Clary Binns

SHEWASSA on Language Comprehension

Human Biology and Decision-Making, June 2017

Abstract

The purpose of this experiment was to see whether and how a person's linguistic background - how many languages they speak and how well - affects their ability to learn a new language through extrapolation from examples versus through a guide to the necessary grammar and vocabulary. The hypothesis was that there would be no correlation between multilingualism and comprehension in the group that received the guide and positive correlation in the examples group. A language, Orkaki, was invented and a test created that focused on a few rules and limited vocabulary of the language. This test was given in a survey format to the participants along with either the list of example sentences or the guide. The participants were asked to translate terms and sentences to and from Orkaki. The results showed that most participants did well; the participants given the guide did significantly better than those given the examples (at a p value of 0.0007); and that, in both groups, high multilingualism scores resulted in high comprehension scores. However, lower multilingual ability resulted in varying levels of comprehension on the test, indicating a lack of correlation (with r values of 0.033 and 0.148). Therefore, the null hypothesis - no correlation - cannot be rejected. Further investigation is necessary to discover what other variables may influence the pattern, possibly by including age, level of education, and level in mathematics in the survey.

Introduction

I am fascinated by everything linguistic: sentence structure; the development, construction, and evolution of languages; etymology; grammar; the neurology of language learning, speaking, and writing; the effect of language on society and vice versa; ancient languages; and probably whole other realms of linguistic phenomenons about which I have yet to learn. I adore learning about languages because of the complex patterns that lie in wait beneath the seemingly random surface. As such, I am curious about how we learn to identify and use these patterns in order to learn new languages. My question was therefore how well different people are able to perform this process of connection based on their linguistic backgrounds, since being practiced in recognising a language's patterns should enable us to do it more readily with new languages. This question has undergone a number of revisions as I have tried to find one that is reasonably answerable with a SHEWASSA, but in its current form, it is: Does multilingualism improve our ability to recognize patterns in an unfamiliar language and therefore to learn that language from examples? My hypothesis was that people who speak a variety of languages on a regular basis would be able to recognise patterns in a new, unfamiliar language better than those who speak fewer and thus learn the language more quickly by example. My null hypothesis was that there would be no

correlation between multilingualism and the ability to comprehend this new language from examples. Since I did not want the results to be influenced by people's ability to recognise cognates or familiarity with romance languages, I could not use a preexisting romance language, like Italian, or a latin-based language like Esperanto. These more common languages also run the risk of my participants already knowing them, and I did not want to narrow my sample size. Other constructed languages are either too simple to be useful or too complex to learn. The language I created in order to perform this test is called Orkaki, and it is based on roots and their modification to form different parts of speech (like Esperanto), but is much simpler than most languages and so should be easy enough to understand.

Experimental Design

In my experiment, I presented people with two methods of learning Orkaki: through a dictionary of sorts with grammatical rules and basic roots/vocabulary, or a list of example sentences and vocabulary (translated into English for them) that allow subjects to extrapolate the same rules. This first method (the dictionary) should serve as a sort of baseline to which to compare the first, though it's not quite a control. I then tested their knowledge with a Comprehension Test and asked them how many languages they are learning and at what levels they are in each language. The variables here are as follows:

- Comprehension score (numerical)
- Multilingualism score (numerical)
- Dictionary vs. Example approach (categorical)

My materials were two rather similar surveys, composed of:
- Either:
  - The dictionary - including subjects, how to conjugate verbs in the present indicative, how to form nouns, how to form plurals, how to form questions, and essential verbs (Appendix A).
  - The examples - simple sentences and questions in the present indicative which involve all of the above concepts (Appendix B).
- The comprehension test, composed of translating 3 nouns into Orkaki, 3 short sentences from Orkaki into English, and 4 short sentences from English into Orkaki (Appendix C).
- The rating of language ability, in which participants are asked to report all of the languages they speak on a regular basis and mark how well on a gradient they speak them (5 being natively, 4 being fluently, 3 being competently, 2 being intermediately, 1 being barely - the exact language is in Appendix D).
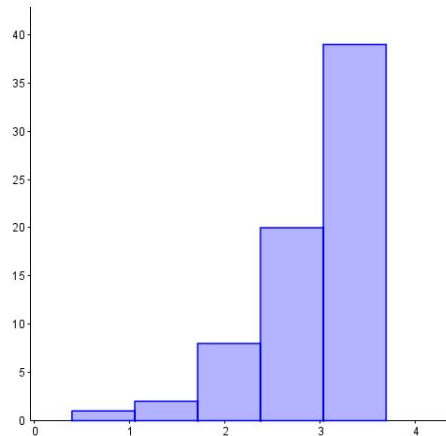
The survey tested a few specific rules: noun formation (from verbs), pluralisation, conjugation, and question formation with a small selection of vocabulary.

