

the

Wolfgang Klein (Nijmegen)

Arguments are to be avoided; they are always vulgar and often convincing.
Oscar Wilde

Many things in the study of human languages are puzzling; but what I find most puzzling is how little we know about them. At present, there are about 6000 languages on earth (that is, about 30 per country, if each language is counted only once). How many of these are well-described? That depends on the criteria for being well-described. Surely, a brief grammar and a fragmentary lexicon by some missionary or adventurous explorer is not enough. Let us say, for the sake of the argument, that a language is well-described if there are at least three grammars and three dictionaries. Over the years, I have asked numerous linguists for their estimate on the number of languages they believe meet this criterion. Answers vary, but they go hardly beyond 100. So, less than 2% of the world's existing languages are well-described, more than 98% are not. This sheds an interesting light on the reliability of many bold statements on universals and typological properties: any claim about grammatical facts in more than 100 languages is on a very weak footing.

One might argue that the criterion "three grammars and three dictionaries" is excessive. I do not share in this view. In fact, one may wonder how many grammars and dictionaries and papers and books are needed to get a reliable picture of a language or, somewhat more modestly, a particular linguistic phenomenon. We can take any language, so why not English, the *drosophila melanogaster* of linguistics, and here its most frequent word - the word *the*. If anything, that word should be well-described. So, what is the meaning of the most-frequent word of the most-investigated language of the world? Bertrand Russell was not the first who tried to uncover the meaning of *the*; but he was the first to become famous for his attempt. His article *On denoting* (1905) ends with the sentence:

I will only beg the reader not to make up his mind against the view [expressed here] - as he might be tempted to do, on account of its apparently excessive complication - until he has attempted a theory of his own on the subject of denotation.

Many have taken up this challenge. Has a better analysis been found after one century of intensive research - an analysis which covers the apparentia and on which the experts are agreed?

The experts are not agreed. This is obvious from a look at the endless literature on definiteness in general and the definite article in particular. Is there an analysis which does justice to the facts, or at least the core facts? No. I have two arguments, or, perhaps less vulgar, two considerations to offer. First, the reader is kindly invited to take any paper or book in which a theory of definiteness is proposed, and then check how many occurrences of *the* in that very paper or book fit the proposed theory. Second, we will go through a few examples, none of them exotic, and see how the best-known theories fare on them.

It is, of course, impossible to do justice to all theories on three pages, and what I have to say is at a very global level. Essentially, there are three - not mutually exclusive - ways to account for definiteness and the definite article:

- (a) The referent of the NP must be *unique*. Russell's account is a typical example.
- (b) The referent of the NP must be *familiar*, i.e., introduced in earlier context or rooted in world knowledge. Christopherson's analysis of the English definite article stands for this idea.
- (c) The referent of the NP must be *identifiable* for speaker and listener. This is probably the oldest account, found in many traditional grammars but also in more recent linguistic work.

These ideas have been elaborated in various forms, not to be discussed here. There are some other, less popular, views, such as Lewis' salience idea. For these, the reader is invited to check whether they can account for the examples or not.

It is clear that mere familiarity does not suffice. In 1, for example, the referent of *the boy* is explicitly introduced in the immediately preceding sentence - still, *the boy* is not felicitous here:

- (1) (A boy came in, then a girl, then another boy.) The boy was about five years old.

There are too many boys. It would seem, therefore, that familiarity or “knownness” is only relevant, if only one referent with that particular properties (here being a boy) is familiar or known. A uniqueness account would predict this. It is less clear whether the criterion “identifiable for speaker and listener” applies here. In a way, the two boys are no less identifiable than the girl in this discourse context. So, the problem is rather what “identifiable for someone” means. If it means “identifiable because unique”, then we are back to the uniqueness feature. If not, I do not understand it.

The uniqueness idea is also in line with 2 where the referent is clearly not introduced but unique for other reasons:

- (2) The only book that Hindenburg ever read is the Bible.

The Bible has the definite article because there is only one Bible in the interlocutors’ world knowledge, and *the only book* is definite because it is the only one that Hindenburg has ever read - it is unique. But why can’t one say then:

- (3) The book that Hindenburg ever read is the Bible.

If the definite article on *the book* indicates that the referent is unique (in context), then this should suffice. But it does not. So, uniqueness cannot be the only criterion, either. Note that this reasoning also applies in sentences without *ever*; it only makes it clear. The point is rather that it would not make sense to add a uniqueness marking such as “only” or “*einzig*” in German, if the noun phrase is unique already.

