# THE METHOD OF QUANTITATIVE EVALUATION OF SYNTACTICAL INVERSION

**or**

**The freedom of the Russian syntax is overestimated**



*Eduard Klyshinsky*
*Higher School of Economics*
*06/07/18, Wroclaw, QUALICO-2018*

# The Goal of the Project


Just for fun.

# The Goal of the Project

How we can measure differences, or similarities, among languages?

# Syntactical Inversion

In some languages the same phrase can have both head initial and head final word orders.
In this work, we will consider syntactical inversion between two syntactically connected word types.

# Syntactical Inversion

W Rudnikach w województwie opolskim znajduje się już też Szkoła Podstawowa im. Andrzeja Wajdy.

Wczoraj warszawska rada miasta niemal jednogłośnie ...

http://wyborcza.pl/7,101707,23638985,powstanie-centrum-kultury-filmowej-im-andrzeja-wajdy-w.htm

# Method of Analysis

1. Let us calculate frequencies of tuples <type of head node, type of a tail node, syntactical connection, direction> of a for a syntactically tagged corpus.
2. Find tuples where head and tail nodes are swapped.
3. Calculate the criteria of symmetry: $s = 1 - \frac{|f_1 - f_2|}{(f_1 + f_2)}$

where $f_1$ and $f_2$ – frequencies of sufficient tuples.

# Method of Analysis

s=1 if $f_1 = f_2$ - completely symmetrical (irregular) connection.

s=0 if $f_1 = 0$ or $f_2 = 0$ - completely unsymmetrical (regular) connection.

# Method of Analysis

4. Calculate the importance of inversion: $i=s*q$, where $q$ – relative frequency of a tuple in the corpus.
5. Select 10 most important tuples for every analysed language. Join all important tuples of all languages in a list.
6. Calculate the sum of importance for the top 10 tuples and for the tuples in the constructed list. These figures are considered as the measure of irregularity of the syntax of considered languages.

# Method of Analysis

7. Calculate the matrix of correlations for all of syntactical connections in the corpus for all languages.

# Used Corpora

Universal Dependencies v 2.0 и 2.1
46 languages: Czech, Slovak, Polish, Russian, Ukrainian, Bulgarian, Croatian, Serbian Slovenian, Old Church Slavonic, Lettish, English, German, Gothic, Dutch, Afrikaans, Norwegian (separately Bokmål and Nynorsk), Swedish, Danish, Spanish, Catalan, French, Portuguese, Brazilian, Galician, Italian, Romanian, Latinic, Finnish, Estonian, Hungarian, Greek (modern and ancient), Hebrew, Arabic, Hindi, Urdu, Farsi, Japanese, Chinese, Korean, Indonesian, Vietnamese, Basque, Turkish.

# Used Corpora

Universal Dependencies v 2.0
Size of corpora starts from 34K tokens (Estonian)
to 1,8M (Czech).

Universal Dependencies v 2.1
Size of corpora starts from 43K tokens (Vietnamese)
to 2,2M (Czech).

# Top 10 Languages (v 2.0) with the Strictest Word Order

| Language | Size of a corpus | Importance | Importance, top10 |
|---|---|---|---|
| **Spanish** | 906 000 | 0,213 | 0,09 |
| **Catalan** | 472 000 | 0,188 | 0,086 |
| **Arabic** | 846 000 | 0,151 | 0,08 |
| **French** | 456 000 | 0,17 | 0,079 |
| **English** | 422 000 | 0,154 | 0,072 |
| **Farsi** | 135 000 | 0,129 | 0,052 |
| **Turkish** | 46 000 | 0,217 | 0,047 |
| **Urdu** | 123 000 | 0,054 | 0,035 |
| **Hindi** | 316 000 | 0,039 | 0,018 |
| **Japanese** | 362 000 | 0,016 | 0,001 |

# Top 10 Languages (v 2.1) with the Strictest Word Order

| Language | Size of a corpus | Importance | Importance, top10 |
|---|---|---|---|
| **French** | 1 099 000 | 0,191 | 0,086 |
| **Arabic** | 1 042 000 | 0,150 | 0,082 |
| **English** | 496 000 | 0,156 | 0,071 |
| **Farsi** | 152 000 | 0,129 | 0,054 |
| **Turkish** | 74 000 | 0,215 | 0,049 |
| **Urdu** | 138 000 | 0,053 | 0,036 |
| **Chinese** | 153 000 | 0,148 | 0,025 |
| **Hindi** | 375 000 | 0,039 | 0,018 |
| **Korean** | 97 000 | 0,019 | 0,006 |
| **Japanese** | 402 000 | 0,016 | 0,001 |

# Top 10 Languages (v 2.0) with the Weakest Word Order

| Language | Size of a corpus | Importance | Importance, top10 |
|---|---|---|---|
| **Estonian** | 34 000 | 0,447 | 0,274 |
| **Finnish** | 324 000 | 0,342 | 0,222 |
| **Slovak** | 93 000 | 0,376 | 0,214 |
| **Polish** | 72 000 | 0,355 | 0,21 |
| **Slovenian** | 145 000 | 0,309 | 0,169 |
| **Dutch** | 290 000 | 0,278 | 0,166 |
| **Czech** | 1 838 000 | 0,317 | 0,159 |
| **German** | 277 000 | 0,278 | 0,157 |
| **Lettish** | 44 000 | 0,353 | 0,155 |
| **Hungarian** | 37 000 | 0,278 | 0,143 |

