

# Lexeme frequency and intercultural communication

**Philip Spaelti**

Department of English, Kobe Shoin Women's University

spaelti[at]shoin.ac.jp

---

## Abstract

This paper considers the question of whether languages have more words for culturally important ideas than for trivial ones, and the effects this has on intercultural communication. Using examples from Germanic, including English as well as from Oceanic, I consider how these languages communicate important cultural concepts. I argue that it is necessary to distinguish cases of specialization which do lead to higher number of lexemes for important concepts, from cases where dialect variation of a local nature leads to higher number of lexemes for marginal concepts. An example of the latter is seen in English which has words like griddle cake, pancake, hotcake, all with essentially the same meaning.

## 語の使用頻度と異文化間コミュニケーション

異文化間コミュニケーションの見地から、その文化で重要な概念を表す語とそうではない語を比べ、前者の方が後者よりも多く存在し、頻繁に使用されるかという問題を考察し、この違いが異文化間コミュニケーションにどのような結果をもたらすかを議論する。英語を含むゲルマン語およびオセアニア語の実例を用い、これらの言語においてその文化および異文化間で重要な概念がコミュニケーションで用いられるあり方について考え、異文化間コミュニケーションの見地から比較検討する。その文化で重要な概念に相当する語が特殊化によってその数が増えるケースと、重要性の低い概念を表す語が方言の多様性によってその数が増えるケースを区別する必要があることを主張する。英語で griddle cake, pancake, hotcake のように同じ物（概念）を指す多様な語が存在することは後者のケースの一例である。

**Keywords:** lexeme frequency, culturally important concepts, Oceanic, Germanic, Sapir-Whorf hypothesis

## 1. Introduction

A commonly held idea is that languages should have many words for important cultural concepts. There is certainly a sense in which this is true, but it should also be said that this is quite unremarkable. Anyone who has tried to engage in any reasonably complex task, such

as taking apart a mechanical clock or the like, will probably have found that expressions such as *thing-a-ma-bob* and *whatcha-ma-call-it* quickly wear thin. During the days when English speaking people used horses for locomotion and as their primary source of literal horse power, the language had an abundance of equine vocabulary which has left its traces in now nearly opaque expressions such “give free rein”, and “champing at the bit”.

In this essay I would like to consider this question a bit more critically. To lighten the discussion I would like to begin with an observation by the late-great George Carlin formulated in his signature clever style:

Griddle cakes, pancakes, hotcakes, flapjacks: why are there four names for grilled batter and only one word for love?  
(Carlin, 2002)

One hates to spoil a good joke, but perhaps this one merits it. Underlying the joke is a view apparently held by many people; languages should have more words for important ideas than for trivial ones.

## 2. What do we mean when we say “many words”

When we consider this idea the first problem that we encounter is what it means to say “more words”. On the face of it this would seem to be a trivial question. But note that even Carlin is careful to say “names” rather than words. No doubt his thinking was that *griddle cakes* is one “name” even if it looks like two words. But why is it that *pancake* and *hotcake* are one word each, while *griddle cake* is two? Surely this would just seem to be an accident of orthography.

With a bit more reflection we can see that this accident is just the tip of the iceberg. The question of whether *griddle cake* should count as one word or two is an example of the several conflicting definitions of *word* in English: Orthographic, Phonological, Semantic or Grammatical. (Murphy, 2010) It is a general pattern of English—in contrast to German—that compound words such as *griddle cake* are spelled as separate orthographic words, though as *pancake* shows, there are exceptions. And while there are good arguments for including such compound words in our definition of *word*, to do so is to quickly lose any hope of narrowing down the question of how many words English has, since the list of compound words is in principle infinite.

The standard solution to this problem is to narrow our sight to *Lexemes*, the entries in our mental dictionary. The appeal of this solution is that words in this sense constitute the foundation of our linguistic knowledge. The first step to learning a language is to learn the words, i.e., the lexemes, and this set must be finite or we could never hope to become competent.

Let us return to the question posed earlier: Is it in any sense true that languages have more words for important ideas? This reminds us of a famous linguistic trope, oft repeated and displayed as an interesting insight that Eskimos have “hundreds” of words for snow. The mythical argument seems to run something like this: Eskimos live in the Arctic. The Arctic is covered by snow for large parts of the year. The Eskimos live and hunt in the snow for much of the year. They even build houses out of the stuff. Therefore snow must figure highly in their life. Therefore it must be important in their culture. Therefore they talk about snow often. They perceive its many variations. They must have many words for it.

Before we continue it must be made clear that the central factoid of this “insight” has a long history of misrepresentation and exaggeration exhibiting an intellectual laziness, that was memorably and hilariously exposed by Pullum (1991). It’s nevertheless clear that Eskimo—or perhaps more accurately Central Alaskan Yupik—does have more than one word for snow. And thus the question has more to do with the hold that this piece of information has, or has

had, on people's imagination: Is it perhaps due, at least in part, to the seeming elegance of the train of thought outlined above. The train of thought has just the right complexity. It isn't completely obvious, but it's just easy enough for anyone to figure out on their own. When presented with the factoid: "Eskimos have X words for snow," people can quickly accommodate the logic and the factoid 'clicks'. "Oh, of course! That makes so much sense!"