Procedure:
- Participants received the survey; links to them were spread on social media and, through email, to the Meridian community, in order to gather a wider diversity of population (in age and number of languages spoken). Both links were present in the online mediums, and participants were asked to select one based on the first letter of their last name (A-L received the examples and M-Z the guide). For student participants in Meridian Academy, the links were sent to them via email (divided randomly) and they were asked to fill out the survey, receiving a starburst upon completion. The survey was in two pages, the first being the preparation (dictionary or examples) and the comprehension test, and the second being the multilingualism questions. Google Forms was used to create and process the survey.
- Each Multilingualism Score (MS) was calculated by the sum of their scores of the languages they speak.
- Their Comprehension Score (CS) was ranked for each sentence in the comprehension test based on closeness to the correct answer, with 5 being exact and 0 being completely off; the single words were scored as correct (1), only partially correct (½), or wrong (0). These scores were then compiled into three means: the mean of the scores of the sentences translated to English, the sentences translated to Orkaki, and the terms. The mean of these three means was taken to form their total CS. The rubric used to score the various elements is in Appendix E.
- The data were then analysed using the following methods:
    - An unpaired, two-tailed T-Test on the difference of means between the rule and example groups.
    - Regression tests on both data sets individually (the independent variable being MS and the dependent being CS), and these two models were compared to see if there is any difference.
        - Regression tests for individual elements of the CS were also run and compared; these elements are, for both the control and the experimental:
            - The vocabulary terms,
            - The sentences translated to English,
            - The sentences translated from Orkaki, and
            - Each individual sentence.

Results

The link to a spreadsheet of the data can be found in Appendix F. In total, 74 responses were received from the survey. The range of multilingualism scores went from 5 to 17, with a mean of 9 and standard deviation of about 2.6 (and a rather unremarkable normal distribution), and this pattern was consistent across both test groups. Comprehension scores ranged from 0.4 to 3.7; their mean was about 3, and standard deviation about .7. However, the distribution was negatively skewed, meaning it was more common to have a higher score, as seen in the histogram to the left. The negative skew can be accounted for by the impossibility of receiving a score higher than 3.7, but the fact that the highest possible score was the mode at all is in itself unexpected. Most people should have an average score, which would result in symmetrical distribution, but more people had higher scores, resulting in this skewed distribution.
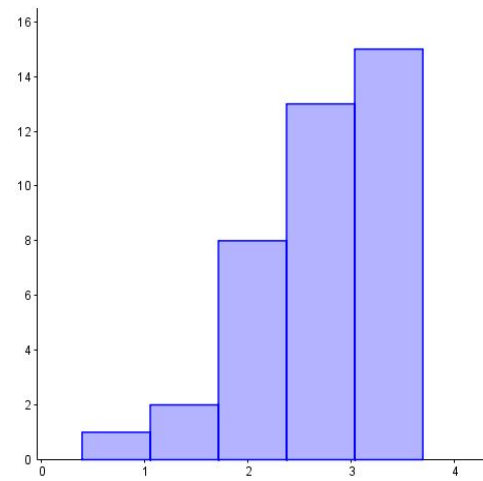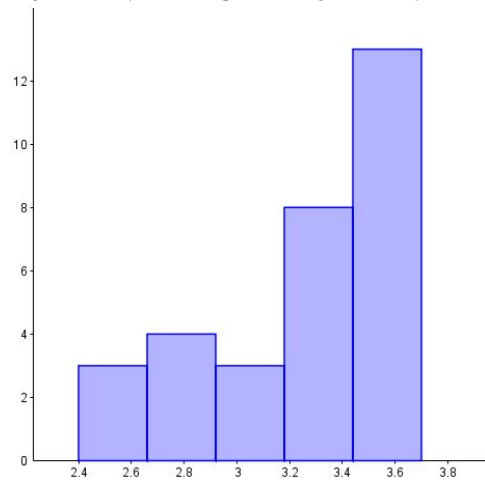

On the left are the results from those participants who received the guide, and on the right the results from those who received the examples.

