

Why We Need Corpus Linguistics in Intuition-Based Semantics

By Leonid Tarasov

This is a pre-print of an article forthcoming in *Grazer Philosophische Studien*

DOI 10.1163/18756735-000052

Abstract:

The following method is popular in some areas of philosophy and linguistics when trying to describe the semantics of a given sentence Φ . Present ordinary speakers with scenarios that involve an utterance of Φ , ask them whether these utterances are felicitous or infelicitous and then construct a semantics that assigns the truth-value True to felicitous utterances of Φ and the truth-value False to infelicitous utterances of Φ . I make five observations about this intuition-based approach to semantics; their upshot is that it should be revised in favour of a more nuanced method. I suggest that this method should be based on corpus linguistics and make some tentative remarks about what it might look like and which questions we need to address in order to develop it.

Key words: Semantics; Linguistic Intuitions; Corpus Linguistics; Methodology.

1. Introduction

The following approach is popular in some areas of philosophy and linguistics when attempting to give a description of the semantics of a given sentence Φ . We should present

ordinary speakers¹ with scenarios that involve an utterance of Φ ² and ask them whether these utterances are felicitous or infelicitous.³ We can call the speakers' responses here linguistic intuitions. We then need to explain why speakers have these intuitions, and a high premium is placed on explanations that appeal only to the semantics of Φ , as opposed to explanations that appeal to pragmatics, speakers' psychology or other factors. More precisely, a high premium is placed on describing a semantics that assigns the truth-value True to Φ in those scenarios that involve an utterance of Φ that ordinary speakers find felicitous and the truth-value False in those scenarios that involve an utterance of Φ that speakers find infelicitous.⁴ We can call this assigning a matching truth-value to utterances of Φ , or truth-value matching for short.

¹ An ordinary speaker is presumed to be someone who is fluent in the language under investigation, but not philosophically or linguistically trained.

² I use 'utterance' in the technical sense of spoken or written production of a word or sentence, i.e. a tokening of a word or sentence type.

³ Felicity is a technical notion which lacks a clear definition in the literature. It goes back at least to J. L. Austin (1975 [1955], 14), who characterises infelicitous utterances as 'types of cases in which something *goes wrong* and the act [performed by the utterance of a sentence] is therefore at least to some extent a failure: the utterance is then, we may say, not indeed false but in general *unhappy*' (original emphasis). The modern notion of felicity is wider than Austin's. If an utterance of a sentence is felicitous, ordinary speakers feel that the utterance complies with the general rules of the language, whether syntactic, semantic or pragmatic. They might use expressions like 'grammatical', 'appropriate', 'right', 'true' or just 'sounds fine' to describe the utterance. If an utterance of a sentence is infelicitous, ordinary speakers feel that the utterance violates some rule(s) of the language, whether syntactic, semantic or pragmatic. They might use expressions like 'ungrammatical', 'inappropriate', 'not right', 'false' or just 'sounds weird' to describe the utterance. For instance, an utterance of any of the following sentences would be infelicitous: 'Furiously sleep ideas green colourless', 'Colourless green ideas sleep furiously' (Chomsky 2002 [1957], §2.2), 'It is raining, but I do not believe it' (Moore 1993, 207-208). Usually, the infelicity of an utterance of the first sentence is traced to violation of syntactic rules, the second to semantic rules and the third to pragmatic rules. Semanticists tend to work with grammatical sentences, so the salient questions here are whether and which semantic rules explain the felicity or infelicity of an utterance of a given sentence.

⁴ Ordinary speakers' linguistic intuitions are usually determined through empirical work or by reference to the readers' linguistic intuitions (or both). The first method, which is common in linguistics, has at least two issues. Firstly, it is not unusual to rely on very small groups of ordinary speakers, so it is possible that their linguistic intuitions are not representative of ordinary speakers' linguistic intuitions more generally. Secondly, even amongst small groups there can be wide variation in linguistic intuitions. The second method, which is common in philosophy, has issues as well. The most obvious is just that a theorist's readers will tend to be other theorists,

There is a debate about how much consideration we should give to speakers' responses to utterances of Φ relative to other factors, which speakers should be involved or which of the speakers' responses are relevant when describing the semantics of Φ (see fn. 4). While I do have something to say in response to these questions (see Tarasov 2014, 17-26), here I would like to make some observations about truth-value matching that are independent of these questions. Together these observations suggest that truth-value matching should be replaced with a more nuanced approach to describing the semantics of a given sentence. I make a few suggestions about what this approach might look like in the conclusion.

