds little light on how one can come to understand an innovative utterance, and thus how one can succeed in interpretation on Davidson's anticonventionalist picture of language.

Davidson (*SP*) appears to suggest as much himself. One view of language can be described as follows (*SP*: 107-9):

1) A "language" is "complex, abstract object, defined by giving a finite list of expressions (words), rules for constructing meaningful concatenations of expressions (sentences), and a semantic interpretation of the meaningful expressions based on the semantic features of individual words".
2) A "language" can exist independently of its exemplification.
3) Only by employing the notion of "language" and theoretical terms such as "word", "predicate" and so on, "can we give a systematic description of linguistic behaviour".

These claims about language are most consonant with what we might expect Davidson to say about the prior theory he attributes to an interpreter prior to an actual communicative exchange.

If we recall, the hearer's prior theory is systematic (claims (1) and (3)), it is prepared (known in advance of a particular exchange) but not shared (*Derangement*: 93). The hearer's prior theory could be modelled as a Tarski-style truth theory with the relevant qualifications and could be said to be of the speaker's "language". However, since it has been constructed on the basis of the hearer's past experience of the speaker, it is perhaps no more than justified conjecture that stands only until it is refuted – a useful model that serves only as long as its predictions bare out. As the predictions are born out, the speaker may gain confidence in his prior theory (*SP*: 111).[1] Moreover, the "language" that the theory is of, is itself not something the speaker possesses, but something that the hearer attributes to the speaker as an aid

[1] By "prediction", I do not mean to suggest that a hearer can predict what the speaker will say, rather that the theory he has prepares him to interpret a range of utterances.

to understanding future linguistic utterances. After all, much of the language that the hearer attributes to the speaker may never be exemplified (Claim 2).

For these reasons I think that it is safest to equate “language” (what a radical interpreter can represent) with what an interpreter ascribes to a speaker in virtue of his prior theory of that speaker, with the qualification that this sort of language is not shared in the way that languages are traditionally considered to be. It is the prior theory, formed by a hearer on the basis of his past experience of the speaker, that provides the hearer with his expectations of how the speaker is to ‘go on’, and it is this that the radical interpreter could model.

However, it would seem that the radical interpreter’s theory does not include the pre-linguistic skills that are also made use of in the day-to-day business of communication, in particular in developing passing theories.

5.1.2 Triangulation

I next turned to triangulation to see if it promises to fill in the gaps left by radical interpretation. I suggested here that an ambitious project, where triangulation was to be seen as *directly* involved in cases where communication threatens to break down could only have very limited application: in particular it seemed limited to occasion sentences at best. Any attempt to extend triangulation, to say, more complex or abstract linguistic interaction, and writing, faced potentially intractable problems. It did not therefore provide the requisite account of interpreting anomalous sentences: while arguably a necessary condition for interpretation, it was not sufficient.

Nevertheless, triangulation viewed more modestly, as constituting part of an account of the constitution of communicative subjects, looked to be more promising. I

suggested that although there is much scope for filling in the details, triangulation provided a good starting point for considering how those learning languages come to acquire the skills necessary to engage in mature linguistic exchanges.

However, triangulation viewed in this light reveals an important role for regularity and shared practices. Moreover, this role is more significant than Davidson appears to admit. If triangulation has a role in endowing the linguistic subject with the skills required for the interpretation of others, she must acquire such skills through her participation in shared practices. The child could not learn what she needs for mature communication unless a teacher was consistent in her use of expressions (her similarity responses to the environment were similar) and the similarity responses of child and teacher were similar to each other. Once the conditioning has set in, it is difficult to see how this would not constrain how a developed linguistic subject could intend to use words, because the subject would be constrained in what she believed possible, and in what possibilities she could entertain.

5.2 General Conclusions

The ease with which we use and interpret linguistic innovations gives rise to Davidson's claim that conventions are not necessary for communication. Perhaps this is even more apparent now than it was when Davidson wrote "A Nice Derangement of Epithets". The English "language" is likely to be in far greater flux now than it ever has been as a result of innovations in media and communications; moreover, the speed with which new phrases and words are coined and subsequently taken up by a vast international community of users, speaks in favour of the Davidsonian picture. I started from the position of accepting that linguistic

conventions do not determine the meaning of utterances, suggesting that Daivdson's observations are evidence enough for this.

This thesis investigated how one could come to know the meanings of another's utterances on a Davidsonian picture of language. Radical interpretation failed as an account of how this could be done. Triangulation, viewed from its role in a subject's linguistic development could provide a useful starting point for developing an account.[2] Yet as a result of triangulating with others, speakers may well face more constraints on how they intend to be interpreted than mere legitimate expectations of being understood.

Nevertheless accounting for a speaker's abilities in this way resolves any potential tension between (1) the substantive role of regularity and shared practices as a general and substantive feature of communication and (2) the claim that prior regularities and shared practices do not determine the meanings of utterances. Appeal to conventions cannot explain the meaning of utterances: legitimate expectations of how one can be interpreted can. But these, in turn, must at least in part, be explained by appeal to regularities and shared practices.