This is the question that I will be trying to narrow down below. Is there any justifiable basis for the idea that we should expect languages to have many words for important concepts? And, if so, under what circumstances?

Before continuing I would like to briefly mention one more point. The seed for this perhaps most famous version of the question of interest is Benjamin Whorf's article "Science and Linguistics" (Whorf, 1940). But if so, we might note that this particular example of Whorf's is now popularly known for the wrong reasons. Whorf's argument didn't really have anything to do with the *number* of words. Rather his concern was with the slicing of the semantic pie. In English a single concept—H<sub>2</sub>O in its crystalline form—a single word. In Eskimo the world sliced into different semantic bits—white fluffy bits falling from the sky vs. white background noise covering the horizon—two concepts, two words. So what Whorf was really after was: different world view, different concepts. But what the popular opinion seems to have taken from this was something else: *important concepts, many words*.

### 3. Important concepts, many words? Two examples

I will now turn to the central question of this essay, whether we should expect languages to have many words for important concepts. To shed light on this question I will consider two examples in more detail.

#### Example 1: The seahorse in Oceanic

The Austronesian language family is by some measures the largest language family on the globe with approximately 1,200 languages spoken over an area stretching two-thirds of the way around the earth, from Madagascar in the West all the way to Easter Island in the East. Research has shown the family originated in China, from where the speakers crossed the strait and settled Taiwan. Then after developing the ability to navigate long distances across the open ocean, Austronesian peoples began moving from island to island, expanding their territory. In the process they developed an ocean-based culture, carrying in their boats plants and animals that could withstand the rigors of travel, and spreading many of them in the process.

Researchers investigating these languages have assembled a trove of over 2,200 lexical items, which reflect the language and culture of these people. As they spread outwards into the islands of the Pacific the number of ocean related terms increased, with terms relating to boats and sailing, all kinds of ocean life, and over one hundred different types of fish.

Along the way the Austronesian people developed a rich culture based on the techniques and plants that allowed them to flourish in this environment. These techniques include the catamaran and the crab claw sail, while plants include many, in particular, breadfruit, pandanus, taro, bananas, and coconut. About the coconut palm, (Ross, Pawley, & Osmond, 2011b) explain:

Just about every part of the palm and the fruit is used in traditional societies, *and many of these parts are named*. Because of the salience of the coconut in Oceanic cultures and because different uses are made of it at different stages of its growth, it is common *for the fruit and palm to be given different names at different growth stages* (Ross, Pawley, & Osmond, 2011b). [emphasis added]

So while the coconut is important, and is the store for a large amount of specialized vocabulary, the coconut itself is generally referred to by just one term, usually a reflex of the Proto-Oceanic root \*niuR.

While such cases of culturally relevant variety are interesting, just as telling is what is absent. In popular imagination indigenous peoples are often assumed to be “close to nature” and practically omniscient about their environment. It is therefore interesting to find that there are notable creatures for which names cannot reliably be found. The seahorse *Hippocampus* is, to Western eyes at least, a striking fish, with its upright posture and curly prehensile tail. However, only few of the Oceanic daughter languages seem to have any kind of name for this species, and no term can be reconstructed for the proto-language. Why did this fascinating animal go unnamed? As noted by the authors of the *Lexicon of Proto-Oceanic* (Ross, Pawley, & Osmond, 2011a):

The seahorse probably did not hold much interest for early Oceanic speakers, *being neither economically useful nor dangerous*. Collected terms for it are rare, as evidenced by responses to a request from Bruce Biggs on the AN-LANG e-mail list in 1999. [emphasis added]

This is a point that will often be surprising to people from cultures with written languages. In traditional cultures, words can drop out of use and be forgotten. It isn't that we should imagine that these people were not aware of the seahorse, but merely that even on the occasion that a name was coined for it, conversations about it were infrequent enough that such a name would soon drop out of use again. Then at the next opportunity a new term might need to be coined ‘on the fly’, only to fall out of use yet again. It is this lack of a continuing habit of usage that leads to such gaps in the lexicon.

In Europe, *Hippocampus*' snout and upright posture reminded people of a horse—a culturally highly significant item. Meanwhile, in East Asia, a perceived resemblance to the also culturally significant mythical dragon even led to its being exploited for imagined health benefits. But just as important, in cultures with a written language one can assemble lists, at which point people develop the habit of “consulting the dictionary”, relying when in doubt on the authority of the written word for the “correct” terms, and thus making the loss of even infrequent words much less likely.

### **Example 2: The butterfly and the bee**

Another example which shows how items are more likely to receive consistent names only if they are either economically useful or dangerous, can be seen in the case of the butterfly and the bee.

The butterfly and the bee are both members of the class *Insecta* and both are important pollinators, feeding on the nectar of flowering plants. However their relation to humans is quite different, and this is reflected in the names we give them.