Truth-value matching is by no means the exclusive approach to semantics. For example, it is also popular to imagine what a speaker intends to communicate when they utter a particular sentence and to try to explain their intentions in terms of the semantics of the sentence (see, e.g., Stanley and Szabó 2000, 219). However, exactly how this latter method compares to truth-value matching falls under the questions of how much consideration we should give to speakers' linguistic intuitions relative to other factors and whether any of the speakers' responses to utterances besides linguistic intuitions are relevant, which I set aside here.

Throughout this paper I draw all my examples from the debate about the semantics of knowledge attributions and denials ('knowledge attributions' collectively), i.e. sentences of the form $\ulcorner x \text{ knows that } \Phi \urcorner$ and $\ulcorner x \text{ doesn't know that } \Phi \urcorner$. This debate can be expressed in terms

i.e. not philosophically or linguistically untrained speakers, and therefore not ordinary speakers in the relevant sense of the term. And, in parallel with the first method, a theorist's readership may be relatively small and may fail to share the same linguistic intuitions. Moreover, the justification for thinking that linguistic intuitions provide any evidence about the semantic features of a sentence is unclear. These issues raise questions about exactly how much consideration we should give to speakers' linguistic intuitions relative to other factors, which speakers should be involved and whether any of the speakers' responses to utterances besides linguistic intuitions are relevant. I set aside these questions here, but for some relevant discussion see, e.g., Alexander and Weinberg (2007), Bach (2005), Culbertson and Gross (2009), Daly and Liggins (2011), Devitt (2006) and Tarasov (2014, 17-26).

of a popular semantic framework (Kaplan 1989) and is notorious for its reliance on linguistic intuitions, so it makes for a good case study. However, truth-value matching is also at work in many other debates. For instance, it is used to argue for all forms of relativist semantics. To give just one example, it is at work in John MacFarlane's case for a relativist semantics for sentences containing predicates of personal taste, like 'tasty' and 'sweet'. MacFarlane (2014, 21-22) makes several observations about utterances of sentences containing 'tasty' that he suggests we need to account for or explain, and then accounts for them by giving a semantics that assigns matching truth-values to utterances of these sentences (MacFarlane 2014, 144-66). Therefore, my examples from the debate about the semantics of knowledge attributions should easily generalise.

2. Truth-value matching in action

To see how truth-value matching works in practice, consider two well-known scenarios, originally due to Keith DeRose (1992, 913; DeRose 2009, 61-66).

Bank Case A

It is Friday; Larry and Jeff's bank will be open tomorrow; Larry was at the bank last Saturday and confidently believes that the bank will be open tomorrow; Jeff is driving past the bank with his partner Suzie; he has a cheque with him, but it is not especially important that he deposits the cheque before Monday. Jeff says, 'Let's deposit the cheque tomorrow'. Suzie says, 'Are you sure? Many banks are closed on Saturdays'. Jeff says, 'No, I asked Larry and he knows that the bank will be open'.

Bank Case B

As with A, except it is very important that Jeff deposits the cheque before Monday. Jeff says, 'Let's deposit the cheque tomorrow'. Suzie says, 'Are you sure? Many banks are closed on Saturdays'. Jeff says, 'No, I asked Larry and he knows that the bank will be open'. Suzie says, 'But banks change their hours'. Jeff says, 'No, Larry knows that the bank will be open'.

If you are anything like DeRose, you will have the intuition that the utterance of the sentence 'He [Larry] knows that the bank will be open' is felicitous in Bank Case A and infelicitous in B. If we are truth-value matching, then we need to give a semantics for the sentence 'He [Larry] knows that the bank will be open' that assigns it the truth-value True in Bank Case A and the truth-value False in Bank Case B. For instance, DeRose would tell us to opt for attributor contextualism. Very roughly put, according to attributor contextualism 'know' makes a different contribution to propositional content relative to different contexts of utterance. Therefore, 'He [Larry] knows that the bank will be open' expresses different propositions in different contexts of utterance. In particular, 'He [Larry] knows that the bank will be open' may express a true proposition in Bank Case A and a false proposition in Bank Case B (see DeRose 2009, 47-66).⁵

