1. Introduction

I should start with a disclaimer. In case anyone is afraid that this talk will actually be confessional, I can ease your minds from the get-go. This talk is actually about the interplay of methodology and theory in linguistics, but if I’d called it that, no one would have wanted to listen to it. More narrowly, it’s about a long-standing question in lexical semantics whether it’s possible in languages to have absolute synonyms, that is, words that mean exactly the same thing and that can therefore substitute for each other in any, or nearly any context. The point that I will try to be making is that a question that has been argued largely on the basis of our intuitions about the way language works can now be studied using enormous searchable corpora, and that the answers we come up with using this methodology are often quite surprising. All of this comes out of the research done for the recently published *Oxford Reference Guide to English Morphology* that I wrote with my colleagues Laurie Bauer of Victoria University in Wellington, New Zealand and Ingo Plag of Duesseldorf University in Germany.

First, some preliminaries. The area of linguistics in which I do my research is called morphology. Narrowly viewed, it has to do with how native speakers of a language form complex words from smaller pieces called morphemes, for example, how it is we can, if we need to, come up with a word like (1), which is comprised of seven pieces, called ‘morphemes’, each of which has its own meaning or function:

(1) non-reconstitutionalized
    non- re-constitute-ion-al-ize-ed

Over the last thirty years, morphologists have had this narrow goal in mind and have concentrated on the precise rules that allow us to put morphemes together to form acceptable new words. In recent years, however, some of us have begun to take a wider view of our task. That is, there are some morphologists, myself among them, who take the view that morphology is actually the study of the mental lexicon, the part of our linguistic knowledge that is concerned with the way in which native speakers of a language create, store, access, and compute both simple words and complex words. Rather than focusing exclusively on the formal mechanisms that put morphemes together, we look at how morphemes and rules are arranged in the mental lexicon, how they are stored, activated, and accessed in a stream of speech or writing in relation to each other.
As this talk progresses, I’ll look first at simple words, and then move on to complex words. As I mentioned at the outset, the specific issue that I’ll focus on is the putative non-existence of absolute synonyms and a theoretical principle in morphology known as ‘blocking’ that arises out of the belief that native speakers of a language generally avoid synonymy.

Let’s start with a simple thought experiment that I often use in my Introduction to Linguistics class: are the words couch and a sofa synonyms? Ask this question to a room full of students and you’d be surprised what heated opinions emerge. Both couches and sofas are articles of furniture that seat more than one person, both are generally upholstered. But students will invariably insist that they are not perfect synonyms: someone will say that one is larger than the other, one fancier and the other comfier, one is a piece of furniture that you have in the family room or play room and the other in the living room, one is in good condition and the other a frat house wreck. Mind you, I doubt that any of these students has ever thought about the words couch and sofa before, but I’m always surprised at the sort of lively debate that the question prompts. Students never agree on what the difference is, but they always agree that there must be some difference.

What is interesting about our thought experiment is that couch and sofa are words that are familiar, and not confined to one dialect or another, or one sociolinguistic register or another. Neither is more learned or more colloquial than the other. So our sense should be that couch and sofa are absolute synonyms. But native speakers of English are quite reluctant to say that they are.

So are linguists. If you look at texts in lexical semantics, you find statements like the following:

> Words are said to be absolute synonyms if they are substitutable in any possible context with no changes in denotation or other aspects of meaning (including connotation…). Using that criterion, it is easy to see that very few words are absolute synonyms. (Murphy 2010:110)

One thing becomes clear once we begin a serious quest for absolute synonyms, and that is that if they exist at all, they are extremely uncommon. Furthermore, it would seem reasonable to predict that if the relationship were to occur, it would be unstable. There is no obvious motivation for the existence of absolute synonyms in a language, and one would expect either that one of the items would fall into obsolescence, or that a difference in semantic function would develop (Cruse 1986: 270).

So the amateurs and the experts seem to have the same intuitions on this point. I used to too. But I don’t anymore, and this is where methodology comes in.

2. Methodology

The methodology that morphologists have typically used in exploring the contents of the mental lexicon is to rely on our intuitions as native speakers; that is, we think up a number of examples, we ask ourselves whether our examples are acceptable or not, that is, whether they are or could
be words, and if they are words what they mean and whether they mean the same thing. On the latter point, if we are conscientious, we might check the OED or other dictionaries. This method of relying on native speaker intuitions (this usually means our own and the five or six people in the offices closest to ours) is the one embraced by those of us schooled in the Chomskian tradition – initially by syntacticians (linguists who study sentence structure) but also by morphologists, who of course study words rather than sentences. So when syntacticians work, they construct sentences that have certain properties and try to decide whether they are acceptable or not, and then construct their theoretical analyses so that the rules they postulate predict that pattern of acceptability:

(2) What syntacticians do:
   a. You saw the program.
   b. What did you see?
   c. I wonder who saw the program.
   d. *What do I wonder who saw?