On the conclusions presented in this thesis, it would seem that Davidson is half-right. I have accepted that conventions do not determine the meanings of utterances: people do not necessarily fail to mean by an utterance what they intended to mean because they fail to speak in accordance with the relevant conventions. However, given the asymmetry between the account of meaning and the account of interpretation on this picture, an account of regularity and shared

[2] Smith (2006a: p. 403) "Davidson has provided a powerful framework for thinking about these issues and pointers to how the story could be developed further. Others will have to take the project forward, but without his contribution we should have had little idea of where to start".

practices - even if these do not count as conventions typically construed - must feature in our complete account of how we communicate linguistically.

Appendix: Plato, Davidson and Dialogue

A.1 Habit and agreement in Cratylus

Recall that at *Cratylus,* 434e-435a, we saw it acknowledged that one person could convey "what they had in mind" to another by either *habit* (or *usage*) or agreement. Socrates, like Davidson, set the criteria for whether one succeeded in using a name as whether one has been successfully interpreted. However, getting "what one has in mind" across to another could be done by either *ethos* or *sunthêke*.[1] The implication for this, however, is that neither *nature*, nor *nomos* ("law", "convention") are *necessary.* I suggested that Socrates appears motivated to this conclusion by very similar considerations to those resulting from Davidson's observations in *Derangement*: there are actual cases where speakers "get it wrong" according to the standards of their community, but nevertheless "get away with it".

A.1.1. The necessity of dialogue for interpreting others on the Hermogenean view

If this *Hermogenean* view can be attributed to Plato, it is certainly consistent with Socrates' professed distaste for both the written word, and the long speech. In a famous passage of the *Phaedrus*, we see Socrates prioritising the spoken word over the written.

> Soc: The painter's products stand before us as though they were alive, but if you question them, they maintain a most majestic silence… It's the same with written words; they seem to talk to you at though they were intelligent, but if you ask them anything about *what they say*, from a desire to be instructed, they go

[1] Socrates claims to reserve judgement on whether *sunthêke* is substantially different from *ethos*, but suggests that, *sunthêke* or at least *ethos* is sufficient for communication (435b-d).

> on telling you the same thing forever. And once a thing is put in writing, the composition, whatever it may be, *drifts all over the place*, getting into the hands of not only those who understand it, but equally of those who have no business with it; it *doesn't know how to address* the right people, and not address the wrong.
>
> *Phaed.*275d-e

The written word is static: it does not know who to address and how to address its various audiences. Socrates in many places also objects to the "long speech" as a method of philosophy; we can supply very similar reasons for this.[2] True communication requires cooperation between two participants: it is not possible to ensure that a mute text has got across "what one had in mind", if reader or audience cannot engage cooperatively with it. This is, of course, not conclusive evidence in favour of the Hermogenean view, although it fits in well with it. However, it is difficult to see why Socrates would disapprove of the written word if he espoused the *naturalist position.*

After all, if *naturalism* were correct, then the *genre* of the linguistic productions would not be at issue. In writing, say, each component name of the passage would pick out its correct nominatum (regardless of the intentions of the writer). Thus to find out what the passage means, one need only discover (perhaps though etymological investigation) the referents of those component words.[3] The written word would be less of a worry for the naturalist, then; however, the same would also be true for the

---

[2] See, for example, *Gorgias* 449b, and *Protagoras* 334cff.

[3] This is clumsily put: however, at the time of the *Cratylus* there is no evidence to suppose that Plato had any conception of a "language" as such: the focus on this dialogue is entirely on names, with the implication that the conception of "language" in operation considers language to be little more than a collection of names. No mention is made of "rules of composition" or grammar, and, moreover, Socrates in this dialogue at least seems not to be aware of a distinction between predicate and referring terms (see, e.g. 429cff and 385b-d). Thus understanding a passage or a sentence is treated as little more than understanding the individual names that comprise it. Contra Fine (1978), who claims that Socrates was to some extent aware of a distinction between predicate and referring terms, but this is not borne out by the passages mentioned.

proponent of the *nomos*-oriented view of language. On this latter view, if one wished to find out the meanings of the words, one would only need to discover what the relevant *nomoi* state, and understanding would again be possible without recourse to trying to work out what the writer "meant". Again, the words on the page could only mean what the *nomoi* determine. It would seem then that the naturalist and *nomoi*-oriented view would seem to chime less well with Socarates' (and perhaps Plato's) distaste for writing and long speeches.

A.2. Dialogue and the Possibility of Disagreement

Thus, the *Hermogenean* view fits nicely with Socrates' express concerns about the written word and long speeches. The more interesting question is whether the *Hermogenean* view is *necessary* for philosophical dialogue. I suggest there are grounds for thinking that it is necessary; however, it need not (and there is reason to believe that Plato thought it would not) lead to any form of "linguistic anarchy".

A point made by Irwin (1995) may help set the ground for this discussion. Irwin claims that Socrates denies conventionalism to the extent that he believes that there exist natural kinds, and that language should represent those kinds. So, for example, language ought to reflect the joints of nature by having distinct names for distinct natural kinds: it would be incorrect to have a single kind term, say 'trock' that picked out trees and rocks indiscriminately. The conventionalist position Socrates does attack, Irwin holds, is a view that "there is nothing about external reality itself that makes it right to classify things in one way rather that another". (Irwin, 1995:148).

