12-1-2013

Dialect or a New Language: A Possible Explanation of the 70% Mutual Intelligibility Threshold

Olga Kosheleva
University of Texas at El Paso, olgak@utep.edu

Vladik Kreinovich
University of Texas at El Paso, vladik@utep.edu

Dialect or a New Language: A Possible Explanation of the 70% Mutual Intelligibility Threshold

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract

In most cases, linguists have a consensus on when people from different regions speak two different dialects of the same language (and can, thus, understand each other reasonably well) or two different languages (in this case, their mutual intelligibility is limited). In most cases, this intuitive consensus corresponds to a 70% mutual intelligibility threshold: if at least 70% of the words from one region are understandable to people from another region, then these are two dialects, otherwise these are two different languages. In this paper, we provide a possible explanation for this 70% threshold.

1 Formulation of the Problem

A dialect or a new language? People in different regions often talk somewhat differently. When this difference is reasonably small and we can still more or less understand each other, we say that these are two dialects of the same language. On the other hand, when communication is difficult, we say that people from these two regions speak different languages.

A formal description of the difference: an empirical 70% threshold. In most cases, linguists reach a consensus on when the way people talk in a new region is a dialect or a new language. It turns out that in most cases, this consensus can be expressed in precise terms (see, e.g., [1] and references therein):

- if 70% or more words from one region are understandable to speakers from the other region, then it is usually a dialect;

- if less than 70% of words are mutually intelligible, then these are usually different languages.

This formal definition is not a universal recipe:

- sometimes, in spite of the mutual intelligibility above 70%, linguists still talk about different languages – this is the case, e.g., for Italian and Spanish;

- in other cases, mutual intelligibility is below 70%, but most linguists still consider this situation dialects of the same language – this is the case, for Chinese.

However, these exceptions notwithstanding, the 70% threshold provides the best possible match for the intuitive understanding of the difference between a dialect and a new language.

Comment. Some linguists use a more strict criterion for classifying two regions as speaking the same language; this stricter criterion corresponds to the 80% threshold.

Problem. Why is 70% (and not any other number) a threshold corresponding to the intuitive distinction between dialects and languages? In this paper, we provide a possible explanation for this empirical threshold.

2 A Possible Solution to the Problem

How to gauge mutual intelligibility. A reasonable way to check intelligibility is to check how well a person from one region can follow instructions issued by a person from another region. If, in general, a person succeeds in following these instructions, this means that we have, in effect, the same language; if this success is not guaranteed, this is probably an indication that people from the two regions speak different languages.

How to gauge mutual intelligibility: a simplified model. Let us describe how the above idea can formalized in the simplest possible way.

In the state space, following instructions means going from the original state A to the desired state B. In general, a state is described by several parameters; as a result, the corresponding state space is multi-dimensional. The simplest situation is when the state is 1-dimensional. In this simplest case, at each moment of time, we either stay in place or more in one direction. For simplicity, we will assume that each movement is either one step to the left or one step to the right.

If B is to the right of A, then correct instructions recommend that the person moves to the right every time. Similarly, if B is to the left of A, then correct instructions recommend that the person moves to the left every time.

Analysis of the simplified model. If 50% or more of the instructions are understood correctly, then, even if the rest of the instructions are misunderstood, the person still moves in the right direction more than half of the time. Thus, overall, the person is moving to the right (i.e., in the right direction). As a result, eventually, the person will reach the desired state B (although, of course, due to possible misunderstandings, this will take longer than in the case when all instructions are properly understood).

Vice versa, if 50% or fewer of the instructions are understood correctly, it may be that all other instructions are misunderstood as going in the wrong direction. In this case, the person is moving to the left, i.e., in the wrong direction.

Conclusions from the above analysis. Based on the above analysis, we can conclude that:

- we have a dialect is at least 50% of instructions are correctly understood, and

- we have a new language is 50% or less of the instructions are correctly understood.

Why 70%? A possible explanation. The simplest instructions consist of two words, like "go right", "grab a sword", etc. If p is the percentage of words which are correctly understood, then for each word, the probability to correctly understand this word is equal to p.

Thus, the probability that both words in a given instruction are correctly understood is equal to p^2. In other words, the fraction p^2 of the instructions will be understood correctly.

According to our analysis, a dialect corresponds to the case when $p^2 > 0.5$. This inequality is equivalent to $p > p_2 \stackrel{\text{def}}{=} \sqrt{0.5}$. The corresponding threshold value p_2 is approximately equal to $0.71 \approx 0.7$. Thus, we indeed have an explanation of why 70% is a threshold separating dialects from languages.

Why 80%? A possible explanation. Similar arguments can explain the stricter 80% threshold proposed by some linguists. Indeed, in case of more complex three-word instructions, the probability of understanding an instruction correctly is equal to p^3. A similar inequality $p^3 > 0.5$ is equivalent to $p > p_3 \stackrel{\text{def}}{=} \sqrt[3]{0.5} \approx 0.8$. This indeed explains the 80% threshold.

3

References

[1] J. Diamond, *The World Until Yesterday: What Can We Learn from Traditional Societies?*, Viking, New York, 2012.