# Top 10 Languages (v 2.1) with the Weakest Word Order

| Language | Size of a corpus | Importance | Importance, top10 |
|---|---|---|---|
| **Gothic** | 55 000 | 0,423 | 0,352 |
| **Anc. Greek** | 414 000 | 0,446 | 0,35 |
| **Old Church Slav.** | 57 000 | 0,411 | 0,348 |
| **Latin** | 491 000 | 0,480 | 0,333 |
| **Estonian** | 106 000 | 0,46 | 0,262 |
| **Finnish** | 377 000 | 0,342 | 0,226 |
| **Slovak** | 106 000 | 0,366 | 0,211 |
| **Polish** | 83 000 | 0,355 | 0,210 |
| **Basque** | 121 000 | 0,361 | 0,198 |
| **Slovenian** | 170 000 | 0,308 | 0,171 |

# Data Comparison(2.0 vs 2.1)

| Language | Size of a corpus | Importance | Importance, top10 |
|---|---|---|---|
| **Estonian** | 72 000 (212%) | 0,012723 | -0,011995 |
| **Finnish** | 53 000 ( 16%) | 0,000478 | 0,003719 |
| **Slovak** | 13 000 ( 14%) | -0,010019 | -0,003008 |
| **Polish** | 11 000 ( 15%) | 0,000480 | 0,000149 |
| **Slovenian** | 25 000 ( 17%) | -0,000918 | 0,002582 |
| **Dutch** | 20 000 ( 7%) | 0,008396 | -0,004849 |
| **Czech** | 384 000 ( 21%) | 0,007520 | 0,007005 |
| **German** | 36 000 ( 13%) | 0,002159 | 0,000583 |
| **Lettish** | 46 000 (139%) | 0,015990 | 0,008563 |
| **Hungarian** | 5 000 ( 14%) | 0,004348 | 0,002743 |