The claim that this is one of the principle conclusions of the *Cratylus* is curious, since there is virtually no evidence in the text to support it (except perhaps

some opaque comments at 393b-c that are situated firmly in the etymological discussion). In fact, Socrates never appears to seriously consider names that "cut across" natural divisions as a target for an attack in this dialogue.

Yet, of course, Plato holds that there are natural divisions that exist independently of us, and these provide the standard to which our judgements about reality must conform if they are to be correct: otherwise there would be no purpose to the elenchus. It may also be fair to ascribe him the view that the one who knows what, say, justice is, will be able to pick out all and only independently existing actions that display that characteristic with the word "justice". Using the word "justice" in this way would in an important sense be *correct*, whereas using the word in such a way that picks out both just and unjust actions would be incorrect. But this "correctness" would not prevent one from communicating *what one has in mind* when one uses the word "justice" when one has an incorrect conception of what justice is.

Furthermore, it does not follow from one using the word "justice" correctly that one will have knowledge of what "justice" is. Just as the lower classes of the Republic are trained to behave correctly (in accordance with the *nomoi* of the philosopher kings), no doubt they will be able to use the word "just" correctly, but they do not do this in either case because they *know* what justice is, knowledge of justice is the province of the Guardians. Thus correct use would here be, at best, a reflection of knowledge: it is not sufficient for attributing knowledge to the user.

Language, or words, are a means to an end: words are the medium through which we can arrive at correct judgements about the world, they are not something which can themselves be investigated to arrive at correct judgements: it is far better that things be sought and learned through themselves than through names" (*Crat*.

439e). To find out what justice is, we cannot investigate what the word “justice” means.

In *Republic VII* we are told that the Philosopher-king’s studies result in his apprehending “by thought itself the nature of the good in itself” (532b): the implication may be that he no longer requires language. But even if the Philosopher-king is ultimately able to dispense with language and “commune” with the forms directly, language is an essential tool in investigating the nature of things for all who have not reached those intellectual heights. It is at the initial stages of intellectual development that we see that if the meaning of a word is determined either by nature or by conventions, it may prove disastrous for Plato’s project.

A.2.1 Getting the philosophical discussion off the ground.

Let us suppose that a lucky interlocutor at the end of an *ideal* discussion with Socrates has through elenchus actually come to have knowledge of justice.

We can say that the interlocutor and Socrates have come to the correct view that “Justice is ‘P’”, where ‘P’ stands for a correct account of what justice is. Armed with such a definition the lucky interlocutor will be able to use “justice” and its cognates to correctly pick out instances of justice and just actions in the world. At this end stage, both the interlocutor and Socrates will be disposed to use “justice” correctly, and *in the same way as each other.*

However, at the opening of the discussion the positions of the interlocutors will presumably be different (for why else would there be motivation to discuss). Take two interlocutors, A and B, starting a philosophical discussion with a disagreement.

A believes “justice is “Q”” and so is disposed to use “justice” (or “just”) to pick out actions, objects etc. that satisfy the criteria set by “Q”.

B believes “justice is “R”” and so is disposed to use “justice” to pick out actions, objects that satisfy the criteria set by “R”.

How is it that the two succeed in communicating to each other? On the *naturalist* view, it would seem impossible; “justice” means just whatever that name *naturally* picks out. Thus anyone using “justice” would either as a matter of fact be talking about whatever justice is, regardless of his own beliefs, or uttering nonsense.[4] The same would be true of the *nomos*-oriented view: if the prevailing *nomos* determined the meaning of the word, again “justice” would pick out just what that *nomos* determined to be its object. Both positions face the problem that we could never actually mis-apply the name. There would therefore be no scope for disagreement: if there was no scope for disagreement, then there would be no opening for the elenchus.

The *Hermogenean* view, which sees words as the tools by which we understand what another has in mind, seems more promising in this regard. Both parties can use one and the same name or sound to communicate different contents and through discussion their diverse beliefs can be brought to the surface and investigated. But it is only if “justice” has no fixed meaning independently of its being uttered and understood as intended, that the word can, through discussion, be used variously, as the subject of discussion changes from incorrect or unclear conceptions of justice to correct and clear ones. Davidson remarks:

> We cannot hope to learn what these virtues are without learning what the words mean, or to learn what the words mean without learning what people, ourselves and others, mean by them. People mean what others can take them to mean; to learn what we mean *is* to learn what others we talk with mean. Understanding others, agreeing with them on basic concepts, clarity about what we mean, come, to the extent that they do, together. The elenchus is a model of our only

[4] As Cratylus says, “moving their mouth in vain, much like someone banging on a brass pot” (430a).

> method for promoting these ends….Dialogue, particularly in the form of the elenchus, provides the forum in which alone words take on meaning and concepts are slowly clarified. The better we know others, the better we know what we think. (SCoT: 249- 250)

I have suggested that the naturalist, or *nomos*-oriented view of language, could not support this understanding of the elenchus, however the conclusions of the *Cratylus*, as I present them, fit in well with what Davidson says here.