Now, it invariably turns out that there are scenarios where a semantic theory for Φ does not assign a matching truth-value to Φ . That is, invariably there are cases where an utterance of Φ is judged felicitous but a semantic theory for Φ assigns the truth-value False to Φ , or cases

⁵ DeRose gives a more detailed presentation of Bank Cases A and B that may affect the strength of an ordinary speaker's response to the cases. He also includes a felicitous utterance of a knowledge denial, rather than an infelicitous utterance of a knowledge attribution, in Bank Case B. Since we are not arguing about the semantics of knowledge attributions here, it does not matter exactly how the cases are presented as long as they elicit broadly equivalent responses.

where an utterance of Φ is judged infelicitous but a semantic theory for Φ assigns the truth-value True to Φ . If a semantic theory does not assign a matching truth-value, it is said that the semantic theory implies commitment to an error theory.⁶ The more cases there are where a semantic theory does not assign a matching truth-value, the larger the error theory implied by the semantic theory. For example, consider Bank Case C, based on some criticisms of attributor contextualism made by John Hawthorne (2004, 98-104).

Bank Case C

It is Friday; Larry's bank will be open tomorrow; he was at the bank last Saturday and confidently believes that the bank will be open tomorrow; he is driving past the bank with his partner Cheryl; he has a cheque with him, but it is not especially important that he deposits the cheque before Monday. Larry says, 'Let's deposit the cheque tomorrow'. Cheryl says, 'Are you sure? Many banks are closed on Saturdays'. Larry says, 'No, I know that the bank will be open, I was there last Saturday'. Later on, Cheryl is talking to Suzie; Cheryl now also has a cheque with her, and it is very important that she deposits it before Monday. Cheryl says, 'Let's deposit the cheque tomorrow'. Suzie says, 'Are you sure? Many banks are closed on Saturdays and yours

⁶ There is a lack of clarity around exactly what the term 'error theory' refers to. According to one view, error theories explain why a given semantics does not assign a matching truth-value to a given sentence, usually on the basis of pragmatic or psychological factors about utterances of that sentence. So in this sense an error theory is a kind of add-on to the semantics. According to another view, a theory T is an error theory about a discourse D if and only if according to T none of the sentences of D are true (Daly and Liggins 2010, 209). In this sense the semantics itself can be the error theory; the pragmatic, psychological or other explanations of the error are still an add-on to the semantics, but they are not called error theories themselves. Since we do not engage with any explanations of error here, it does not matter if we leave the term 'error theory' ambiguous in this respect. Elsewhere (Tarasov forthcoming) I have used the term 'non-semantic theory' to refer to the pragmatic, psychological and other explanations of cases where the relevant semantics does not assign matching truth-values.

might have changed its hours'. Cheryl says, 'Well, I asked Larry and he said that he knows that the bank will be open'.

If you are anything like Hawthorne, you will have the intuition that the utterance of the sentence 'I asked Larry and he said that he knows that the bank will be open' is felicitous. It is well-known that here attributor contextualism fails to assign a matching truth-value to this sentence; specifically, it assigns the truth-value False to 'I asked Larry and he said that he knows that the bank will be open'. The reason is that according to attributor contextualism the proposition expressed by 'I [Larry] know that the bank will be open' as uttered by Larry is different from the proposition expressed by 'He [Larry] knows that the bank will be open' as uttered by Cheryl as part of her utterance of the sentence 'I asked Larry and he [Larry] said that he [Larry] knows that the bank will be open'. Since the two propositions are different, Cheryl does not correctly report that Larry said that he knows that the bank will be open on Saturday. Therefore, attributor contextualism implies commitment to an error theory.⁷

Since it is invariable that there are cases where a semantic theory does not assign a matching truth-value, a semantic theory is rarely rejected on these grounds. In other words, that a semantic theory implies an error theory is usually tolerated. In these cases pragmatic, psychological or other non-semantic factors may be called in to explain why speakers have the linguistic intuitions that they do. However, since the aim is to construct a semantics that assigns matching truth-values, it follows that any error theory implied by a given semantic theory should be as small as possible, and explanations that appeal to pragmatic, psychological or