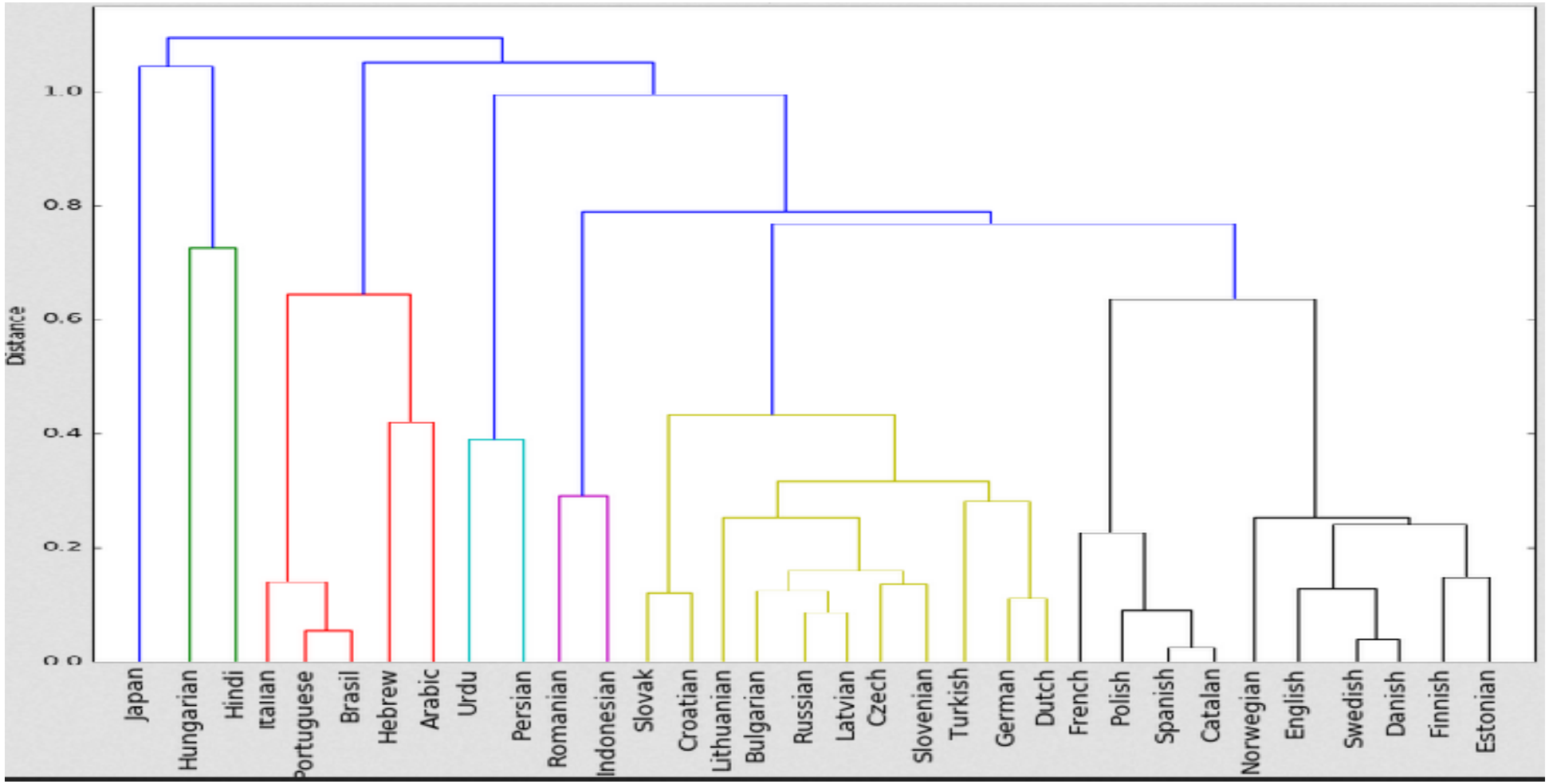
# Interrogative sentences

| | Quest. marks | Corpus size | Importance |
|---|---|---|---|
| **Latin** | 57 | 491 000 | 0,480 |
| **Estonian** | 485 | 106 000 | 0,46 |
| **Ancient Greek** | 2 | 414 000 | 0,446 |
| **Gothic** | 0 | 55 000 | 0,423 |
| **Old Church Slavonic** | 0 | 57 000 | 0,411 |
| **Slovak** | 398 | 106 000 | 0,366 |
| **Basque** | 240 | 121 000 | 0,361 |
| **Polish** | 615 | 83 000 | 0,355 |
| **Finnish** | 455 | 377 000 | 0,342 |
| **Slovenian** | 517 | 170 000 | 0,308 |