Butterflies are striking creatures often with highly visible aesthetic appearance, and a fascinating life cycle. They also move about in a relatively leisurely manner. Linguistically speaking it has been noted that they are referred to by a great variety of names. This idea seems to appeal to many people and it is not uncommon to find lists displaying this variety. Here for example is a list showing this variety among the Germanic languages:

English	butterfly
Dutch	vlinder
German	Schmetterling
Danish	sommerfugl
Norwegian	sommerfugl
Swedish	fjäril
Icelandic	fiðrildi

In this list the identical terms used in Danish and Norwegian are most likely due to the fact that Danish served as the written standard for Norway until the nineteenth century. Apart from that, what is most striking is how dissimilar these terms are.

It should be noted that this variety may be a bit overstated, or, depending on one's viewpoint, understated. The German form *Schmetterling*, which is superficially very different from the English term, is in fact from a dialectal form of German, where *Schmetter-* refers to 'fat, butter', thus making this form more closely parallel to the English expression. (Dudenredaktion (Bibliographisches Institut), 2001) This name is ostensibly given due to the insect's propensity for landing on greasy foods. And while this name is the most common expression used to refer to the butterfly in Standard German, other forms are known as well. Southern Germans and Swiss people are likely to know and use *Sommervogel*, which can be transparently analyzed as "Summer bird" presumably due to the butterfly only appearing during the warm part of the year. This name is clearly similar or identical to the one used in Danish/Norwegian. Yet another name is *Falter*, historically also from a form with doubling *Fifalter*, based on the verb *flattern* 'flap', parallel to and/or calqued from Latin *papilio*.

And in fact one could extend this list using yet more languages. French has *papillon*, which is obviously from its parent language Latin, while Italian has the similar, but also strikingly different *farfalla*, while Spanish calls the same *mariposa*, and Portuguese *borboleta*.

Of course all of these forms come from different languages, so shouldn't we expect them to be different? Maybe. But it will be instructive to compare this with the case of the terms in use among the Germanic languages for the bee (University of Texas at Austin).

Old English:	beó, bī, beón
Middle English:	bee, been
English:	bee
Old Dutch:	bie
Dutch:	bij(e)
Low German:	bigge
Old High German:	bī(n)a, pīa
Middle High German:	bīn, bīe
German:	Biene, Beie
Old Norse:	bý
Icelandic:	bý
Danish:	bi
Swedish:	bi

These terms show stability across great periods of time. In fact arguably this can and has been extended back even further leading to the reconstructed Proto-Indo-European form *\*bhei-* 'bee (insect)' (University of Texas at Austin). We are left with the conclusion that people have been using the same form to refer to this insect in an unbroken tradition stretching back thousands of years.

So why this difference? Well, as was noted in the earlier discussion of Austronesian, names for things that are economically useful and/or dangerous are likely to be stable. In this case, the bee is both. The bee has been pursued and cultivated for centuries to take advantage of the honey that it accumulates. At the same time bees are dangerous. Their stings are at the very least painful, and on occasion even fatal. The butterfly is neither. This would seem to indicate that the great profusion of names for the butterfly, while exciting the romantic imagination of many, is simply the result of its name being periodically ‘forgotten’ and later re-coined or re-borrowed.

Which brings us back to Carlin and his griddle cakes. Surely this case has more in common with the example set by the butterfly, then as a marker of great cultural significance. The variety of terms was probably born through the repeated re-coining of new terms, perhaps as regionalisms. Most people are likely to stick to their favorite from among these assorted expressions, and will probably not have extended conversations on the topic, unless they are employed in their local “serve breakfast all day” restaurant.

#### 4. Summary

To summarize: I would like to propose that frequency of lexemes

- Words referring to concepts of cultural significance are likely to be stable over long periods of time
- As a general rule a large set of synonyms is not an indication that a concept is culturally important.
- When a culture has a variety of terms for a single concept with high dialect variation of a local nature, this is likely due to its relative *unimportance*.

#### References

- Carlin, G. (2002). *Napalm & Silly Putty*. Hyperion.
- Dudenredaktion (Bibliographisches Institut). (2001). *Duden Herkunftswörterbuch*. Mannheim: Dudenverlag.
- Murphy, M. L. (2010). *Lexical Meaning*. Cambridge University Press.
- Pullum, G. K. (1991). *The Great Eskimo Vocabulary Hoax, and other irreverent essays on the study of language*. University Of Chicago Press.
- Ross, M., Pawley, A., & Osmond, M. (2011a). *The Lexicon of Proto Oceanic: Animals* (4). Pacific Linguistics.
- Ross, M., Pawley, A., & Osmond, M. (2011b). *The Lexicon of Proto Oceanic: Plants* (3). University of Texas at Austin, L. R. C. Indo-European Lexicon. Retrieved 2021-02-18, from <https://liberalarts.utexas.edu/lrc/resources/ielex.php>
- Whorf, B. L. (1940). Science and linguistics. *Technology Review*, 42, 229-231.

(Received: January 10, 2021)