⁷ Bank Case C is an instance of a general problem that attributor contextualists have with embedding constructions, e.g. 'x said that y knows that Φ ', 'x believes that y knows that Φ ', 'x suspects that y knows that Φ ', and so on (x can be identical with or distinct from y). A number of solutions have been offered, which I have discussed at length elsewhere (Tarasov 2014: 67-75). I do not think any of the solutions are satisfactory. I am grateful to Johannes Brandl for urging me to note some of the broader background for Bank Case C.

other factors should be invoked as a last resort. Similarly, as far as this aim is concerned, one semantic theory is preferable to another if the former implies a smaller error theory than the latter. As we might expect then, many of the arguments in debates that are based on truth-value matching focus on creating scenarios in which a rival semantic theory of Φ does not assign matching truth-values to utterances of Φ , and around comparing theories for the number of scenarios in which these theories assign matching truth-values. For instance, according to assessor relativism, ‘know’ makes the same contribution to propositional content in all contexts, but its extension varies across contexts of assessment. As a result, ‘He [Larry] knows that the bank will be open’ expresses the same proposition in all contexts of utterance, but has a different truth-value relative to different contexts of assessment. In particular, assessor relativists would claim that the sentence ‘He [Larry] knows that the bank will be open’ expresses the same proposition in Bank Case B as it does in Bank Case A, but it is true in A and false in B. Moreover, according to assessor relativism the proposition expressed by ‘I [Larry] know that the bank will be open’ as uttered by Larry is the same as the proposition expressed by ‘He [Larry] knows that the bank will be open’ as uttered by Jeff as part of his utterance of the sentence ‘I asked Larry and he [Larry] said that he [Larry] knows that the bank will be open’. So according to assessor relativism Jeff correctly reports that Larry said that he knows that the bank will be open on Saturday. Therefore, assessor relativism is able to assign a matching truth-value in Bank Cases A, B and C (e.g. MacFarlane 2014, 187-90). On this basis, assessor relativists often argue that their view is preferable to attributor contextualism (e.g. MacFarlane 2014, 198).

To sum up then, in practice truth-value matching works as follows. We present an ordinary speaker with scenarios that involve utterances of Φ and ask them whether these utterances are felicitous or infelicitous. We then construct a semantic theory for Φ that assigns matching truth-values to Φ . If this is not possible, and usually there are cases where it is not,

we acknowledge that the semantic theory implies commitment to an error theory. We then compare our semantic theory against its rivals and look for the one that implies the smallest error theory; this semantic theory is then taken to describe the semantics of Φ .

3. Five observations about truth-value matching

We are now in a position to make some observations about truth-value matching. Let me begin with a question: why would we think that a semantic theory for a sentence Φ that is constructed on the basis of linguistic intuitions given in response to scenarios involving utterances of Φ describes the semantics of Φ ? The obvious answer is just that we work, explicitly or implicitly, on the basis of an assumption that there is a relation between linguistic intuitions that arise in response to scenarios involving utterances of Φ and the semantics of Φ . In fact, I think it is fairly obvious that we work on the basis of a stronger assumption, viz. that the utterances of Φ in the scenarios we give to ordinary speakers to canvass their linguistic intuitions are indicative of the way we actually use Φ . If we did not work on the basis of this assumption, it would be difficult to understand what relevance we thought the method of creating scenarios involving utterances of Φ and canvassing speakers' linguistic intuitions in response to these utterances has to finding out the semantics of Φ .

Let us call utterances of Φ in scenarios that we construct and give to speakers to canvass their linguistic intuitions 'test utterances' and the utterances that constitute the actual use of Φ 'live utterances'. For example, Bank Case A involves a test utterance of a knowledge attribution that we assume to be representative of live utterances of that knowledge attribution. We can summarise our present observation as follows:

- (i) Truth-value matching is based on the assumption that a test utterance of Φ is representative of live utterances of Φ .

This observation is a simplification. In reality, when theorists truth-value match, in most cases they assume that a test utterance of Φ is representative not only of live utterances of Φ , but also of live utterances of sentences that are in relevant respects similar to Φ . For instance, they would assume that the utterance of ‘He [Larry] knows that the bank will be open’ in Bank Case B is representative not only of live utterances of ‘He [Larry] knows that the bank will be open’, but also live utterances of, say, ‘Smith knows that the plane will stop in Chicago’ (Cohen 1999: 58) in cases similar to Bank Case B. To put it another way, the assumption is that a test utterance of a sentence type is representative of both live utterances of that sentence type and live utterances of relevantly similar sentence types. However, in order to keep things relatively straightforward, we will ignore this point and stick with the simplified observation for the rest of this paper.