These histograms are fairly similar, both showing negatively-skewed distribution, but the one for the guide group has a higher minimum (the scale does not show this difference very well), at 2.4, and an overall smaller range (2.4 - 3.7), with a standard deviation from the mean of only 0.4. The example group's range is 0.4 to 3.7 and its standard deviation is 0.8, twice that of the guide group. It should be noted that the example group contains 8 more data points than the guide group, which may have affected these ranges. When the two groups are compared in a t-test, the means (3.3 for the guide group and 2.8 for the example group) have a meaningful difference, with a p value of 0.0007. All these statistics

demonstrate that the people receiving the guide did significantly better on average than those who received the list of the examples. The negative skew also indicates that people were more consistent in their range of scores, and that people were most likely to do well in each group and overall.

Below are regression tests for multilingualism versus comprehension for both groups (guide on the left and examples on the right). None of the other regression tests displayed different patterns, so the test for this variable should suffice as a summary.
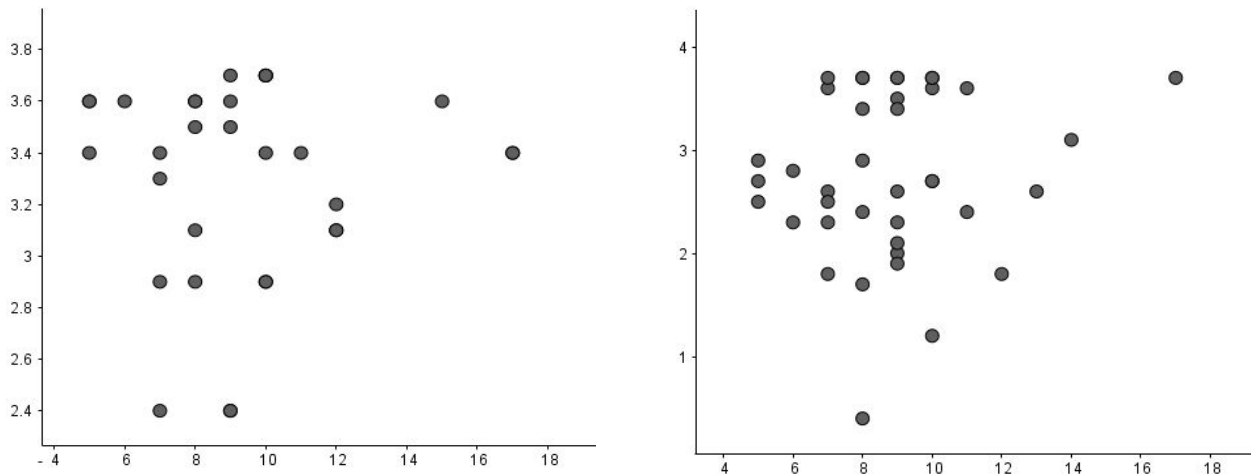
Neither of these data sets have significant correlation coefficients (.033 and .148), so it is clear that neither have a linear correlation, or, based on observation, any other type of correlation. Both graphs, and particularly the results from the example group, have a similar trend of lower multilingualism scores resulting in a variety of comprehension scores, and higher multilingualism scores resulting in only higher comprehension scores. This latter element could be explained by the theory that speaking more languages does indeed assist comprehension of the new one. If this idea is followed, the phenomenon that lower multilingualism scores displayed no correlation whatsoever to lower comprehension scores suggests that there may be some other variable in play that increases ability along with language speaking. It is possible that this variable is level of ability in mathematics, as maths require the same type of analytical thinking and pattern recognition as language learning does. There are also other variables that could have affected the data, such as the age of the participants, which was not included in the survey, or their conditions when taking the survey, as that could not be controlled in the online format. Therefore, while we cannot reject the null hypothesis (no correlation), there is certainly room for further investigation into the factors causing this pattern to emerge.

Conclusion

It is clear that the group that received the guide to Orkaki had higher, more consistent scores, but also that both groups tended towards higher scores over lower ones, which is somewhat surprising as the

test was quite challenging. The data show no correlation between low multilingualism scores and low comprehension scores, which indicates that the alternative hypothesis was unsupported, but they do seem to display a correlation between high multilingualism scores and high comprehension scores, as a sort of lower limit emerges in a positive linear fashion. This trend is clearer in the group that received examples than the guide group, but still present in both. It indicates that there may be another variable affecting the results, such as age, level of education generally, or, possibly most interestingly, ability in maths, which is a form of language. Going forward, therefore, I should like to repeat the experiment and collect data on age and education as well as mathematical ability to include in the multilingualism score and see if correlation does emerge more strongly.