In the example in (2), we see that all of the sentences are acceptable except for the last one, and we would then ask ourselves why it’s not possible to form a question like the one in (2d). Our intuitions in such cases are usually pretty clear.

Like syntacticians, morphologists have always assumed that our intuitions about whether a word is acceptable or whether two words mean the same thing are every bit as good as our intuitions about whether a sentence is acceptable or not, or what it means.

But are our intuitions reliable? What I want to do next is to show how sorely mistaken we have been in the assumption that they are. I’ll start out by illustrating this with our couch ~ sofa example, but then go on to talk about related issues that have arisen repeatedly in the analysis of complex words, the real territory of morphologists.

Luckily these days we have methodological options beyond our own intuitions, thanks to a new tool at our disposal, the linguistic corpus. Linguistic corpora are massive searchable collections of language that ideally include both spoken and written text from a wide variety of genres. Corpora of English began being available around twenty years ago. But until 2008, all of these corpora (for English at least) were proprietary (you had to pay to use them), and worse, were only searchable if the user had some degree of computer programming ability. Since 2008, however, thanks largely to the work of Mark Davies at Brigham Young University, several large corpora have become freely available, and even better, with a user interface that just about anyone can figure out. The largest of these corpora – currently at 450 million words – is called the Corpus of Contemporary American English or COCA for short. It contains both written and spoken American English with the written English balanced for kinds of writing – academic, newspaper, magazine, fiction, and so on. A corpus of British English is available as well, but it’s not nearly as large, and it’s not updated yearly, as COCA is. So I’ll stick to COCA for the purposes of this talk.

[screen shot of COCA]
Let me show you a bit of what we find if we search for the words couch and sofa in COCA:

*Country Living 2012*: My kitten often "kneads" things: blankets, the couch, even me!

*Esquire 2012*: A family could inhabit this couch, could lean all the way back and laugh.

*Mother Jones 2012*: I settled onto a plush leather couch and absorbed the decor.

*New England Review 2011*: She planted one hand on a couch cushion and tried to hoist herself up.

*BK: Deader Homes and Gardens 2012*: He found some wine and decided to take a nap on the sofa.

*Southern Review 2012*: George is sitting on a sofa reading the newspaper when Will appears in the lobby ten minutes later.

*Good Housekeeping 2012*: If your sofa is looking a little winter-weary, bring on some spring with Pottery Barn's 18-inch-square Birdcage Pillow (potterybarn.com) in a cotton-linen blend

*Parenting Early 2012*: Make music with pots and pans, build a fort out of sofa cushions.

These are only a few of the thousands of available examples with these words, but trust me that I have not cherry-picked my examples. In every case it looks to me like you can substitute couch for sofa and vice versa. And lest someone propose that there are idioms in which one word cannot substitute for the other, for example that you can be a couch potato but not a sofa potato, I offer the following visual counterexample:

photo of tee shirt saying “sofa potato”

My conclusion is that both lay persons and linguists are loathe to believe in the existence of synonymy, but that nevertheless synonymy does seem to be possible.

This introduction is meant to be taken as a cautionary tale about relying on our intuitions about the meanings of words. But the real subject of this talk is the use of intuition when it comes to the possibility of coining and using new complex words. This is an issue that is of great importance to current linguistic morphologists, as I will show.

2. Blocking

The words couch and sofa are what linguistic morphologists might call simplex, that is, they cannot be divided into separate morphemes. Another way of describing words like couch and sofa is to call them monomorphemic. For a linguistic morphologist like myself, the question of whether synonymy exists becomes even more interesting when we move from simplex to complex words, that is, to words that consist of more than one morpheme.
One reason that the existence of synonymy in complex words is particularly interesting in English is because of its complex history. English has a range of prefixes and suffixes, some that are native to English (coming from our original Anglo-Saxon lexicon) and others that are borrowed from French, Latin, or Greek. The result of this complex history is that English often seems to have multiple affixes (affixes is a term that includes both prefixes and suffixes) that have the same meaning or function. So it’s especially interesting to ask whether the avoidance of synonymy, if it exists at all, extends to complex words where there are so many more opportunities for synonymy to arise.

Indeed, it is quite clear that linguistic morphologists believe that the avoidance of synonymy extends to complex words. In fact, this belief has been elevated into a theoretical principle widely known as *blocking*:

*Blocking* is the nonoccurrence of one form due to the simple existence of another. (Aronoff 1976: 43).

