

Re: where do so many tenses come from?

Re: where do so many tenses come from?

Source: <http://sci.tech-archive.net/Archive/sci.lang/2006-04/msg00184.html>

- *From:* hrubin@xxxxxxxxxxxxxxxxxxxx (Herman Rubin)
 - *Date:* 3 Apr 2006 13:40:48 -0400
-

In article <874q1a7m8c.fsf@xxxxxxxxxxxxxxxxxxxxxxxxxxxx>, Lee Sau Dan <danlee@xxxxxxxxxxxxxxxxxxxxxxxxxxxx> wrote:

"Joachim"
== Joachim
Pense
<spam-collector@xxxxxxxxxxxxxxxxxxxx>
writes:

Joachim> Information theory gives the standard definition of
Joachim> "information". The information of a sign (in this case a
Joachim> phoneme) within a sequence of signs is the negative
Joachim> dyadic logarithm of the probability that this sign occurs
Joachim> in this position.

Well... you have to get these probabilities first.

That is if you want to compute the information. However, the amount of information exists, even if we do not know how to compute it.

Joachim> You can describe it as the number of bits you need to
Joachim> encode it if your coding system is optimized to produce
Joachim> optimally short bit strings for the data source
Joachim> (=language in our case).

What a big IF!

That is not a big if, and is SLIGHTLY wrong. But it gives the general idea.

Re: where do so many tenses come from?

Re: where do so many tenses come from?

Joachim> So if there are less signs to choose from, then for each
Joachim> the probability to occur in any given position is higher
Joachim> (on average), so the information is less.

On average? Does this average translate to an assumption that the probabilities are more or less EVENLY DISTRIBUTED among every sign? I'm afraid this assumption isn't that valid. Take English as an example. Does the phoneme /T/ occur almost as frequently as /i/?

It makes no such assumption. There is no requirement that each sign gets the same length of code. And the best encoding would use encodings for major groups of signs, not for single signs. Computer compression algorithms do not just attack single signs.

But it is still true that in general the larger the alphabet the fewer characters will be needed. For speech, the alphabet consists of the phonemes, or phoneme combinations. The most efficient known oral method of communication is by pitch, for those with perfect pitch. It has occasionally been used for communication, but not often.

Joachim> I guess this definition encloses both syntagmatics and
Joachim> paradigmatics.

This address is for information only. I do not claim that these views are those of the Statistics Department or of Purdue University.
Herman Rubin, Department of Statistics, Purdue University
hrubin@xxxxxxxxxxxxxxxxx Phone: (765)494-6054 FAX: (765)494-0558

.

Re: where do so many tenses come from?