I take it that a test utterance of Φ is representative of live utterances of Φ only if there has been at least one live utterance of Φ in circumstances that are sufficiently similar to the scenario in which the test utterance is made. For instance, the test utterance of ‘He [Larry] knows that the bank will be open’ in Bank Case A is representative of live utterances of ‘He [Larry] knows that the bank will be open’ only if there has been at least one live utterance of ‘He [Larry] knows that the bank will be open’ in circumstances that are sufficiently similar to Bank Case A. For our purposes we do not need to worry about what exactly sufficient similarity amounts to here. However, we do need to consider the sufficient condition(s) for a test utterance to be representative of live utterances. Clearly, the condition above is not both necessary and sufficient. That is, a test utterance of Φ is not representative of live utterances of Φ *if and only if* there has been at least one live utterance of Φ in circumstances that are

sufficiently similar to the scenario in which the test utterance is made. This is far too weak. For one, it includes one-off or anomalous uses of Φ , i.e. uses which quite clearly do not need to be accounted for in a description of the semantics of Φ . Instead, I suggest that something like the following is true: a test utterance of Φ is representative of live utterances of Φ if and only if there has been a sufficiently high number of live utterances of Φ in circumstances that are sufficiently similar to the scenario in which the test utterance is made. Again, there is an issue about what exactly a sufficiently high number is. I will not address this here, but I would like to point out that these conditions raise an interesting issue about the status of the test utterances we use to canvass speakers' linguistic intuitions. Specifically, they suggest that not all test utterances are on a par. To show what I mean here, take Bank Cases A, B and C again. Suppose that there is a very large number of live utterances of 'He [Larry] knows that the bank will be open' in circumstances that are sufficiently similar to Bank Case A, but very few live utterances of 'He [Larry] said that he [Larry] knows that the bank will be open' in circumstances that are sufficiently similar to Bank Case C. Then there is a clear sense in which the utterance of 'He [Larry] knows that the bank will be open' in Bank Case A is representative of more live utterances of 'He [Larry] knows that the bank will be open' than the utterance of 'He [Larry] said that he [Larry] knows that the bank will be open' in Bank Case C is representative of live utterances of 'He [Larry] said that he [Larry] knows that the bank will be open'. More generally then, the following observation is true:

(ii) Some test utterances of Φ are representative of more live utterances of Φ than other test utterances.

Observation (iii) is a corollary of (ii):

(iii) Test utterances may be ranked on the basis of the number of live utterances they represent, from most to least.

Let us call the ranking of test utterances on the basis of how many live utterances they represent a representation ranking. We will say that a test utterance scores highly on the representation ranking if there is a large number of live utterances of Φ in sufficiently similar circumstances. It scores low on the representation ranking if there is a small number of live utterances of Φ in sufficiently similar circumstances.

As we have seen, usually we cannot give a description of the semantics of Φ that assigns a matching truth-value to all test utterances of Φ ; invariably, a description of the semantics of Φ implies a commitment to an error theory. We have also seen that usually the way we choose between rival semantic theories is by comparing the number of test utterances the theories assign matching truth-values to; we choose the theory that assigns matching truth-values to the highest number of test utterances. However, I suggest that, in light of (i) to (iii), this view should strike us as overly simplistic. In particular, it is natural to consider not only how many test utterances a semantic theory assigns matching truth-values to, but also how highly the test utterances that the theory assigns matching truth-values to score on the representation ranking. For example, suppose we have just five test utterances of Φ ; call them U_1 through U_5 . They score high to low on the representation ranking, with U_1 being the highest scoring and U_5 the lowest. Now imagine we have two rival semantic theories of Φ – T_1 and T_2 – where T_1 assigns matching truth-values to U_1 and U_2 , and T_2 assigns matching truth-values to U_3 , U_4 and U_5 . If we should choose a theory that assigns matching truth-values to the highest number of test utterances, then we should choose T_2 . However, if we should choose a theory on the basis not only of how many test utterances it assigns matching truth-values to, but also how

highly these test utterances score on the representation ranking, then it may be that we should choose T_1 . Let us put this observation as follows:

(iv) If we need to choose between rival semantic theories of Φ , we should choose on the basis not only of how many test utterances of Φ the theories assign matching truth-values to, but also how highly the test utterances these theories assign matching truth-values to score on the representation ranking.