APPENDIX A: GUIDE TO ORKAKI

Subject pronouns:

- I - a
- We - aki
- You - i
- She/He - e
- They (for animals or other living/personified objects) - eki

Grammar:

- All verbs are conjugated by adding the subject's pronoun to the end of the infinitive.
- Placing a specific noun directly after a conjugated verb (combined into the same word) makes it the subject.
- Placing "u" at the end of a verb makes it a noun.
- Placing "ki" at the end of a noun makes it plural.
- Placing "sa" at the end of a sentence makes it a question.

Vocabulary:

- Ork - to speak, talk, say
- Suk - to build
- Rok - to bake
- Lak - to like
- Tok - to give

APPENDIX B: PHRASES AND SENTENCES IN ORKAKI

roka
I bake.

laki rok sa
Do you like to bake?

laki roku sa
Do you like cake?

lakaki rokuki
We like cakes.

sukeki suku
They build a house.

tokaki rokuki
We give cakes.

roke rokuki
He/She bakes cakes.

roki roku sa
Do you bake a cake?

laka ork
I like to talk.

orkaki
We talk.

suke roku
He/She builds a cake.

lakeki suk sa
Do they like to build?

toka roku
I give cake.

tokeki sa
Do they give?

APPENDIX C: COMPREHENSION TEST

How do you say "speech" in Orkaki?
Answer: orku

How do you say "houses" in Orkaki?
Answer: sukuki

How do you say "gift" in Orkaki?
Answer: toku

Translate the following sentences into English.

Suka sukuki
Answer: I build houses.

lakeki rokuki sa
Answer: Do they like cakes?

toki orku
Answer: You give (a) speech.

Translate the following sentences into Orkaki.

She bakes.
Answer: roke

Do they talk?
Answer: orkeki sa

We like gifts.
Answer: lakaki tokuki.

Do you like to build?
Answer: laki suk sa

APPENDIX D: TEXT FROM SURVEYS

Introduction to Test for Rules Group:

"Hi! Thank you so much for agreeing to help me in my experiment. Please make sure you're in quiet surroundings in which you can focus, and this should only take 10 minutes. Study the following guide to Orkaki, a constructed language, for a few minutes and then go ahead and answer the questions based on the rules you've learned. You can look back at the guide as much as is necessary."

Introduction to Test for Examples Group:

"Hi! Thank you so much for agreeing to help me in my experiment. Please make sure you're in quiet surroundings in which you can focus, and this should only take 10 minutes. Study these example sentences in Orkaki, a constructed language, for a few minutes and then go ahead and answer the questions based on the patterns you observe. You can look back at the guide as much as is necessary."

Multilingualism Determination:

"Please rank each language you speak ON A REGULAR BASIS (being in a class counts, as does speaking it at home). 1 means you're just learning it (basic grasp, first year of Spanish, for example), 2 means you're intermediate (could carry out a rather broken conversation, Spanish NH-I1), 3 means you're competent (could carry out a conversation, if slowly - Spanish I2 or Advanced), 4 means you're fluent (could keep up a conversation with a native speaker), and 5 means you're a native speaker."

APPENDIX E: RUBRIC

Criteria	Score
Terms	
Exactly Correct	1
Right idea on the grammar front, but incorrect root OR correct root, but incorrect modification or not smushed	1/2
Totally off	0
Sentences to English	
Exactly correct	5
Mostly right, but one small error (element misunderstood, incorrect pluralisation, order slightly off, or form of word wrong)	4
Got the jist, but medium-sized error or multiple small errors (incorrect conjugation, adding an element or getting one wrong, misunderstanding a few elements slightly)	3
Some of the basic jist, but incorrect grammar or elements	2
One element correct (like a root, or verb conjugated correctly, but the other parts are off)	1
Totally off	0
Sentences to Orkaki	
Exactly correct	5
Very close, but minor error (incorrect pluralisation, words not smushed together, extra or missing letter, single incorrect element content-wise)	4
Grammatical error (conjugated incorrectly, missing an element, not made into a noun or question correctly, order of elements is completely wrong)	3
Mostly wrong elements with some (2-3, depending on length) correct or correct grammar	2
One element correct (like a root, or verb conjugated correctly, but the other parts are off)	1
Totally off - wrong grammar, wrong elements	0

APPENDIX F: DATA

Fake Data

Real Data

Area of interest: linguistics, etymology, and language comprehension

Research Question: Does multilingualism improve our ability to recognize patterns in an unfamiliar language and to learn that language from examples?