Blocking involves two expressions, one potential and one actual. We say that a potential expression is prevented from occurring because another expression with the same meaning and function already exists. (Aronoff & Fudeman 2011:235)

Blocking may be due to the prior existence of another word with the meaning that the putative word would have (Aronoff, 1976). Usually perfect synonyms are avoided (Katamba & Stonham 2006:75).

What this means for complex words is that given two word formation processes that seem to have the same meaning or function, the existence of a word formed with one would preclude the creation of another word with the other; at the very least, if we find such doublets, they would have to have different meanings. What I want to do next is to present a number of cases that have been prominent in the morphological literature over the years to illustrate the concept of blocking.

**Case 1** is the formation of plural nouns in English. Most everyone would agree that given an irregular plural like *feet*, a regular plural like *foots* is blocked. Or that if both the regular and irregular plurals exist, that they must have different meanings – for example, *mice* – little four-footed vermin – versus *mouses* – pointing devices used with computers. Our intuitions tell us fairly strongly that this is the case.

The formation of plurals is what is known as an inflectional process, that is, a kind of morphology that conveys grammatical distinctions rather than making new words. But the theory of blocking is also meant to apply to morphological processes that do create new words, which is called derivational morphology.

This leads us to **Case 2**. In English we form agent or instrument nouns – that is nouns that denote a person or thing that performs the action – by adding the suffix *-er* to verbs. So from *write* we get *writer* and from *print* we get *printer* and so on. Common wisdom among
morphologists has it that given a simplex word for an agent noun, the corresponding *er* form is blocked. So for example, given thiefer, some linguists argue that the word stealer is prevented from occurring.

A more interesting class of examples arises with what are called ‘rival affixes’. These are cases of pairs or even triples of affixes that attach to the same kind of roots and appear to have the same overall meaning. I’m going to divide these into two different types, which will give us Case 3 and Case 4. Sets of rival affixes might be *ness* and *ity*, both of which form abstract nouns from adjectives, or *ize* and *ify*, both of which form verbs from either adjectives or nouns. Sets of rival affixes provide us with Cases 3 and 4.

**Case 3** is perhaps the strictest interpretation of morphological blocking. Under this interpretation of blocking we should actually never find perfectly synonymous affixes; that is, there should be no rival affixes.

(3) Rival affixes:  
*ness* and *ity*happiness, purity  
Both suffixes attach to adjectives and create abstract nouns

Whatever affixes mean (and this is not at all a straightforward matter), if blocking or avoidance of absolute synonymy in its strictest form is correct, either pairs of rival affixes should never be attested on the same bases, or if they are, the resulting words should not mean precisely the same thing.

**Case 4** is another case of rival affixes.

(4) *ize* and *ify*civilize, purify; unionize, codify  
Both attach to adjectives or nouns and create verbs

Here, the theory of blocking would tell us that if two affixes do have exactly the same meaning, they should not be found at all on the same root words, or if they do, that the resulting words should have different meanings. So given affixes like *ize* and *ify*, both of which attach to noun or adjective bases and form verbs, we should never find both purize and purify or both unionize and unionify. You’ll see in a moment why I separate Cases 3 and 4.

To summarize so far: both linguists and non-linguists tend to believe that synonymy is avoided in simplex words and that blocking is a theoretical principle that governs the existence and coinage of complex words. Our intuitions lead us comfortably to both the commonsense or theoretical positions.

3. **Oops**
The problem is that our intuitions turn out to be dead wrong in almost every case.

Let’s go through our four cases systematically.

Case 1 concerns inflectional morphology, specifically, the claim that if a noun has an irregular plural in English it cannot have a regular plural, at least not with the same meaning. So for mouse, we assume that one plural goes with the vermin meaning and the other with the computer meaning. But this assumption turns out to be wrong, at least for the case of computer pointing devices:

Good Housekeeping 2004: Thanks to new optical and wireless computer mice, which don't need pads, it's easy to set up your desk so that you won't strain muscles.

Technology for Teaching 2010: Imagine a classroom filled with Finale music notation software users; there are novices as well as experienced professionals, Macintosh and Window users, some with laptops, others with full keyboards and mice.

Business 2000: IBM's program director for mobile market development, Sperano says Bluetooth can eventually be used to disconnect keyboards, mouses and printers, among other things.

CBS_Early 2009: They should be wiped down probably twice a day with -- with alcohol or Clorox. In the hospital, the nurses stations do that every day, mouses, keypads, fax machines, anything that your hands are on.