However, there is a potential problem for this view. This picture is all very well when considering how the meanings of *individual* words are altered and clarified over the course of the elenchus, but what of the many other words that are used in the course of the discussion; if the meaning of every word was determined by just how its speaker was disposed to use it, surely this would render discussion impossible. Even in the case of the single word "justice": if two interlocutors had sufficiently different things in mind when using "justice" surely they would be speaking at cross-purposes. What is to say this will not happen? If it did a kind of linguistic anarchy would result: rendering meaningful communication all but impossible. Davidson provides a satisfying solution to this problem, and, moreover, a solution that I do not think would be a stretch to attribute to Plato.[5]

A.3. Avoiding Linguistic Anarchy: Triangulation and The Principle of Charity.

In considering the role of dialogue in Plato's philosophy, Davidson takes his cue from a puzzle posed by Vlastos (1983). Vlastos' famous analysis of the *Socratic*

[5] See Doyle (2010) who does indeed attribute similar acknowledged Davidsonian views to Plato.

elenchus sees Socrates first elicits some claim from an interlocutor, and then, through a process of question and answer, extracts further beliefs from the interlocutor that entail the falsity of the initial claim. *Prima facie,* the aim of the elenchus is to show that the initial beliefs of the interlocutor are false. Yet Socrates has only succeeded in showing that those beliefs are inconsistent: "There is no principle of logic that prompts us to abandon our original assumption. Giving up one or more of the subsequent admissions would do as well" (*SCoT*: 42).

So why does Socrates feel entitled to his moral beliefs? Vlastos (1983) suggests we attribute to Socrates an additional tacit assumption that "if a respondent gives a false answer, there are always further beliefs he has in light of which his answer is false". Davidson retorts, however, that this still does not provide sufficient grounds for supposing the original statement false: modifying Vlastos, Davidson suggests Socrates needs to add the following assumption:

C: There is a presumption that one's seriously held moral beliefs are true.

Davidson (*SCoT*: 244) claims that it follows from this that the elenchus will probably result in correct moral beliefs but **C** still requires justification. He believes that this can be done by appeal to the Principle of Charity: "The elements of correct interpretation are such as to ensure a large degree of truth […] creatures with thoughts must be largely consistent and correct in their beliefs" (245). An appeal to triangulation supports this view, since the evidence we depend on for learning a language and developing thought in the first place is, for Davidson, based on triangulation between the similarity of learner's and teacher's responses to the environment, and the world, "which determines the meanings of those basic

sentences, and hence also determines the contents of thoughts expressed by those sentences" (245).

Basic sentences determined by perceptually available evidence are a far cry from sentences in which the evaluative moral judgements that Socrates is interested in are stated. Davidson does not, however, see this as a problem: "our understanding of the values of others depends on enough that is shared to make sense of what is not shared, and to provide legitimate grounds for the resolution of moral differences" (SCoT: 246).

The supposition that we are basically right about the world, and our evaluative judgements, supported by Davidson's triangular externalism, provides justification for Socrates' dependence on the elenchus in the search into moral truths. Given the Davidsonian staple that to be interpretable one must have largely true beliefs, the elenchus, properly practiced, leads one to beliefs that are justified, and probably true, even though one cannot confer certainty on such beliefs.[6]

This kind of approach also shows us why linguistic anarchy would not follow from the Hermogenean view of language. Nor does it require us to attribute any anachronistic philosophical insights to Plato.[7] It is reasonable to suppose on the grounds suggested by Davidson and Vlastos above, that through general interaction with the world and their communities, the majority of a person's beliefs are true, *including her beliefs about how her communities use words in general*. People, as a matter of fact, will use words in the same ways as their fellows. It is when mere

---

[6] Socrates is notably sceptical about his beliefs, but this would give us reason to think that the results of the elenchus are superior to the starting points. At *Gorgias* 509a, Socrates comes as close as he ever does to professing certainty in his moral beliefs, claiming that the are held fast with "reasons of steel and adamant".

[7] N.B Doyle (2010: 21), for the plausibility of attributing a similar view to Plato.

interaction with the world and one's fellows does not give rise to clear beliefs (for example in the case of the virtues) that further, philosophical dialogue is required to clarify the subject matter. Nevertheless there is enough truth in one's beliefs from the start: this enables discussion to get off the ground.

> In the Socratic dialogues respondents find themselves *saying* things that are contradictory. But what this reveals is not that they have clear ideas that are in conflict; the appearance of contradiction more commonly betrays real confusion. There is a genuine sense in which one who is confused can't be straightforwardly wrong, for he simply does not know what he thinks. (SCoT: 250)

The superiority of dialogue, Davidson suggests on Socrates behalf, lies in its "supplying the nexus" for an organic development of thoughts. It is a mistake to suppose that fully-formed thoughts are exchanged through dialogue: "without language, thoughts have no clear shape; but the shape language gives them emerges only in the context of active communication" (SCoT: 249).

Dialogue enables us to clarify thought and develop clearer conceptions of the issues at stake. But this is achieved when we alter our use of words to match out evolving conceptions of the issues at stake.

