Why Ambiguity?

Ariel Cohen
Ben-Gurion University
arikc@bgu.ac.il

All human languages are ambiguous. Ambiguity is not restricted to some special constructions that linguists are fond of discussing, but is quite ubiquitous. It is hard, in fact, to find a sentence that is *not* ambiguous. This fact is all too familiar to computational linguists: "One often hears in computational linguistics about completely unremarkable sentences with hundreds of parses, and that is in fact no exaggeration." (Abney, 1996). The question is, simply, why? Why is language ambiguous?

Krifka (2002) raises a similar question concerning vagueness. He shows that there are good reasons for language to allow, and even encourage vagueness. Vagueness, however, is not ambiguity. When we are told that the theater is far from here, we may be unsure as to the precise distance; but we know what the speaker intends, and we can draw inferences based on this, e.g., that we should take a cab instead of walking to the theater.

But with an ambiguous term, the intended meaning is not merely insufficiently specified; it is not known, until the term is disambiguated. When we hear that John has a kid, we draw very different inferences if John is a father or if he owns a young goat.

Thus, the advantages of vagueness do not seem to apply to ambiguity, and we are back to the question: why ambiguity?

Some may say that it doesn't matter. Humans possess very powerful mechanisms for disambiguation; these mechanisms resolve ambiguities, so that we are not normally aware of them. But this is not really an answer: these powerful mechanisms are there *because* language is ambiguous. If it weren't, they wouldn't have developed. Moreover, these mechanisms must require extra processing time. We know that all senses of an ambiguous word are accessed first, and only then does disambiguation occur (Swinney, 1979). If language were not ambiguous, we would need to access fewer meanings, and we would be saved the additional process of disambiguation. And, of course, powerful though they unquestionably are, these disambiguation mechanisms are not perfect, and misunderstandings do occasionally occur, with consequences varying from the hilarious to the disastrous. Thus, although our capacity for disambiguation may attenuate some of the disadvantages of ambiguity, these disadvantages are still there. Why would language burden itself with such a handicap?

It might be felt that language, because of its complexity, *must* be ambiguous—that any naturally developing formal system of that level of complexity must be ambiguous. While artificial formal languages are designed not to be ambiguous, their complexity is lower than that of natural language by orders of magnitude; and even with these relatively simple languages, lack of ambiguity is accomplished by cumbersome and somewhat unnatural methods (e.g., brackets in mathematics and logic). Perhaps it is very hard to design an unambiguous complex language, so that for languages that evolved naturally, not by design, there is no hope of avoiding ambiguity.

Perhaps there is some merit to this claim, perhaps not; but what *is* clear is that language could easily have been, if not completely unambiguous, then much less ambiguous than it actually is. We know that languages are subject to *some* pressure to eliminate ambiguity: the devices of word order, case markings, agreement, and the like do not seem to have any other function than the avoidance of ambiguity. And if a language loses one of these devices, e.g., case markings, then another device, e.g., word order, becomes more powerful and restrictive.

With very simple modifications, such devices could have evolved to reduce ambiguity to a much greater extent than they actually do. Agreement with indirect objects, for example, would substantially reduce PP attachment ambiguities. And yet, such a device is relatively rare.

Another popular answer follows Zipf's (1949) Law. The idea is that the current state of language is the result of the interaction between two competing goals: the good of the speaker and the good of the hearer. From the point of view of the hearer, the ideal language would contain no ambiguities, so that there would be no misunderstandings. In contrast, the ideal language from the point of view of the speaker is maximally ambiguous, so that one word would mean everything. Thus, speakers would require no effort at all in constructing their utterances: every utterance will always be the same word. The current state of language (namely ambiguous, but not maximally so), it is claimed, is the result of a balance between these two opposing forces.

This idea sounds appealingly simple, but is ultimately problematic. It rests on the assumption that ambiguity reduces the number of expressions in a language, hence making the speaker's job, of choosing between them, easier. But there is no evidence that this is so: the fact that *kid* means *child* as well as *young goat* does not mean that the word *child* is about to disappear from the language. Quite the contrary, in fact: languages tend to be *redundant*, i.e. use several expressions to express the same meaning. This means that, if anything, there is pressure to *increase*, rather than *decrease*, the number of expressions in a language.

