Formal Concept Analysis meets grammar typology

Kow Kuroda
Medical School, Kyorin University
FCA meets grammar typology at NLP 21

Introduction

Motivations, Goals and Outline
Why this work?

❖ In pursuit of truly effective methods of English teaching/learning, I wanted

❖ to measure the similarity among the grammars of languages, against which relative difficulty of a target language can be estimated.

❖ This should gives what I will call **relativized learnability index**.

❖ and then to answer, **Which language is the most similar to Japanese in terms of grammar?**

❖ To achieve this goal, I **needed a new measure that successfully replaces so-called “language distance”** which turned out to be too biased toward shared vocabulary/lexemes.
Outline of presentation

❖ Data and Analysis
  ❖ 15 languages are selected and manually encoded against 24 grammatical/morphological features.
  ❖ Formal Concept Analysis (FCA) was performed against a formal context with the 15 languages as objects and the 24 features as attributes.

❖ Results
  ❖ A series of experiments suggested a few optimal results, one of which I expect is informative enough to define relativized learnability index.
  ❖ Comparison between optimal and suboptimal FCA’s is revealing in typological studies of language.
  ❖ A tentative answer to, “Which language is most similar to Japanese in terms of grammar?”

❖ Discussion
FCA meets grammar typology at NLP 21

Data and Analysis

How data was set up and analyzed
Data setup

❖ The following **15 languages** are selected and manually encoded against **24 attributes** (to be shown later):

❖ Bulgarian, Chinese, Czech, English, French, Finnish, German, Hebrew, Hungarian, Japanese, Korean, Latin, Russian, Swahili, and Tagalog

❖ **Design criteria**

❖ aims to **cover as wide a variety of languages as possible**, 
❖ aims to **include as many phylogenically unrelated languages as possible**, and 
❖ aims to **provide a good background against which Japanese is well profiled**.

❖ **Caveats**

❖ All the criteria are far from fully satisfied in this study and generated a serious sampling bias in the results, admittedly.
24 attributes/features used in coding

- A1 Language has **Definite Articles**
- A2 Language has **Indefinite Articles**
- A3 **Noun** encodes Plurality
- A4 **Noun** encodes Class
- A5 **Noun** encodes Case
- A6 **Relative clause** follows **Noun**
- A7 **Language has Postpositions**
- A8 **Language has Prepositions**
- A9 **Adjective** agrees with **Noun-plurality**
- A10 **Adjective** agrees with **Noun-class**
- A11 **Adjective** agrees with **Noun-case**
- A12 **Adjective** follows **Noun**
- A13 **Object** must follow **Verb**
- A14 **Language requires Subject**
- A15 **Verb** encodes Voice
- A16 **Verb** encodes Tense
- A17 **Verb** encodes Aspect
- A18 **Verb** agrees with **Subject**
- A19 **Verb** encodes Person
- A20 **Verb** encodes Plurality
- A21 **Verb** encodes **Noun-class**
- A22 **Verb** infinitive is derived
- A23 **Verb** agrees with **Object**
- A24 **Language has Tense Agreement**
## Data coding