> It is easy to confuse what goes on in a live conversation with what we find in a written dialogue. If we read that someone, under questioning, says, "Justice is doing good to one's friends and evil to one's enemies", and subsequently is honestly persuaded that it is not just to do evil to anyone, using the same word, "justice", we are almost certain to conclude that the speaker has radically changed his mind. […] The original statement and the later admission, we say to ourselves, contradict each other. And so they do if the word "justice" means the same thing in both occurrences. We are almost certain to take fixity of meaning for granted because as passive readers we attach the same meaning to the word each time we read it – or at least we assume the word has the same meaning from start to finish, for otherwise what are the discussants discussing? Yet in truth it often happens that what is being discussed is exactly at issue. (*D&D*: 266-7).

The best method of investigating things, according to Socrates, is through themselves, and though each other, not through names.[8] Socrates cannot learn the true nature of justice through merely finding out the conventions of his linguistic community to use "justice" in a particular way: their demotic and "incorrect" use of the word is precisely what comes under attack in our dialogues.

One cannot, in general, investigate the nature of something, without using the language available to one; nevertheless, the words we use must be our slaves, not our masters, if our investigations are to bear fruit.

[8] *Crat.* 438.

Abbreviations

Abbreviations of Articles by Davidson:

BM	Davidson, D. (2001/1974) 'Belief and the Basis of Meaning.' In, *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 141-154
C&C	Davidson, D. (2001/1984) 'Communication and Convention.' In, *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 265-280
CToT	Davidson, D. (2001/1992) 'A Coherence Theory of Truth and Knowledge.' In *Subjective, Intersubjective, Objective,* New York: Clarendon Press. pp. 137-153
Derangement	Davidson, D. (2005/1986) 'A Nice Derangement of Epitaphs.' In *Truth, Language and History,* New York: Clarendon Press. pp. 89-108
EoT	Davidson, D. (2001/1997) 'The Emergence of Thought.' In *Subjective, Intersubjective, Objective,* New York: Clarendon Press. pp. 123-134
Externalisms	Davidson, D. (2001) 'Externalisms.' In Kotatko, P., Pagin, P. and Segal, G. (eds) *Interpreting Davidson*, Center for the Study of Language and Information: Stanford. pp. 1-16
HIT	Davidson, D. (1999) 'Interpretation: Hard in Theory, Easy in practice.' In De Caro, M. (ed.), *Interpretations and Causes,* Dordrecht: Kluwer. pp. 31-46
Inscrutability	Davidson, D. (2001/1979) ' The Inscrutability of Reference.' In *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 227-242
Intending	Davidson, D. (2006/1978) 'Intending.' In *The Essential Davidson,* New York: Clarendon Press. pp. 122-137
JJHD	Davidson, D. (2005/1989) 'James Joyce and Humpty Dumpty.'In *Truth, Language and History,* New York: Clarendon Press. pp. 143-158
LLL	Davidson, D. (2005/1993) 'Locating Literary Language.' In *Truth, Language and History,* New York: Clarendon Press. pp. 167-182
M&M	Davidson, D. (2005/1993) 'Method and Metaphysics.' In *Truth, Language and History,* New York: Clarendon Press. pp. 39-46
MTE	Davidson, D. (2005/1990) 'Meaning, Truth, and Evidence.' In *Truth, Language and History,* New York: Clarendon Press. pp. 47-62
RA	Davidson, D. (2001/1982) 'Rational Animals.' In *Subjective, Intersubjective, Objective,* New York: Clarendon Press. pp. 95-106
RI	Davidson, D. (2001/1973) 'Radical Interpretation.' In *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 125-140
RII	Davidson, D. (1994) 'Radical Interpretation Interpreted.' In *Philosophical Perspectives, vol. 8: Logic and Language,* Atascadero: Ridgeview. pp.121-128
RTF	Davidson, D. (2001/1976) 'Reply to Foster.' In *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 171-180
SAL	Davidson, D. (2005/1994) 'The Social Aspect of Language.' In *Truth, Language and History,* New York: Clarendon Press. pp. 109-126

SCoT	Davidson, D. (2005/1992) 'The Socratic Concept of Truth.' In *Truth, Language and History,* New York: Clarendon Press. pp. 241-250
SP	Davidson, D. (2001/1992) 'The Second Person.' In, *Subjective, Intersubjective, Objective,* New York: Clarendon Press. pp. 107-122
T&M	Davidson, D. (2001/1967) 'Truth and Meaning.' In *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 17-42
TMLL	Davidson, D. (2001/1965) 'Theories of Meaning and Learnable Languages.' In *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 3-16
TVK	Davidson, D. (2001/1991) 'Three Varieties of Knowledge.' In *Subjective, Intersubjective, Objective,* New York: Clarendon Press. pp. 205-220
VI	Davidson, D. (2001/1974) 'On the Very Idea of a Conceptual Scheme.' In, *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 183-198

Other Abbreviations:

PoC	Principle of Charity
MI	Malaprop Interpreter
RI	Radical Interpreter
Crat.	Plato's *Cratylus*.