The justification for (iv) is implied by (i). That is, we have seen that truth-value matching is based on the assumption that test utterances are representative of live utterances. If we did not in some way rely on this assumption, then it would be unclear what relevance we thought the method of creating scenarios involving utterances of Φ and canvassing speakers' linguistic intuitions in response to these utterances has for finding out the semantics of Φ . By the same token, we should favour test utterances that are representative of a high number of live utterances than those that are representative of a low number; just as we want to avoid accounting for one-off or anomalous uses of a sentence when describing its semantics, so we want to avoid accounting for test utterances that score low on the representation ranking at the expense of those that score higher.

Finally, the last observation I would like to make is this:

(v) Observations (i) to (iv), and especially (ii) to (iv), are not generally recognised by those who truth-value match.

There are some hints at observations (i) to (iv) in the literature. For example, in his response to the problem for attributor contextualism posed by Bank Case C, DeRose (2009,

161-74) suggests that we can construct scenarios where a knowledge attribution is uttered, but an utterance of a report of this knowledge attribution is infelicitous, and he argues on this basis that it is unclear what conclusions we should draw from Bank Case C. Others have suggested that when we describe the semantics of knowledge attributions, we can ignore their utterances in sceptical scenarios, i.e. scenarios that involve the attribution of knowledge that one is not a brain-in-a-vat, that one is not being deceived by an evil demon and so on (see, e.g., Hazlett 2010). One way to understand these sorts of suggestions is as an implicit acknowledgement of something like (i) to (iv) above. For example, perhaps the reason we can ignore utterances of knowledge attributions in sceptical scenarios is that they are representative of a very low number of live utterances of knowledge attributions. Nonetheless, it is fair to say that observations (i) to (iv) are not generally recognised. The evidence for this comes from just how common it is to choose a semantic theory *only* on the basis of how many test utterances it assigns matching truth-values to.

4. Conclusion

At least one consequence of observations (i) to (v) above is obvious, viz. even if we do not dispute the role that ordinary speakers' linguistic intuitions play in semantics, truth-value matching presents a flawed methodology. When we present ordinary speakers with test utterances of a given sentence and give a semantic theory for that sentence based on speakers' linguistic intuitions in response to these test utterances, we should pay attention not only to the number of test utterances the theory assigns matching truth-values to, but the number of live utterances the test utterances represent.

There are several difficult questions to answer here. First, what exactly does it take for a test utterance to represent a live utterance? Specifically, following our comments in the

previous section, what exactly are a sufficiently high number of live utterances and a sufficient similarity between circumstances? Second, how should we weigh the number of test utterances a theory assigns matching truth-values to against how highly these test utterances score on the representation ranking? For instance, do we favour a theory that assigns matching truth-values to just a few test utterances, but ones that score very highly on the representation ranking, or a theory that assigns matching truth-values to many test utterances, but ones that score only reasonably highly on the representation ranking? And third, how exactly should we go about measuring how many live utterances a given test utterance represents?

I do not have answers to the first two questions. However, I believe that we can begin to answer the third. Corpus linguistics is the method of collecting and analysing linguistic corpora, i.e. bodies of written or spoken text (see, e.g. McEnery and Wilson 2001, 29). In our present terminology, it is the method of collecting and analysing sets of live utterances.⁸ Corpus linguistics is therefore ideally placed to establish whether a given test utterance of Φ represents live utterances of Φ and how many live utterances it represents (see Devitt 2012, 27-32).⁹ We need to build a corpus of live utterances of Φ in circumstances that are sufficiently similar to the scenario in which the test utterance is made and determine

⁸ Presumably, most live utterances are felicitous; otherwise they would not be uttered. For instance, presumably, if an ordinary speaker finds the utterance of the sentence ‘No, Larry knows that the bank will be open’ in Bank Case B infelicitous, this is because the speaker would not themselves utter it in these circumstances (see fn. 3). If this is right, then linguistic corpora are unlikely to contain many, if any, infelicitous live utterances (see fn. 5). This raises another potential issue for truth-value matching: if truth-value matching should be based on linguistic corpora, and if linguistic corpora do not contain infelicitous utterances, to what extent should we bother writing a semantics that assigns matching truth-values to infelicitous utterances?