Let me now turn to a very different type of example:

- (4) Each natural number larger than five is the sum of three prime numbers.

This is Goldbachs Vermutung (in its original form), still not proven after 250 years. The relevant noun phrase here is *the sum of three prime numbers*. There are infinitely many prime numbers (Euklid!), hence, infinitely many sums of three prime numbers. Therefore, the uniqueness idea does not work here. The referent is not “introduced” or “familiar”. Is it identifiable by the listener? What, for Christ’s sake, would this mean? In a way, every number is identifiable for every reasonable person. It is not generic, if you ask me. What is the generic sum of three prime numbers? Nor do I think that the referent is particularly salient, for adepts of David Lewis. In other words, no existing account seems to fare very well with examples of this sort - and they are not exotic, although they do not reflect a type of texts normally investigated in linguistics.

This brings me to a last type of examples (in fact, these are the reason for me to write this squib) - extracts from legal texts. Here I switch to German, because I am more familiar with German texts of this type. Here are the first two sentences of *Bürgerliche Gesetzbuch* - the German Civil Law - generally considered to be a masterpiece of juridical expression (by lawyers):

- (5) §.1. Die Rechtsfähigkeit des Menschen beginnt mit der Vollendung der Geburt.
§ 2. Die Volljährigkeit tritt mit der Vollendung des achtzehnten Lebensjahres ein.

Six noun phrases - all of them definite. None of their referents was introduced before, except in a very global sense (the BGB is about legal norms). So, familiarity as an explanation for the definite article is more or less ruled out here. How about uniqueness? This is somewhat less clear. Surely, none of these notions is unique in the sense in which *the author of “Waverly”* is unique. Could these noun phrases be generic? I do not think so. Take, for example *die Geburt*: this term does not refer to the generic birth, in the sense in which *the bush wolfe* in *The bush wolfe is almost extinct* is generic. We are

talking here about the particular completion of a particular birth of a particular person, whose “Rechtsfähigkeit” is at stake. So, none of the usual criteria seems to apply here (I leave to the reader to decide how the identifiability account, the saliency account, the individuation account, or his or her own account, fare better).

Note that in some of these cases, the definite article can be replaced by some other determiner. Thus, we can also say *Die Rechtsfähigkeit eines Menschen beginnt mit der Vollendung seiner Geburt*. But we cannot have *einer Geburt*, nor can we have *einer Vollendung* or *eines achtzehnten Lebensjahres*. Take, for example, ... *mit einer Vollendung der Geburt*. This somehow suggests the idea that there could be several *Vollendungen einer Geburt* - so, we have a kind of “relativised uniqueness” - the uniqueness of *Vollendung* is relative to the attribute *Geburt* in the same noun phrase. But note that we can also have *Eine Vollendung der Geburt war nicht zu erwarten*. Here, the same type of “relativised uniqueness” is given, and still, the indefinite noun phrase is perfectly appropriate, just as its definite counterpart.

Usages of the definite article such as in 5 are in no way peculiar. Almost every paragraph in a law or any other legal text shows them - and they seem to escape the familiar treatments of definiteness. This, I believe, is not atypical of a great deal of linguistic analyses even in the best-described languages: they are built on an unbalanced diet of examples. Nor is it specific to the definite article. By far most literature on tense and aspect, for example, focuses on single events in the past or present, with occasional glances at the future and, less so, generic events or states. But this is only one type of discourse in which temporality is expressed; we also find it, for example, in route directions, in film retellings or even in legal texts. There is little hope to understand how the expression of temporality works if only one type of discourse is considered.

I started with the somewhat gloomy idea that we are at best well-informed about 2% about the world’s languages. But are English, German, French or Chinese well-described? We can safely say that they are much-described, and anything else depends again on what “well-described” means. In 1935, Bertrand Russell stated a few principles on when someone is entitled to defend a particular opinion:

- (1) that when the experts are agreed, the opposite opinion cannot be held to be certain; (2) that when they are not agreed, no opinion can be regarded as certain by a non-expert; and (3) that when they all hold that no sufficient grounds for a positive opinion exist, the ordinary man would do well to suspend his judgment.

By that criterion, the ordinary man is well-advised to suspend his judgment on most linguistic issues for quite a while. But everybody is encouraged to change this - there is no risk to run out of supply.