# Correlation among Languages

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 Czech | 100 | 90 | 80 | 88 | 93 | 90 | 77 | 61 | 87 | 41 | 86 | 58 | 81 | 38 | 84 | 66 | 65 | 73 | 64 | 66 | 78 | 85 | 6 | 62 | 22 | 16 | 49 | 78 | 72 | 71 | 84 | 39 | 75 | 62 | 13 | 1 | 14 | 8 | 18 | -4 | 36 | 14 | 42 | 25 | 81 | 75 |
| 2 Slovak | 90 | 100 | 62 | 84 | 87 | 85 | 89 | 76 | 83 | 39 | 81 | 53 | 80 | 34 | 79 | 68 | 61 | 66 | 60 | 63 | 59 | 67 | 1 | 43 | 15 | 8 | 35 | 64 | 59 | 65 | 79 | 33 | 67 | 55 | 11 | 0 | 9 | 4 | 7 | -4 | 42 | 6 | 36 | 20 | 80 | 70 |
| 3 Polish | 80 | 62 | 100 | 60 | 69 | 65 | 50 | 38 | 59 | 33 | 54 | 39 | 49 | 31 | 50 | 38 | 44 | 58 | 41 | 43 | 94 | 92 | 12 | 75 | 26 | 27 | 59 | 68 | 69 | 44 | 52 | 26 | 56 | 46 | 8 | 1 | 6 | 5 | 26 | -3 | 17 | 7 | 25 | 18 | 56 | 53 |
| 4 Russian | 88 | 84 | 60 | 100 | 96 | 91 | 78 | 61 | 89 | 44 | 94 | 50 | 83 | 40 | 85 | 66 | 64 | 69 | 59 | 59 | 60 | 69 | 4 | 42 | 15 | 6 | 36 | 64 | 69 | 66 | 83 | 33 | 70 | 63 | 12 | 2 | 13 | 10 | 10 | -3 | 38 | -1 | 41 | 24 | 73 | 75 |
| 5 Ukrainian | 93 | 87 | 69 | 96 | 100 | 91 | 77 | 60 | 91 | 45 | 93 | 56 | 87 | 41 | 89 | 70 | 66 | 72 | 64 | 64 | 68 | 76 | 6 | 51 | 19 | 11 | 42 | 68 | 73 | 72 | 88 | 39 | 74 | 67 | 12 | 2 | 15 | 9 | 10 | -3 | 39 | -2 | 39 | 23 | 77 | 79 |
| 6 Bulgarian | 90 | 85 | 65 | 91 | 91 | 100 | 73 | 58 | 85 | 47 | 90 | 65 | 80 | 42 | 80 | 59 | 73 | 78 | 73 | 72 | 68 | 77 | 3 | 52 | 20 | 13 | 43 | 73 | 67 | 70 | 85 | 30 | 80 | 62 | 14 | 1 | 6 | 7 | 10 | -4 | 30 | 2 | 35 | 19 | 74 | 73 |
| 7 Croatian | 77 | 89 | 50 | 78 | 77 | 73 | 100 | 92 | 78 | 32 | 75 | 34 | 74 | 28 | 72 | 77 | 50 | 54 | 41 | 44 | 45 | 54 | 1 | 28 | 8 | 0 | 25 | 46 | 53 | 50 | 66 | 29 | 51 | 48 | 6 | 0 | 7 | 6 | 8 | -3 | 40 | -2 | 30 | 19 | 75 | 59 |
| 8 Serbian | 61 | 76 | 38 | 61 | 60 | 58 | 92 | 100 | 62 | 28 | 58 | 23 | 58 | 23 | 53 | 60 | 41 | 47 | 33 | 35 | 35 | 43 | 0 | 18 | 5 | -1 | 19 | 35 | 36 | 35 | 51 | 24 | 44 | 34 | 8 | 2 | 0 | 0 | 2 | -3 | 30 | -3 | 19 | 11 | 63 | 46 |
| 9 Slovenian | 87 | 83 | 59 | 89 | 91 | 85 | 78 | 62 | 100 | 37 | 88 | 55 | 86 | 33 | 87 | 74 | 63 | 65 | 63 | 62 | 56 | 66 | 2 | 44 | 14 | 5 | 33 | 62 | 63 | 72 | 84 | 34 | 66 | 62 | 9 | 0 | 17 | 9 | 9 | -3 | 39 | 2 | 39 | 23 | 77 | 71 |
| 10 Church Slavo | 41 | 39 | 33 | 44 | 45 | 47 | 32 | 28 | 37 | 100 | 43 | 35 | 30 | 97 | 31 | 14 | 35 | 40 | 36 | 32 | 36 | 40 | 5 | 26 | 12 | 9 | 24 | 41 | 70 | 30 | 38 | 18 | 47 | 71 | 7 | -2 | 8 | 18 | 4 | -4 | 9 | -3 | 22 | 15 | 53 | 33 |
| 11 Latvian | 86 | 81 | 54 | 94 | 93 | 90 | 75 | 58 | 88 | 43 | 100 | 57 | 85 | 40 | 85 | 70 | 71 | 75 | 63 | 62 | 54 | 65 | 3 | 38 | 14 | 8 | 33 | 60 | 68 | 76 | 88 | 37 | 73 | 70 | 12 | 3 | 18 | 26 | 26 | -3 | 38 | -4 | 36 | 22 | 74 | 73 |
| 12 English | 58 | 53 | 39 | 50 | 56 | 65 | 34 | 23 | 55 | 35 | 57 | 100 | 58 | 32 | 55 | 40 | 78 | 72 | 93 | 87 | 52 | 58 | 0 | 60 | 26 | 24 | 37 | 65 | 44 | 82 | 78 | 7 | 81 | 50 | 14 | -3 | 11 | 29 | 26 | -3 | 10 | 8 | 26 | 7 | 52 | 55 |
| 13 German | 81 | 80 | 49 | 83 | 87 | 80 | 74 | 58 | 86 | 30 | 85 | 58 | 100 | 28 | 95 | 80 | 74 | 71 | 69 | 72 | 51 | 60 | 4 | 42 | 18 | 9 | 33 | 55 | 59 | 76 | 88 | 34 | 64 | 58 | 8 | 0 | 22 | 13 | 14 | -4 | 43 | -3 | 37 | 21 | 72 | 74 |
| 14 Gothic | 38 | 34 | 31 | 40 | 41 | 42 | 28 | 23 | 33 | 97 | 40 | 32 | 28 | 100 | 29 | 15 | 32 | 36 | 33 | 30 | 34 | 37 | 1 | 25 | 9 | 6 | 21 | 37 | 72 | 28 | 35 | 15 | 40 | 67 | 5 | -3 | 7 | 17 | 8 | -4 | 15 | -3 | 19 | 18 | 48 | 31 |
| 15 Dutch | 84 | 79 | 50 | 85 | 89 | 80 | 72 | 53 | 87 | 31 | 85 | 55 | 95 | 29 | 100 | 80 | 69 | 67 | 67 | 69 | 50 | 60 | 1 | 41 | 13 | 4 | 30 | 55 | 63 | 76 | 88 | 39 | 61 | 61 | 7 | -2 | 23 | 8 | 5 | -3 | 43 | -4 | 36 | 23 | 73 | 74 |