⁹ Michael Devitt advocates the use of corpus data in discussions of theories of reference. He calls live utterances ‘the linguistic usage’, and what I call representation of live utterances checking ‘the linguistic reality’ (Devitt 2012, 27). Devitt (2012, 29-30) also suggests that corpora can be supplemented with what he calls ‘elicited production’: instead of presenting ordinary speakers with utterances in certain scenarios and asking the speakers if those utterances are felicitous or infelicitous, we can present ordinary speakers with incomplete scenarios and ask them to complete the scenarios with whatever utterance they feel would be felicitous in the circumstances. I am grateful to a reviewer for alerting me to Devitt’s work on corpus linguistics.

whether the number of live utterances contained in this corpus is sufficiently high.¹⁰ Corpus linguistics also gives us an easy handle on the representation ranking, viz. test utterances would rank higher or lower on the representation ranking depending on the size of the linguistic corpora of the live utterances they represent.¹¹

¹⁰ In the last ten years there has been a surge in so-called Experimental Philosophy (X-Phi) (see Knobe and Nichols 2017). I am grateful to a reviewer for suggesting that X-Phi, instead of (or in addition to) corpus linguistics, might be able to address the worries I have raised for truth-value matching. While there is a wide variety of methods within X-Phi, they are united by the aim of approaching certain philosophical problems by in some sense canvassing ordinary speakers' (in the sense of fn. 1) responses with respect to those problems. For example, a very crude X-Phi experiment might present ordinary speakers with Bank Cases A and B and canvass their intuitions in response to utterances of knowledge attributions in those cases. This kind of X-Phi experiment is not suitable as an alternative to truth-value matching because it still only considers test utterances, albeit directly from the point of view of ordinary speakers (see fn. 4). Put another way, replacing philosophers or linguists with a cross-section of ordinary speakers does not make the utterances that are tested any more representative of the day-to-day use of the expressions in question; the test utterances might still be incorporated into scenarios that have no day-to-day parallels. More nuanced X-Phi methodologies might involve building and analysing corpora of ordinary speakers' live utterances or their intuitions in response to live utterances. I am enthusiastic about these methodologies, although I confess ignorance about whether and to what extent they differ from corpus linguistics or other empirically driven disciplines.

¹¹ I am grateful to a reviewer for noting that the use of corpus data may itself involve an intuition-based matching exercise: 'If I examine a corpus, I need to decide which utterances belong in the set of live utterances for the test utterance. This may involve a judgment of similarity in meaning, in the way or sense in which a term is used, in its syntactic structure, etc. Might I then be faced with the task of developing an analysis of the corpus which is evaluated to the extent to which it aligns with my intuitions about which utterances in the corpus are or are not part of the set of [relevant] live utterances?'. The worry, I take it, is two-fold. First, there is a concern about the practicalities of introducing corpus linguistics into truth-value matching: how do we judge whether a given utterance in a corpus is relevant or not? Second, there is a concern about my principles: if the point of introducing corpus data is to move away from the kind of intuition mongering inherent in constructing iterations of Bank Cases and the like without regard for live utterances, then it looks like this point is undermined if we now have to engage in intuition mongering about which corpus data are relevant.

The first concern properly falls under the question of what is it for a test utterance to represent a live utterance. Once we know the answer to this, we will know which utterances belong in the set of live utterances for the test utterance, i.e. just those live utterances which are represented by the test utterance. As noted in the main text, I do not have an answer to the question of what it is for a test utterance to represent a live utterance. However, I wholeheartedly agree with the reviewer that this is a legitimate and significant worry, so I add some very speculative thoughts here that hopefully go a little way towards addressing it.

We are left with the following picture of the alternative to truth-value matching. Present ordinary speakers with test utterances of Φ and canvass the speakers' linguistic intuitions in response to these utterances. Develop a semantic theory that assigns matching truth-values to Φ . In doing so, take account of the number of test utterances the theory assigns matching truth-values to and how highly these test utterances score on the representation ranking, based on an analysis of what it is for a test utterance to represent a live utterance, a theory of the correct way to weigh the number of test utterances against their scoring on the representation ranking, and on corpora of live utterances. What remains, then, is the substantial task of fleshing out this approach.