Bibliography

Akrill, J. L. (1994) 'Language and Reality in Plato's *Cratylus*.' In *Essays on Plato and Aristotle*. Oxford: OUP. pp. 33-52

Anagnostopoulos, G. (1972) 'Plato's *Cratylus*: The Two Theories of the Correctness of Names.' *Review of Metaphysics,* 25 (4): 691-736

Barber, A. (2001) 'Idiolectical Error.' *Mind and Language,* 16 (3): 263-283

Barber, A. (2004) 'Idiolects.' In Zalta, E. (ed.) *Stanford Encyclopaedia of Philosophy* [online]. Avaialble from: http://plato.stanford.edu/entries/idiolects/ [Accessed 9 December 2008]

Barney, R. (1997) 'Plato on Conventionalism.' *Phronesis,* 42 (2): 143-62

Bar-On, D. and Risjord, M. (1992) 'Is there Such Thing as a Language?' *Canadian Journal of Philosophy,* 22 (2): 163-90

Baxter, T. M. S. (1992) *The* Cratylus*: Plato's Critique of Naming*, Leiden: Brill.

Bestor, T. W. (1980) 'Plato's Semantics and Plato's *Cratylus*.' *Phronesis,* 25 (3): 306-330

Brink, I. (2004) 'Joint Attention, Triangulation and Radical Interpretation: A Problem and its Solution.' *Dialectica,* 58 (2): 179-206

Burge, T. (1975) 'On Knowledge and Convention.' *The Philosophical Review,* 84 (2): 249-55

Burnyeat, M. F. (1997) 'First Words: A Valedictory Lecture.' *Proceedings of the Cambridge Philological Society*, 43: 4-8

Cook, J. (2009) 'Is Davidson a Gricean?' *Dialogue,* 48: 557-575

Dalmiya, V. (1990) 'Coherence, Truth and the 'Omniscient Interpreter.' *The Philosophical Quarterly,* 40 (158): 86-94

Davidson, D. (1994) 'Radical Interpretation Interpreted.' In *Philosophical Perspectives, vol. 8: Logic and Language,* Atascadero: Ridgeview. pp.121-128

---------- (1999) 'Interpretation: Hard in Theory, Easy in practice.' In De Caro, M. (ed.), *Interpretations and Causes,* Dordrecht: Kluwer. pp. 31-46

---------- (2001) 'Externalisms.' In Kotatko, P., Pagin, P. and Segal, G. (eds) *Interpreting Davidson*, Center for the Study of Language and Information: Stanford. pp. 1-16

---------- (2001/1965) 'Theories of Meaning and Learnable Languages.' In *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 3-16

---------- (2001/1967) 'Truth and Meaning.' In *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 17-42

---------- (2001/1973) 'Radical Interpretation.' In *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 125-140

---------- (2001/1974) 'Belief and the Basis of Meaning.' In, *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 141-154

---------- (2001/1974) 'On the Very Idea of a Conceptual Scheme.' In, *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 183-198

---------- (2001/1975) 'Thought and Talk.' In *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 155-170

---------- (2001/1976) 'Reply to Foster.' In *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 171-180

---------- (2001/1978) 'What Metaphors Mean.' In, *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 245-264

---------- (2001/1979) ‘ The Inscrutability of Reference.’ In *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 227-242
---------- (2001/1982) ‘Rational Animals.’ In *Subjective, Intersubjective, Objective,* New York: Clarendon Press. pp. 95-106
---------- (2001/1984) ‘Communication and Convention.’ In, *Enquiries into Truth and Interpretation,* New York: Clarendon Press. pp. 265-280
---------- (2001/1984) ‘First Person Authority.’ In *Subjective, Intersubjective, Objective,* New York: Clarendon Press. pp. 3-14
---------- (2001/1991) ‘Three Varieties of Knowledge.’ In *Subjective, Intersubjective, Objective,* New York: Clarendon Press. pp. 205-220
---------- (2001/1992) ‘A Coherence Theory of Truth and Knowledge.’ In *Subjective, Intersubjective, Objective,* New York: Clarendon Press. pp. 137-153
---------- (2001/1992) ‘The Second Person.’ In, *Subjective, Intersubjective, Objective,* New York: Clarendon Press. pp. 107-122
---------- (2001/1997) ‘The Emergence of Thought.’ In *Subjective, Intersubjective, Objective,* New York: Clarendon Press. pp. 123-134
---------- (2005/1985) ‘Plato’s Philosophy.’ In *Truth, Language and History,* New York: Clarendon Press. pp. 223-240
---------- (2005/1986) ‘A Nice Derangement of Epitaphs.’ In *Truth, Language and History,* New York: Clarendon Press. pp. 89-108
---------- (2005/1989) ‘James Joyce and Humpty Dumpty.’In *Truth, Language and History,* New York: Clarendon Press. pp. 143-158
---------- (2005/1990) ‘Meaning, Truth, and Evidence.’ In *Truth, Language and History,* New York: Clarendon Press. pp. 47-62
---------- (2005/1992) ‘The Socratic Concept of Truth.’ In *Truth, Language and History,* New York: Clarendon Press. pp. 241-250
---------- (2005/1993) ‘Locating Literary Language.’ In *Truth, Language and History,* New York: Clarendon Press. pp. 167-182
---------- (2005/1993) ‘Method and Metaphysics.’ In *Truth, Language and History,* New York: Clarendon Press. pp. 39-46
---------- (2005/1994) ‘Dialectic and Dialogue.’ In *Truth, Language and History,* New York: Clarendon Press. pp. 251-261
---------- (2005/1994) ‘The Social Aspect of Language.’ In *Truth, Language and History,* New York: Clarendon Press. pp. 109-126
---------- (2006/1978) ‘Intending.’ In *The Essential Davidson,* New York: Clarendon Press. pp. 122-137
Davis, W., (2003), *Meaning, Expression and Thought*, Cambridge: CUP.
De Caro, M. (1999) ‘Davidson in Focus.’ In De Caro, M. (ed.) *Interpretations and Causes,* Dordrecht: Kluwer. pp. 1-30
De Caro, M. (ed.) (1999), *Interpretations and Causes*, Dordrecht: Kluwer.
De Rosa, R. (1999) ‘Is there a problem about Davidson’s Holism vis a-vis his Holism.’In De Caro, M. (ed.) *Interpretations and Causes*, Dordrecht: Kluwer. pp. 201-216
Donnellan, K. S. (1966) ‘Reference and Definite Descriptions.’ In *The Philosophical Review*, 77: pp. 281-302
Donnellan, K. S. (1968) ‘Putting Humpty Dumpty Together Again.’ In *The Philosophical Review*, 77: 203-215
Doyle, J. (2010) ‘Socrates and Gorgias.’ *Phronesis* 55 (1): 1-25