| 16 Afrikaans | 66 | 68 | 38 | 66 | 70 | 59 | 77 | 60 | 74 | 14 | 70 | 40 | 80 | 15 | 80 | 100 | 48 | 44 | 45 | 47 | 37 | 44 | -1 | 32 | 8 | -1 | 20 | 39 | 54 | 60 | 71 | 25 | 40 | 45 | -1 | -5 | 18 | 14 | 19 | -3 | 45 | -3 | 31 | 20 | 68 | 56 |
| 17 Norwegian B | 65 | 61 | 44 | 64 | 66 | 73 | 50 | 41 | 63 | 35 | 71 | 78 | 74 | 32 | 69 | 48 | 100 | 95 | 84 | 84 | 52 | 60 | 1 | 47 | 19 | 15 | 34 | 56 | 48 | 74 | 82 | 14 | 74 | 52 | 11 | -2 | 7 | 31 | 33 | -4 | 15 | -2 | 24 | 9 | 53 | 64 |
| 18 Norwegian N | 73 | 66 | 58 | 69 | 72 | 78 | 54 | 47 | 65 | 40 | 75 | 72 | 71 | 36 | 67 | 44 | 95 | 100 | 79 | 80 | 62 | 70 | 3 | 51 | 21 | 19 | 41 | 61 | 53 | 73 | 81 | 22 | 80 | 56 | 14 | 0 | 6 | 28 | 38 | -4 | 13 | 0 | 22 | 9 | 55 | 66 |
| 19 Swedish | 64 | 60 | 41 | 59 | 64 | 73 | 41 | 33 | 63 | 36 | 63 | 93 | 69 | 33 | 67 | 45 | 84 | 79 | 100 | 96 | 51 | 60 | 2 | 57 | 26 | 21 | 37 | 65 | 43 | 80 | 84 | 12 | 82 | 48 | 16 | -1 | 4 | 13 | 13 | -3 | 12 | 1 | 25 | 7 | 52 | 61 |
| 20 Danish | 66 | 63 | 43 | 59 | 64 | 72 | 44 | 35 | 62 | 32 | 62 | 87 | 72 | 30 | 69 | 47 | 84 | 80 | 96 | 100 | 55 | 63 | 3 | 59 | 27 | 21 | 38 | 63 | 42 | 77 | 84 | 10 | 79 | 41 | 14 | -1 | -1 | 6 | 14 | -3 | 25 | 3 | 26 | 10 | 51 | 67 |
| 21 Spanish | 78 | 59 | 94 | 60 | 68 | 68 | 45 | 35 | 56 | 36 | 54 | 52 | 51 | 34 | 50 | 37 | 52 | 62 | 51 | 55 | 100 | 96 | 22 | 89 | 40 | 40 | 71 | 76 | 69 | 48 | 57 | 15 | 63 | 45 | 13 | 4 | 0 | 7 | 31 | -3 | 17 | 14 | 33 | 19 | 59 | 56 |
| 22 Catalan | 85 | 67 | 92 | 69 | 76 | 77 | 54 | 43 | 66 | 40 | 65 | 58 | 60 | 37 | 60 | 44 | 60 | 70 | 60 | 63 | 96 | 100 | 18 | 84 | 37 | 34 | 67 | 77 | 70 | 56 | 67 | 21 | 72 | 51 | 14 | 2 | 2 | 7 | 27 | -3 | 21 | 11 | 32 | 19 | 65 | 63 |
| 23 Galician | 6 | 1 | 12 | 4 | 6 | 3 | 1 | 0 | 2 | 5 | 3 | 0 | 4 | 1 | 1 | -1 | 1 | 3 | 2 | 3 | 22 | 18 | 100 | 46 | 92 | 85 | 80 | 19 | 6 | 2 | 10 | -3 | 18 | 5 | 39 | 54 | -6 | 15 | 3 | -3 | -4 | 0 | 34 | -1 | 7 | 5 |
| 24 French | 62 | 43 | 75 | 42 | 51 | 52 | 28 | 18 | 44 | 26 | 38 | 60 | 42 | 25 | 41 | 32 | 47 | 51 | 57 | 59 | 89 | 84 | 46 | 100 | 65 | 63 | 86 | 77 | 55 | 47 | 53 | 0 | 59 | 34 | 19 | 17 | -2 | 12 | 27 | -4 | 9 | 21 | 43 | 16 | 49 | 46 |
| 25 Portuguese | 22 | 15 | 26 | 15 | 19 | 20 | 8 | 5 | 14 | 12 | 14 | 26 | 18 | 9 | 13 | 8 | 19 | 21 | 26 | 27 | 40 | 37 | 92 | 65 | 100 | 95 | 90 | 37 | 19 | 21 | 28 | -1 | 38 | 15 | 41 | 50 | -6 | 16 | 8 | -4 | -2 | 7 | 38 | 1 | 19 | 19 |
| 26 Brasil | 16 | 8 | 27 | 6 | 11 | 13 | 0 | -1 | 5 | 9 | 8 | 24 | 9 | 6 | 4 | -1 | 15 | 19 | 21 | 21 | 40 | 34 | 85 | 63 | 95 | 100 | 85 | 34 | 16 | 23 | 21 | -2 | 32 | 12 | 38 | 48 | -4 | 23 | 17 | -4 | -6 | 10 | 31 | -2 | 15 | 11 |
| 27 Italian | 49 | 35 | 59 | 36 | 42 | 43 | 25 | 19 | 33 | 24 | 33 | 37 | 33 | 21 | 30 | 20 | 34 | 41 | 37 | 38 | 71 | 67 | 80 | 86 | 90 | 85 | 100 | 62 | 45 | 33 | 43 | 7 | 52 | 31 | 36 | 41 | -3 | 18 | 20 | -4 | 4 | 15 | 46 | 9 | 41 | 37 |
| 28 Romanian | 78 | 64 | 68 | 64 | 68 | 73 | 46 | 35 | 62 | 41 | 60 | 65 | 55 | 37 | 55 | 39 | 56 | 61 | 65 | 63 | 76 | 77 | 19 | 77 | 37 | 34 | 62 | 100 | 63 | 59 | 65 | 13 | 71 | 52 | 13 | -2 | 2 | 12 | 25 | -4 | 17 | 41 | 72 | 22 | 74 | 54 |
| 29 Latin | 72 | 59 | 69 | 69 | 73 | 67 | 53 | 36 | 63 | 70 | 68 | 44 | 59 | 72 | 63 | 54 | 48 | 53 | 43 | 42 | 69 | 70 | 6 | 55 | 19 | 16 | 45 | 63 | 100 | 53 | 60 | 32 | 52 | 81 | 1 | -4 | 23 | 30 | 31 | -5 | 36 | 3 | 36 | 25 | 75 | 55 |
| 30 Finnish | 71 | 65 | 44 | 66 | 72 | 70 | 50 | 35 | 72 | 30 | 76 | 82 | 76 | 28 | 76 | 60 | 74 | 73 | 80 | 77 | 48 | 56 | 2 | 47 | 21 | 23 | 33 | 59 | 53 | 100 | 89 | 32 | 76 | 61 | 13 | 0 | 30 | 37 | 34 | -4 | 26 | 5 | 28 | 12 | 62 | 67 |