Acknowledgements: I would like to thank Johannes Brandl, Thomas Uebel and an anonymous referee for discussion, comments and suggestions. I would also like to thank Rob Knowles, Graham Stevens and Jack Winter for discussion of ideas that motivated this paper.

What might it mean for the utterance of 'He [Larry] knows that the bank will be open' in Bank Case A to represent live utterances? As the reviewer notes, we need to consider both similarities between sentences and similarities between the circumstances in which they are uttered. For the sake of the argument, this may be divided into syntactic similarity, semantic similarity, the risk attached to the object proposition (in this case that the bank will be open), the number of error possibilities made salient (in this case one: that many banks close on Saturdays) and the risk attached to each possibility, and the evidence that the subject has for the relevant proposition (in this case the evidence Larry has for the proposition that the bank will be open). So I suggest that here representation is a matter of passing a threshold of similarity of syntax, semantics, risk, salience and evidence. Crucially, there are already ways of measuring and modelling these factors (e.g. risk and evidence are modelled in decision science, salience is studied in cognitive sciences, etc.), so we know we do not need to involve an intuition-based matching exercise. This point should also alleviate the worry about my principles: the idea is not to engage in intuition mongering about which corpus data are relevant, but rather to select this data based on a holistic approach to measures of syntax, semantics, risk, salience and evidence. The central question – and I admit a very large question – is just how to develop these measures and models in a way that suit our present purposes.

References

- Alexander, Joshua and Jonathan Weinberg 2007. “Analytic Epistemology and Experimental Philosophy.” *Philosophy Compass* 2, 56–80.
- Austin, J. L. 1975. *How to Do Things With Words*. Oxford: Oxford University Press.
- Bach, Kent 2005. “Context Ex Machina”. In: *Semantics vs. Pragmatics*, edited by Zoltán Szabó, Oxford: Oxford University Press, 15-44.
- Chomsky, Noam 2002. *Syntactic Structures*. Berlin: Mouton de Gruyter.
- Cohen, Stewart 1999. “Contextualism, Skepticism, and the Structure of Reasons.” *Noûs* 33, 57–89.
- Culbertson, Jennifer and Steven Gross 2009. “Are Linguists Better Subjects?” *The British Journal for the Philosophy of Science* 60, 721–736.
- Daly, Chris and David Liggins 2010. “In Defence of Error Theory.” *Philosophical Studies* 149, 209–230.
- 2011. “Differentialism.” *Philosophical Studies* 156, 321–37.
- DeRose, Keith 1992. “Contextualism and Knowledge Attributions.” *Philosophy and Phenomenological Research* 52, 913–929.
- 2009. *The Case for Contextualism: Knowledge, Skepticism, and Context, Vol. 1*. Oxford: Oxford University Press.
- Devitt, Michael 2006. “Intuitions in Linguistics.” *The British Journal for the Philosophy of Science* 57, 481–513.
- 2012. “Whither Experimental Semantics?” *Theoria* 27, 5-36.
- Hawthorne, John 2004. *Knowledge and Lotteries*. Oxford: Oxford University Press.
- Hazlett, Allan 2010. “The Myth of Factive Verbs.” *Philosophy and Phenomenological Research* 80, 497–522.

- Kaplan, David 1989. "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals." In: *Themes from Kaplan*, edited by Joseph Almog, John Perry, & Howard Wettstein, Oxford: Oxford University Press, 481-563.
- Knobe, Joshua and Shaun Nichols 2017. "Experimental Philosophy." In: *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), edited by Edward N. Zalta, <https://plato.stanford.edu/archives/win2017/entries/experimental-philosophy/>.
- MacFarlane, John 2014. *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford: Oxford University Press.
- McEnery, Tony and Andrew Wilson 2001. *Corpus Linguistics: An Introduction*. 2nd Revised edition. Edinburgh: Edinburgh University Press.
- Moore, G. E. 1993. "Moore's Paradox." In: *G. E. Moore: Selected Writings*, edited by Thomas Baldwin, London: Routledge, 207-212.
- Stanley, Jason and Zoltán Szabó 2000. "On Quantifier Domain Restriction." *Mind & Language* 15, 219–261.
- Tarasov, Leonid 2014. *The Semantics of Knowledge Attributions: A Defence of Moderate Invariantism*. Manchester: The University of Manchester (PhD thesis).
- forthcoming. "Projective Adaptivism." *Philosophical Papers*.