Dummett, M. (1986)'"A Nice Derangement of Epitaphs": Some Comments on Davidson and Hacking.' In Lepore, E. (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson,* Blackwell. pp. 459-476.
Evans, G. (1975) 'Identity and Predication.' *Journal of Philosophy*, 72 (13): 343-363
Ferrari, G. R. F. (1990) *Listening to the Cicadas: A Study of Plato's* Phaedrus, Cambridge: CUP
Fine, G. M. (1977) 'Plato on Naming.' *Philosophical Quarterly*, 27 (109): 289-301
Fodor, J. and Leopre, E. (1994) 'Is Radical Interpretation Possible?' In *Philosophical Perspectives, vol. 8: Logic and Language,* Atascadero: Ridgeview. pp. 101-119
Geach, P. T. (1966) 'Plato's *Euthyphro:* An Analysis and Commentary," *The Monist*, 50: 369-82
George, A. (1990), 'Whose Language is it Anyway? Some notes on idiolects.' *Philsophical Quaterly*, 40 (160): 275-98
Gilbert, M. (1983) 'Agreements, Conventions, and Language.' *Synthese*, 54 (3): 375-407
---------- (1992) *On Social Facts*, Princeton: Princeton UP.
Grice, H. P. (1957) 'Meaning.' *Philosophical Review*, 66 (3): 377-88
Gustafsson, M. (1999) 'Systematic Meaning and Linguistic Diversity: The Place of Meaning Theories in Davidson's Later Philosophy.' *Inquiry*, 41 (4): 435-53
---------- (Unpublished) 'Familiar Words in Unfamiliar Surroundings: Davidson's Malapropisms, Cavell's Projections.' [online] Available from: https://www.abo.fi/media/6840/150210gustafsson.pdf. [Accessed 7 July 2010]
Guthrie, W. K. C. (1971) *The Sophists*, Cambridge: CUP.
Hacker, P. M. S. (1997) 'Davidson on First Person Authority.' *The Philosophical Quarterly,* 47 (188): 285-304
Hacking, I. (1986) 'The Parody of Conversation. ' In Lepore, E. (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell. pp. 447-458.
Kahn, C. H. (1973) 'Language and Ontology in the *Cratylus*.' In Lee, E. N., Mourelatos, A. P. D. and Rorty, R. M. (eds.) *Exegesis and Argument: Studies in Greek Philosophy Presented to Gregory Vlatsos; Phronesis Supplementary Volume*, 1. Leiden: Brill. pp. 152-176
Kemmerling, A. (1990) 'The Philosophical Significance of a Shared Language.' In Stoecker, R. (ed.), *Reflecting Davidson: Donald Davidson Responding to an International Forum of Philosophers,* Berlin: Walter de Gruyter. pp. 85-116
Kotatko, P. (1998) 'Two Notions of Utterance Meaning.' *Proceedings of the Aristotelian Society*, 98 (3): 225-239
Kretzmann, N. (1971) 'Plato on the Correctness of Names.'*American Philosophical Quarterly*, 8 (2): 126–138
Kripke, S. (1977) 'Speakers Reference and Semantic Reference.' *Midwest Studies in Philosophy*, 2: 255-276
Laurence, S. (1996) 'A Chomskian Alternative to Convention-Based Semantics.' *Mind*, 105: 269-301
Lepore, E. and Ludwig, K. (2005) *Donald Davidson: Meaning, Truth, Language, Reality*, Oxford: OUP.
Levin, S. B. (2001) *The Ancient Quarrel Between Philosophy and Poetry Revisited*, Indianapolis: Hackett.
Lewis, D. (1969) *Convention*, Cambridge MA: Harvard UP.