| 31 Estonian | 84 | 79 | 52 | 83 | 88 | 85 | 66 | 51 | 84 | 38 | 88 | 78 | 88 | 35 | 88 | 71 | 82 | 81 | 84 | 84 | 57 | 67 | 10 | 53 | 28 | 21 | 43 | 65 | 60 | 89 | 100 | 34 | 81 | 64 | 16 | 6 | 18 | 25 | 22 | -3 | 33 | 1 | 34 | 17 | 71 | 78 |
| 32 Hungarian | 39 | 33 | 26 | 33 | 39 | 30 | 29 | 24 | 34 | 18 | 37 | 7 | 34 | 15 | 39 | 25 | 14 | 22 | 12 | 10 | 15 | 21 | -3 | 0 | -1 | -2 | 7 | 13 | 32 | 32 | 34 | 100 | 26 | 37 | 4 | 0 | 30 | 8 | -2 | -3 | 18 | -3 | 1 | 5 | 35 | 31 |
| 33 Greek | 75 | 67 | 56 | 70 | 74 | 80 | 51 | 44 | 66 | 47 | 73 | 81 | 64 | 40 | 61 | 40 | 74 | 80 | 82 | 79 | 63 | 72 | 18 | 59 | 38 | 32 | 52 | 71 | 52 | 76 | 81 | 26 | 100 | 57 | 23 | 6 | 6 | 16 | 19 | -4 | 11 | 4 | 36 | 8 | 59 | 66 |
| 34 Ancient Gree | 62 | 55 | 46 | 63 | 67 | 62 | 48 | 34 | 62 | 71 | 70 | 50 | 58 | 67 | 61 | 45 | 52 | 56 | 48 | 41 | 45 | 51 | 5 | 34 | 15 | 12 | 31 | 52 | 81 | 61 | 64 | 37 | 57 | 100 | 3 | -3 | 32 | 43 | 27 | -5 | 26 | -2 | 31 | 14 | 73 | 49 |
| 35 Hebrew | 13 | 11 | 8 | 12 | 12 | 14 | 6 | 8 | 9 | 7 | 12 | 14 | 8 | 5 | 7 | -1 | 11 | 14 | 16 | 14 | 13 | 14 | 39 | 19 | 41 | 38 | 36 | 13 | 1 | 13 | 16 | 4 | 23 | 3 | 100 | 58 | -6 | 2 | 0 | -4 | -6 | 3 | 4 | -3 | 7 | 11 |
| 36 Arabic | 1 | 0 | 1 | 2 | 2 | 1 | 0 | 2 | 0 | -2 | 3 | -3 | 0 | -3 | -2 | -5 | -2 | 0 | -1 | -1 | 4 | 2 | 54 | 17 | 50 | 48 | 41 | -2 | -4 | 0 | 6 | 0 | 6 | -3 | 58 | 100 | -5 | 10 | -2 | -3 | -5 | -4 | -1 | -4 | -3 | 1 |
| 37 Hindi | 14 | 9 | 6 | 13 | 15 | 6 | 7 | 0 | 17 | 8 | 18 | 11 | 22 | 7 | 23 | 18 | 7 | 6 | 4 | -1 | 0 | 2 | -6 | -2 | -6 | -4 | -3 | 2 | 23 | 30 | 18 | 30 | 6 | 32 | -6 | -5 | 100 | 50 | 12 | -3 | 11 | -3 | 0 | 3 | 25 | 15 |
| 38 Urdu | 8 | 4 | 5 | 10 | 9 | 7 | 6 | 0 | 9 | 18 | 26 | 29 | 13 | 17 | 8 | 14 | 31 | 28 | 13 | 6 | 7 | 7 | 15 | 12 | 16 | 23 | 18 | 12 | 30 | 37 | 25 | 8 | 16 | 43 | 2 | 10 | 50 | 100 | 63 | -3 | 2 | 6 | 9 | 3 | 28 | 7 |
| 39 Persian | 18 | 7 | 26 | 10 | 10 | 10 | 8 | 2 | 9 | 4 | 26 | 26 | 14 | 8 | 5 | 19 | 33 | 38 | 13 | 14 | 31 | 27 | 3 | 27 | 8 | 17 | 20 | 25 | 31 | 34 | 22 | -2 | 19 | 27 | 0 | -2 | 12 | 63 | 100 | -2 | 2 | 21 | 16 | 9 | 19 | 13 |
| 40 Japanese | -4 | -4 | -3 | -3 | -3 | -4 | -3 | -3 | -3 | -4 | -3 | -3 | -4 | -4 | -3 | -3 | -4 | -4 | -3 | -3 | -3 | -3 | -3 | -4 | -4 | -4 | -4 | -4 | -5 | -4 | -3 | -3 | -4 | -5 | -4 | -3 | -3 | -3 | -2 | 100 | -3 | -3 | -3 | -3 | -4 | -5 |
| 41 Chinese | 36 | 42 | 17 | 38 | 39 | 30 | 40 | 30 | 39 | 9 | 38 | 10 | 43 | 15 | 43 | 45 | 15 | 13 | 12 | 25 | 17 | 21 | -4 | 9 | -2 | -6 | 4 | 17 | 36 | 26 | 33 | 18 | 11 | 26 | -6 | -5 | 11 | 2 | 2 | -3 | 100 | -1 | 21 | 14 | 46 | 26 |
| 42 Korean | 14 | 6 | 7 | -1 | -2 | 2 | -2 | -3 | 2 | -3 | -4 | 8 | -3 | -3 | -4 | -3 | -2 | 0 | 1 | 3 | 14 | 11 | 0 | 21 | 7 | 10 | 15 | 41 | 3 | 5 | 1 | -3 | 4 | -2 | 3 | -4 | -3 | 6 | 21 | -3 | -1 | 100 | 44 | 12 | 23 | 3 |
| 43 Indonesian | 42 | 36 | 25 | 41 | 39 | 35 | 30 | 19 | 39 | 22 | 36 | 26 | 37 | 19 | 36 | 31 | 24 | 22 | 25 | 26 | 33 | 32 | 34 | 43 | 38 | 31 | 46 | 72 | 36 | 28 | 34 | 1 | 36 | 31 | 4 | -1 | 0 | 9 | 16 | -3 | 21 | 44 | 100 | 23 | 56 | 30 |
| 44 Vietnamese | 25 | 20 | 18 | 24 | 23 | 19 | 19 | 11 | 23 | 15 | 22 | 7 | 21 | 18 | 23 | 20 | 9 | 9 | 7 | 10 | 19 | 19 | -1 | 16 | 1 | -2 | 9 | 22 | 25 | 12 | 17 | 5 | 8 | 14 | -3 | -4 | 3 | 3 | 9 | -3 | 14 | 12 | 23 | 100 | 25 | 31 |
| 45 Basque | 81 | 80 | 56 | 73 | 77 | 74 | 75 | 63 | 77 | 53 | 74 | 52 | 72 | 48 | 73 | 66 | 53 | 55 | 52 | 51 | 59 | 65 | 7 | 49 | 19 | 15 | 41 | 74 | 75 | 62 | 71 | 35 | 59 | 73 | 7 | -3 | 25 | 28 | 19 | -4 | 46 | 23 | 56 | 25 | 100 | 56 |
| 46 Turkish | 75 | 70 | 53 | 75 | 79 | 73 | 59 | 46 | 71 | 33 | 73 | 55 | 74 | 31 | 74 | 56 | 64 | 66 | 61 | 67 | 56 | 63 | 5 | 46 | 19 | 11 | 37 | 54 | 55 | 67 | 78 | 31 | 66 | 49 | 11 | 1 | 15 | 7 | 13 | -5 | 26 | 3 | 30 | 31 | 56 | 100 |