Introduction: I am fascinated by everything linguistic: sentence structure; the development, construction, and evolution of languages; etymology; grammar; the neurology of language learning, speaking, and writing; the effect of language on society and vice versa; ancient languages; and probably whole other realms of linguistic phenomenons about which I have yet to learn. I adore learning about languages because of the complex patterns that lie in wait beneath the seemingly random surface. As such, I am curious about how we learn to identify and use these patterns in order to learn new languages. My question is therefore one of how well different people are able to perform this process of connection based on their linguistic backgrounds, since being in this practice of recognising a language's patterns should enable us to do it more readily with new languages.

Hypothesis: People who speak a variety of languages on a regular basis will be able to recognise patterns in a new, unfamiliar language better than others and thus learn the language more quickly by example.

Null Hypothesis: There will be no correlation between multilingualism and the ability to comprehend this new language from examples.

Experimental Design:

Variables:

- Correctness in comprehension
- Approach - example or rule
- Multilingualism/how many languages they're learning (or speaking regularly) and how well they speak them

Materials:

- "Dictionary": Reference papers with basic grammatical rules and all necessary vocabulary of Orkaki.
- Example Sentences: Examples of sentences in Orkaki, translated into English.
- Comprehension Test: On this handout, participants are asked to translate 5 short sentences into English from Orkaki and 5 short sentences from English into Orkaki. At the end, it asks for the participant to write down all of the languages they speak (that are currently in practice) and mark how well on a gradient they speak them (5 being natively, 4 being fluently, 3 being competently, 2 being intermediately, 1 being learning).

Procedure:

1. A google
1. Every other group is in the control group and the rest are in the experimental group.
2. Participants sit at tables and are asked not to communicate with one another; the room is quiet.
3. The control participants are given the Dictionary and the experimental participants are given the Example Sentences.
4. They are given five minutes to read through these.
5. They are then presented with the Comprehension Test and asked to complete it, retaining their original handout as a reference.
6. Their Multilingualism Score (MS) is calculated by the sum of the scores of the languages they speak.
7. Their Comprehension Score (CS) is ranked for each sentence based on closeness to the correct answer, with 5 being exact and 0 being completely off. These scores are then compiled into two means, the mean of those translating to English and the mean of those translating to Orkaki. These two means are added up to form the CS for the subject.
8. The data are then analysed using the following methods:
   a. An unpaired, two-tailed T-Test on the difference of means between the control and experimental groups as wholes.
   b. Regression tests on both data sets individually (the independent variable being MS and the dependent being CS), and these two models are compared to see if there is any difference.

- Research Question: Does a specific or broad grasp of a language assist more in comprehending it?
- Or: Does having to produce or understand a language result in having better comprehension of it?
- Variables:
    - multilingual-ness (level of fluency in other languages); I will ask for people to tell me how many languages they speak and how well they speak them so I can try to check if this is affecting the results
    - Cognates? I intend to choose a totally unfamiliar language to avoid this.
- Control: A group of people (probably smaller than the experimental group since that will have two groups within it) who receive no preparation or reference for the task (this wouldn't apply for my first design)
- Possible experimental designs:
1. Everyone gets the same task, with different modes of preparation. This ending task is to write down the gist of a few different sentences in the strange language, and then a paragraph of 2-4 sentences. For preparation, they are split into three groups: one gets nothing (control), the next gets a page of essential terms in the language, and the third gets a page of sentences in both English and the language.
2. Everyone is given a "dictionary" for the strange language (some basic terms, a few sample sentences, and all necessary vocabulary), and then given a sheet of sentences in the strange language. Half of them are told to translate these sentences into English, the other half told to write the sentences' gists.
    a. Possibly, at the end, they have to try to understand a short paragraph or write a few simple statements in the language. Either of these would be the final test, or the experiment would just be the aforementioned sentence thingy.
3. Either of these, but instead of asking them to generate their own answer, give them a few from which to choose.
4. Production v. Understanding: Ask one group to translate into the language and ask the other to translate from the language, then ask them to try to understand a short paragraph.