It might be argued that, while ambiguity may not reduce the total number of words in a language, it reduces the number of *long* words. It is well known that short words tend to be more ambiguous than long words. Perhaps, then, the role of ambiguity is to "recycle" the short words, so that long words would only be used infrequently. In this way, speakers' job is made easier, since the words they need are usually short, hence easy to produce.[1]

However, as mentioned above, it appears that the driving force behind language is redundancy rather than economy, and this seems to be the case here as well. If there were substantial pressure on language users to reduce the number of long words, we would expect many more nonsense short words to be used up, before turning to ambiguity. Why go to the trouble of ambiguity, when there are many strings that do not even have a single interpretation? Why not, in English, hang some meaning onto, e.g., *nisk, nime, dace, bish, nast*, etc., before attaching it to a word that is already used?

We seem to be forced to the conclusion that there is some pressure on languages to be ambiguous: there is some advantage to using an ambiguous language. But what could this advantage be?

Wasow *et al.* (2005) make some tentative suggestions. They consider the case of two distinct dialects spoken in adjacent regions, and of a people living close to the border between the regions. Then, if some expression E has different meanings in the two regions, it will be to the advantage of these people to use E ambiguously in their dialect. Another case Wasow *et al.* consider is that of speakers who wish to mislead without actually lying, by using an expression that can be interpreted in more than one way.

While ambiguity may well be advantageous in such cases, it does not seem plausible that such a prevalent phenomenon in all of the world's languages is motivated by a handful of rather special cases, and that this is sufficient to override the substantial disadvantages of ambiguity.

As far as I can see, the reason for the ambiguity of language remains a puzzle—we simply don't know why language is ambiguous. An idea presented in a science fiction novel (Brin, 1995, p. 583) appears to provide as good an answer as any:

> "Pray, what could language *be* for, if not to maintain a culture's cohesion and foster communication?"

> "There is another desirable thing," Sara replied. "Another product of language, just as important, in the long run, as cohesion."

---

[1] I am indebted to Hans-Martin Gaertner for this suggestion.

"And that is?"

"*Creativity.* If I'm right, it calls for a different kind of grammar. A completely different way of looking at error."

"One that *welcomes* error, embraces it." Dedinger nodded. "This part of your paper I had trouble following. You say Anglic [a future descendant of English] is better [than the unambiguous alien languages] because. . . errors and ambiguity creep into every phrase or paragraph. But how can *chaos* engender inventiveness?"

"By shattering preconceptions. By allowing illogical, preposterous, even obviously *wrong* statements to parse in reasonable-sounding expressions. . ."

"This is good?"

"It's how creativity works."

References

Abney, S. P.: 1996, 'Statistical Methods and Linguistics'. In: J. L. Klavans and P. Resnik (eds.): *The Balancing Act: Combining Symbolic and Statistical Approaches to Language.* Cambridge, MA: MIT Press, pp. 1–26.

Brin, D.: 1995 *Brightness Reef.* New York: Bantam Spectra.

Krifka, M.: 2002, 'Be brief and vague! And how Bidirectional Optimality Theory allows for verbosityand precision'. In: D. Restle and D. Zaefferer (eds.): *Sounds and systems: studies in the structure and change. A Festschrift for Theo Vennemann.* : Berlin/New York: Mouton de Gruyter, pp. 439–458.

Swinney, D.: 1979, 'Lexical access during sentence comprehension: (Re)consideration of context effects'. *Journal of verbal learning and verbal behavior* **18**, 645–660.

Wasow, T., A. Perfors, and D. Beaver: 2005, 'The Puzzle of Ambiguity'. In: O. Orgun and P. Sells (eds.): *Morphology and the Web of Grammar: Essays in Memory of Steven G. Lapointe.* Stanford: CSLI Publications

Zipf, G. K.: 1949, *Human Behaviour and the Principle of Least Effort.* New York: Addison-Wesley.