# Language Clustering



# The Most Important Connections

| Tail word | Head word | Label | English |
|---|---|---|---|
| ADV | VERB | advmod | 0,0229272 |
| NOUN | VERB | obl | 0,0066593 |
| VERB | VERB | advcl | 0,00656007 |
| NOUN | NOUN | nmod | 0,00653802 |
| NOUN | VERB | nsubj | 0,00363285 |
| PRON | VERB | obj | 0,00297133 |
| PROPN | NOUN | nmod | 0,0024807 |
| ADV | NOUN | advmod | 0,0023925 |
| NUM | NOUN | nummod | 0,00203417 |
| ADJ | NOUN | amod | 0,00194597 |

# The Most Important Connections

| Tail word | Head word | Label | Russian |
|---|---|---|---|
| NOUN | VERB | obl | 0,0371461 |
| NOUN | VERB | nsubj | 0,0224947 |
| ADV | VERB | advmod | 0,00981738 |
| VERB | VERB | advcl | 0,00641601 |
| PRON | VERB | obl | 0,00607148 |
| NOUN | VERB | obj | 0,005625 |
| VERB | NOUN | acl | 0,00468809 |
| PRON | VERB | obj | 0,00413086 |
| NOUN | VERB | nsubj:pass | 0,00324844 |
| PRON | NOUN | nmod | 0,00324316 |

# The Most Important Connections

| Tail word | Head word | Label | Gothic |
|---|---|---|---|
| CCONJ | VERB | cc | 0,073344 |
| VERB | VERB | advcl | 0,0449217 |
| NOUN | VERB | nsubj | 0,0272581 |
| NOUN | VERB | obl | 0,0249297 |
| NOUN | VERB | obj:dir | 0,0214372 |
| PRON | VERB | obj:dir | 0,0204737 |
| ADV | VERB | advmod | 0,0167001 |
| VERB | VERB | xcomp | 0,0150943 |
| ADJ | NOUN | amod | 0,0135287 |
| ADV | VERB | discourse | 0,012766 |

# The Most Important Connections (Gothic)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | Jah | jah | CCONJ | C- | _ | 5 | cc |
| 2 | atsteigands | at-steigan | VERB | V- | 5 | advcl | |
| 3 | in | in | ADP | R- | _ | 4 | case |
| 4 | skip | skip | NOUN | Nb | 2 | obl | |
| 5 | ufarlaiþ | ufar-leiþan | VERB | 0 | root | | |
| 6 | jah | jah | CCONJ | C- | _ | 5 | cc |
| 7 | qam | qiman | VERB | 5 | conj | | |
| 8 | in | in | ADP | R- | _ | 10 | case |
| 9 | seinai | *seins | ADJ | 10 | nmod | | |
| 10 | baurg | baurgs | NOUN | 7 | obl | | |

# The Most Important Connections (Old Church Slavonic)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | ⍰ | и | CCONJ | C- | _ | 6 | cc |
| 2 | вьлѣзъ | вълѣсти | VERB | | 6 | advcl | |
| 3 | въ | въ | ADP | R- | _ | 4 | case |
| 4 | корабь | корабль | NOUN | | | 2 | obl |
| 5 | йсъ | исоусъ | PROPN | | 6 | nsubj | |
| 6 | прѣѣде | прѣ⍰хати | VERB | | 0 | root | |
| 7 | ⍰ | и | CCONJ | C- | _ | 6 | cc |
| 8 | приде | прити | VERB | | 6 | conj | |
| 9 | въ | въ | ADP | | 11 | case | |
| 10 | свои | свои | ADJ | 11 | nmod | | |
| 11 | градъ | градъ | NOUN | | 8 | obl | |