For the computer pointing device, you find a fairly even split between the regular and the irregular. Notice that you cannot dismiss any of these examples as mistakes; they all come from respected, well-edited publications. And notice as well that we cannot yet argue that one of the uses is likely to supercede the other historically, as both are still in use in this decade. They seem to be coexisting quite nicely.

For the other kind of mouse, what we find in COCA is that the irregular plural is consistent in adult speech for plural of the word for furry little vermin, but the regular plural is the norm when we have more than one Mickey or Minnie. Plural cartoon critters are Mickey Mouses and Minnie Mouses. So even here, both forms are attested, although these are more clearly separated by context.

Case 2 makes the prediction that given the existence of a simplex noun like thief that means person who steals stuff, we would not also coin the word stealer. Here the COCA results are similarly unequivocal. The word stealer is in fact abundantly attested. Not only do we find lots of compound words like base stealer, soul stealer, and child stealer, but we also find a substantial number of examples like these:
Because now he is a thief, and not just any thief, but a **stealer** of dreams and wishes.

I'm imagining Shakespeare, the greatest **stealer** of other people's stories who ever lived, lying on his bed at New Place, smoking his clay pipe full of the latest consignment of hash from the New World, wondering whose work he might plunder and embellish next season at the Globe.

And lest we let our intuitions tell us that we tend to find **stealer** only followed by a prepositional phrase with *of*, as in the two examples in the last slide, and **thief** in contexts not followed by preposition *of*, the following examples should persuade you that that’s not the case:

Or he would have been destroyed. He would have been a drug dealer. He would have been a crack head. He would have been a **stealer**.

John Laroche is a serial monomaniac who learns everything about Ice Age fossils; then chucks fossils for orchids, becomes a celebrated breeder, then a **thief of** wild orchids; and finally turns his back on the plant world and its obsessed hobbyists and dives into computers.

These are of course not isolated examples. It seems that the existence of a simplex word with a particular meaning does not generally preclude the coinage of a complex word with the same meaning.

**Case 3** is somewhat trickier to show. Remember that in this case we were looking at rival affixes, that is two or more affixes that appear to mean precisely the same thing. Blocking tells us that if we find a form with one of the rival affixes, we should not find the other rival affix on the same base. Blocking in its strongest form therefore dictates that if we find a form like *curiosity*, we should not also find a form like *curiousness*. The problem here is that even using intuitions, the word *curiousness* doesn’t seem all that bad, nor does *pureness* next to *purity*. Indeed, if we look in COCA, both *curiousness* and *pureness* are attested alongside the expected *curiosity* and *purity*.

To maintain the theory of blocking linguists are thus forced to retreat to the position that if they find both of the rival affixes on the same base, there must be a difference in meaning. And this is just what was argued in Riddle (1985), for example. Specifically, Riddle argues for a subtle difference between the two suffixes. According to her, whereas *ness* attaches to an adjective that denotes an ‘embodied trait’, *ity* creates forms denoting abstract or concrete entities. What she means by the distinction is not entirely clear, but it seems to hinge on the degree of reification of the quality in question: *ness* forms abstract nouns with a lower degree of reification than *ity*. Riddle gives a number of reasonably persuasive minimal pairs to support this hypothesis. For example, whereas *hyperactivity* names a diagnosable condition, *hyperactiveness* denotes only a property or set of properties that can be displayed by a particular individual at a particular moment. Similarly, *ethnicity* is a reified state of belonging to a
particular group, whereas *ethnicness* denotes only a quality or property that describes a particular group or location (Riddle 1985: 440).

Persuasive though these specific examples are, not all possible *ness* and *ity* pairs seem to display a clear difference; indeed, most do not. *Purity* and *pureness*, for example, are used interchangeably, as are *exclusivity* and *exclusiveness*, *passivity* and *passiveness*, and many more. The examples from COCA are representative:

*New York Times 1998*: He is a fan, and it is the pureness and passion of his enthusiasm that has defined his accomplishment.

*Bicycling 2009*: The purity of having climbed Galibier is almost too much to bear - so raw, so honest that we need to step away, walk off and again convince ourselves that our jobs are important and our world depends, after all, on paying our phone bills and washing our cars and answering our e-mails.

If you’re unconvinced, I refer you to Chapter 12 of Bauer, Lieber & Plag where there are lots more examples. The suffixes *ity* and *ness* are just one rival pair, but in every case where my colleagues and I compared COCA data for sets of rival affixes, we came up with the same thing: there is no such thing as blocking. Apparent morphologically complex synonyms do not block each other.