MackKenzie, M. M. (1986) 'Putting the *Cratylus* in its Place.' *Classical Quarterly*, 36: 124-150
McCabe, M. M. (2008) 'Plato's Ways of Writing.' In Fine, G. (ed.) *Oxford Handbook to Plato*. Oxford: OUP. pp. 88-113.
McGinn, C., (1977), 'Charity, Interpretation and Belief.' *The Journal of Philosophy*, 74 (9): 521-535
McKay, A. F. (1968) 'Mr. Donnellan and Humpty Dumpty on Referring.' *Philosophical Review*, 77 (2): 197-202
Milikan, R. (2005) *Language: A Biological Model*, Oxford: OUP.
Miller, A. (1998) *Philosophy of Language*, New York: Routledge.
Mole, C. (2006) 'Attention.' In Zalta, E. (ed.) *Stanford Encyclopaedia of Philosophy* [online]. Avaialble from: http://plato.stanford.edu/entries/attention/ [Accessed 16 September 2010]
Pagin, P. (2001) 'Semantic Triangulation.' In Kotatko, P., Pagin, P. and Segal, G. (eds) *Interpreting Davidson*, Center for the Study of Language and Information: Stanford. pp. 199-212
Picardi, E (1999) 'Sensory Evidence and Shared Interests.' In De Caro, M. (ed.) *Interpretations and Causes*, Dordrecht: Kluwer. pp. 171-186
Pietroski, P. (1994) 'A Defence of *Derangement*.' *Canadian Journal of Philosophy*, 24: 95-118
Quine, W. V. O. (1960), *Word and Object*, Boston: MIT Press.
Quine, W. V. O. (1981) *Theories and Things,* Cambridge MA: Harvard UP.
Ramberg, B. T. (1991) *Donald Davidson's Philosophy of Language: An Introduction,* New York: Wiley-Blackwell.
Reeve, C. D. C. (1999) *Plato:* Cratylus, Cambridge: CUP.
Reimer, M. (2004) 'What Malapropisms Mean: A reply to Donald Davidson.' *Erkenntnis*, 60: 317-334
Rescorla, M. (2007) 'Conventions.' In Zalta, E. (ed.) *Stanford Encyclopaedia of Philosophy* [online]. Avaialble from: http://plato.stanford.edu/entries/convention/ [Accessed 2 February 2010]
Robinson, R. (1969) *Essays in Greek Philosophy*, Oxford: OUP
Schofield, M. (1982) 'The Dénouement of the *Cratylus.*' In Schofield, M. and Nussbaum, M. (eds.), *Language and Logos*, Cambridge: CUP. pp. 61-81
Sedley, D. (2003) *Plato's Cratylus*, CUP.
Simpson, D. (2003) 'Interpretation and Skill: On Passing Theory.' in Preyer, G., Peter, G, & Ulkan, M. (eds.) *Concepts of Meaning: Framing an Integrated Theory of Linguistic Behaviour,* Dordrecht: Kluwer. pp. 251-266
Smith, B. C. (2006a) 'Davidson, Interpretation and First-Person Constraints on Meaning.', pp., in, *International Journal of Philosophical Studies*, 14 (3): 385-406
---------- (2006b)'What I Know when I Know a Language.' In Lepore, E. and Smith, B. (eds.) *The Oxford Handbook of Philosophy of Language.* pp. 941-982
Smith, I. (2008) 'False Names, Demonstratives, and the Refutation of Linguistic Naturalism in Plato's *Cratylus,* 427d1-431c3.' *Phronesis*, 53 (2): 125-151
Talmage, C. J. L. (1994) 'Literal Meaning, Conventional Meaning, and First Meaning.' *Erkenntnis*, 40 (2): 213-225
Talmage, C. J. L. (1996) 'Davidson and Humpty Dumpty.' *Nous,* 30 (4): 537-544
Talmage, C. J. L. (1997) 'Meaning and Triangulation.' *Linguistics and Philosophy*, 20 (2): 139-145.

Tarski, A. (1944) ‘The Semantic Conception of Truth and the Foundations of Semantics.’ *Philosophy and Phenomenological Research*, 4: 341-76
Verheggen, C. (2007) ‘Triangulation with Davidson.’ *The Philosophical Quaterly*, 57 (226): 95-103
Vermazen, B. (1982) ‘General Beliefs and the Principle of Charity.’*Philosophical Studies*, 42 (1): 111-118
Vlastos, G. (1983)‘The Socratic Elenchus.’ *Oxford Studies in Ancient Philosophy*, 1: 27-58
Wikforss Å. (2010/2001) ‘Semantic Noramativity.’ In Byrne, D. and Köbel, M. (eds.), *Arguing about Language*, London: Routledge. pp. 455-469
Williams, B. (1982) ‘Cratylus' Theory of Names and Its refutation.’In Schofield, M. and Nussbaum, M. (eds.) *Language and Logos*, Cambridge: CUP. pp. 83-93
Wittegenstein, L. (1973/1950). *Philosophical Investigations*, London: Prentice-Hall.