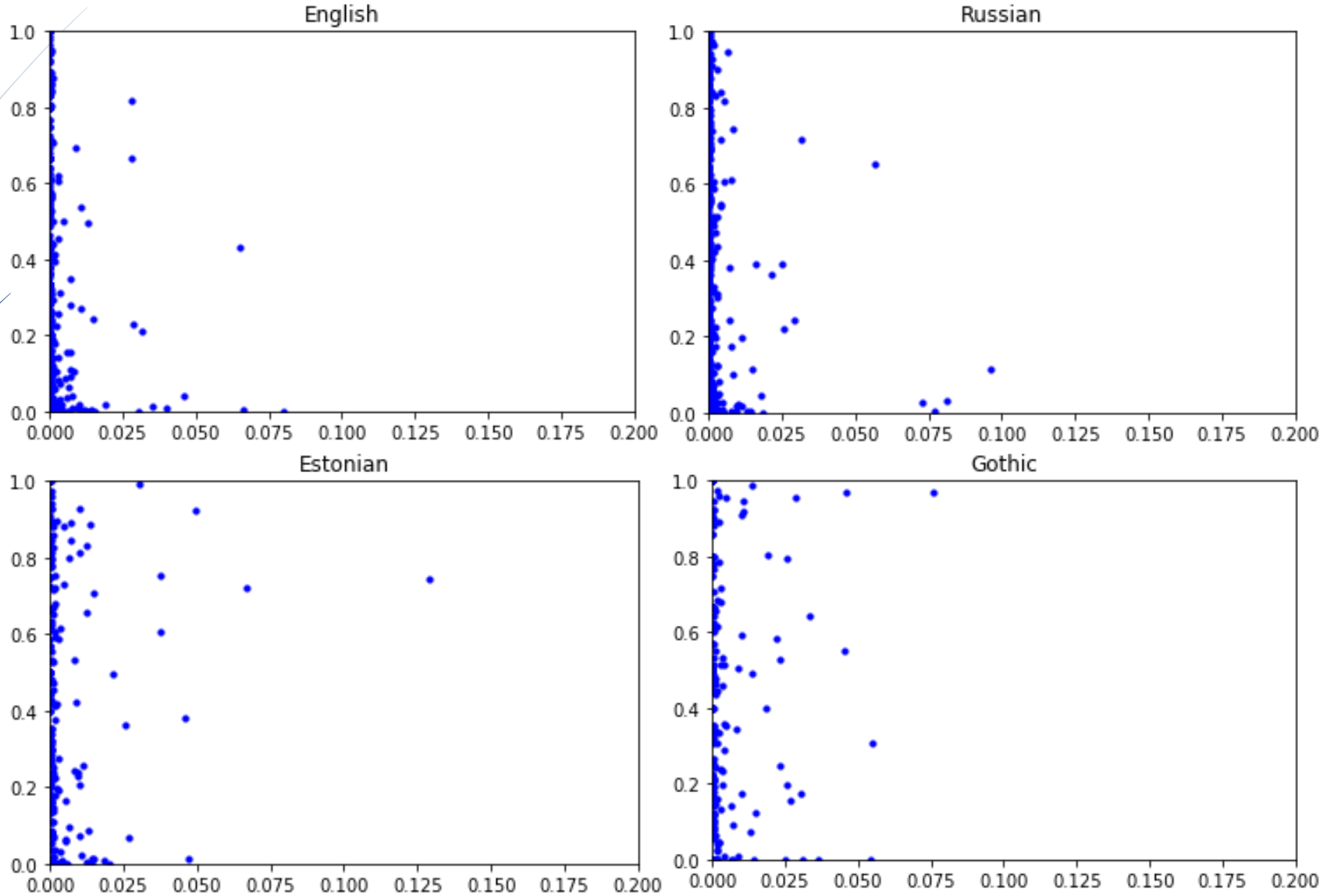
# The Most Important Connections

<NOUN, VERB, obl>, bigger than 0.001 in 39 languages; obsolete for Japanese, Korean and Brazilian (but presented in Portuguese).
<NOUN, VERB, nsubj>, bigger than 0.001 in 38 languages; obsolete for Japanese and Korean.
<ADV, VERB, advmod>, bigger than 0.001 in 37 languages; obsolete for Japanese, Vietnamese.
<VERB, VERB, advcl>, bigger than 0.001 in 35 languages; obsolete for Japanese, Korean, Afrikaans and Turkish.

# The Most Important Connections

<NOUN, VERB, obl>, bigger than 0.01 in 29 languages;
<NOUN, VERB, nsubj>, bigger than 0.01 in 24 languages;
<ADV, VERB, advmod>, bigger than 0.01 in 26 languages;
<VERB, VERB, advcl>, bigger than 0.01 in 5 languages.

# The Most Important Connections



# Conclusion

1. It is easy to evaluate the differences among syntactical irregularities of different languages.
2. It is easy to find where languages are the same.
3. It looks like a correct method since it adequately arranges languages.

# We Can Do It! Together.

I don’t know all of these languages, but possibly we can explain some linguistic effects together.