This brings us to the exception to the rule, which is not actually an exception. Our *Case 4* is the case of the rival verb-forming suffixes *ize* and *ify*. Both attach to adjectives or nouns and result in verbs with the same range of meanings. As in other cases, we ought to expect pairs that share the same base but differ in the suffix used. But in this case we do not. These two suffixes never occur on the same base, and we might be inclined to say that they constitute a real example where use of one suffix blocks the use of the other. But that would be wrong. It is well known that the suffixes *ize* and *ify*, while alike in almost all respects, have different restrictions on the prosodic shape of the bases they can attach to. So *ize* requires bases that display a trochaic or dactylic stress pattern:

(4) trochaic base (stressed-unstressed): random randomize
dactylic base: (stressed-unstressed-unstressed) hospital hospitalize

And the suffix *ify* requires either a monosyllabic base or one with an iambic stress pattern:

(5) monosyllabic base (stressed): pure purify
    iambic base (unstressed-stressed): bourgeois bourgeoisify

I leave it to you as a thought experiment to verify that you can’t substitute *ify* for the first two forms or *ize* for the second two. The upshot here is that if blocking does not really exist, and if *ize* and *ify* are like *ness* and *ity*, we would expect to look in COCA and find pairs with the rival suffixes that mean the same thing. But in this case other things are not equal. The two suffixes carve up the territory of potential bases they could attach to between them on the basis of stress patterns. Any given base will go with one or the other, but not both. But this has
nothing to do with blocking and everything to do with the prosodic requirements of the suffixes themselves.

The overall conclusion we are forced to is that the theoretical principle of blocking is simply wrong, that synonymy exists all over the place, and that our intuitions on the matter are not very good at all.

4. What this tells us about the mental lexicon

What does this tell us about the mental lexicon? Why is the reliance on native-speaker intuition a sound methodological practice for syntacticians but a rather poor methodological practice for morphologists? The answer, of course, lies in the nature of the mental lexicon.

Let me talk a bit about syntax first. Perhaps the most important characteristic of syntax is the fact that native speakers of a language can create infinite numbers of new sentences, and that almost every sentence we create is unique. Aside from a small number of formulaic sentences like *How are you? What’s up?* and the like, we generate new sentences on line and we almost never have need to use individual sentences again. There is no need to store sentences, because by their very nature they are ephemeral.

As native speakers we can also generate infinite numbers of new words, but there is a difference between words and sentences. Whereas individual sentences are rarely if ever repeated, words -- at least some words -- are frequently repeated. We have little need to store sentences in our memory, but we can and do store new words in what is called our mental lexicon. Indeed, psycholinguists have shown that it is possible to find evidence of a memory trace in a speaker’s mental lexicon after as little as one exposure to a word (the evidence for this is complex, so I won’t have time to go into it.)

This has consequences for the principles of avoidance of absolute synonymy and of morphological blocking. Both synonymy avoidance and blocking presuppose that the mental lexicon has limits with respect to storage. Specifically, they are based on the assumption that the mental lexicon is organized a bit like a pegboard: think of each meaning as a hole and each word as a peg. Obviously once you’ve filled a hole, you can’t insert another peg into the same hole. But what the use of corpora shows us is that the mental lexicon is not at all like a pegboard. It is far more expansive and far less rigid than the pegboard analogy would suggest: you may have words that are characteristically associated with a specific meaning, but that doesn’t preclude coining a word with precisely the same meaning.

So why are our intuitions so good in the case of sentences and so bad in the case of words? This very likely has to do with the interplay of the effects of frequency and what’s called lexical access. The more frequently a word is encountered, the easier it is to access that word and pull it out of our mental lexicons when we need it. And the easier the access is to a word, the more natural it sounds, especially out of a specific context. If we rely on our intuitions -- for example, asking ourselves if *stealer* sounds like an acceptable word -- we are inclined to access the more frequently encountered word *thief* and decide that the infrequently heard *stealer* is rather odd in comparison. But if we’re writing or speaking on line we may nevertheless produce *stealer* and
in context it will seem perfectly natural and unremarkable. One reason we might do this is that lexical access to the more frequent word is temporarily unavailable – think of this as a lexical senior moment. Another is that some speakers may not have a word stored for a complex meaning and may generate one using the word formation rules of English – and precisely because we have multiple affixes with the same meaning, might come up with a form that is not the conventional one.

5. Conclusions

- Words are not like sentences: while our intuitions about the grammaticality or acceptability of sentences might be sound, our intuitions about the acceptability of words can lead us sorely astray.

- The only way to discover this is by embracing the use of vast databases of language.

- Contrary to thirty years of morphological theorizing, we find that native speakers of English, and presumably of other languages, do not avoid synonymy, and that there is no such thing as blocking, in any of its various forms.

References