| Language  | has_definite | has_indefinite | N_en_codes_plurality | N_en_codes_case | relative_cl_foliation | has_postposition | has_preposition | A_agrees_w_Nplurality | A_agrees_w_Ncase | A_follows_with_Su | O_must_follow | V_agrees_with_Su | V_encodes_plurality | V_encodes_case | V_encodes_s_voiced | V_encodes_s_paired | V_encodes_asparagus | V_infinite_is_important | V_agrees_with_O imperative | tens_agrees_with_O imperative | check_sum |
|-----------|--------------|----------------|----------------------|----------------|-----------------------|------------------|------------------|-----------------------|------------------|----------------|---------------|----------------|------------------|----------------|------------------|-------------------|----------------|-------------------|---------------------|---------------|
| Bulgarian| 1            | 0              | 1                    | 1              | 0                     | 1                | 1                | 0                     | 0                | 1               | 0             | 1               | 1                | 1               | 1                | 0                 | 1             | 0                 | 13                  |               |
| Chinese  | 0            | 0              | 0                    | 0              | 0                     | 1                | 0                | 0                     | 0                | 1               | 0             | 0               | 0                | 1               | 0                | 0                 | 0             | 0                 | 3                   |               |
| Czech    | 0            | 0              | 0                    | 0              | 0                     | 1                | 0                | 0                     | 1                | 0               | 1             | 1               | 1                | 1               | 1                | 0                 | 1             | 0                 | 16                  |               |
| English  | 1            | 1              | 1                    | 0              | 0                     | 1                | 0                | 0                     | 0                | 1               | 1             | 1               | 1                | 1               | 1                | 0                 | 1             | 0                 | 13                  |               |
| Finnish  | 0            | 0              | 1                    | 1              | 1                     | 0                | 0                | 1                     | 0                | 1               | 1             | 1               | 1                | 1               | 1                | 0                 | 1             | 0                 | 13                  |               |
| French   | 1            | 1              | 1                    | 0              | 1                     | 0                | 1                | 1                     | 1                | 1               | 1             | 1               | 1                | 1               | 1                | 0                 | 1             | 0                 | 13                  |               |
| German   | 1            | 1              | 1                    | 1              | 1                     | 0                | 1                | 1                     | 1                | 1               | 1             | 1               | 1                | 1               | 1                | 1                 | 1             | 0                 | 13                  |               |
| Hebrew   | 1            | 0              | 1                    | 0              | 1                     | 0                | 1                | 1                     | 1                | 0               | 1             | 1               | 1                | 1               | 1                | 1                 | 0             | 1                 | 17                  |               |
| Hungarian| 1            | 1              | 0                    | 0              | 1                     | 0                | 0                | 0                     | 0                | 1               | 1             | 0               | 1                | 1               | 1                | 1                 | 1             | 0                 | 13                  |               |
| Japanese | 0            | 0              | 0                    | 0              | 0                     | 1                | 0                | 0                     | 0                | 0               | 0             | 0               | 0                | 0               | 1                | 0                 | 0             | 1                 | 4                   |               |
| Korean   | 0            | 0              | 0                    | 0              | 0                     | 1                | 0                | 0                     | 0                | 0               | 0             | 0               | 0                | 0               | 1                | 0                 | 0             | 1                 | 4                   |               |
| Latin    | 0            | 1              | 1                    | 1              | 1                     | 0                | 1                | 1                     | 1                | 1               | 0             | 1               | 1                | 1               | 1                | 1                 | 1             | 0                 | 16                  |               |
| Russian  | 0            | 0              | 1                    | 1              | 1                     | 0                | 1                | 1                     | 1                | 0               | 1             | 1               | 1                | 1               | 1                | 1                 | 0             | 1                 | 16                  |               |
| Swahili  | 0            | 0              | 1                    | 1              | 1                     | 0                | 1                | 1                     | 1                | 1               | 1             | 1               | 1                | 1               | 1                | 1                 | 1             | 0                 | 17                  |               |
| Tagalog  | 0            | 0              | 0                    | 0              | 1                     | 0                | 0                | 0                     | 1                | 1               | 0             | 1               | 1                | 1               | 0                | 1                 | 0             | 1                 | 9                   |               |

---

N.B. All attributes encode general **tendancies** rather than strict **rules**.
<table>
<thead>
<tr>
<th>Language</th>
<th>has_definite</th>
<th>has_indefinite</th>
<th>N_en_codes_plurality</th>
<th>N_en_codes_cc</th>
<th>relative_cl_follows</th>
<th>has_postposition</th>
<th>has_prepositions</th>
<th>A_agrees_w_Npluralit</th>
<th>A_agrees_w_Nclass</th>
<th>A_agrees_w_Ncase</th>
<th>A_follows</th>
<th>O_must_follow</th>
<th>check_sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulgarian</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>12.7</td>
</tr>
<tr>
<td>Chinese</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Czech</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>English</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Finnish</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>French</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>German</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Hebrew</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Hungarian</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Japanese</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Korean</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Latin</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Russian</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Swahili</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Tagalog</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Count: 6 4 11 8 5 12 4 12 9 8 5 5 6
Average: 0.4 0.3 0.73 0.53 0.33 0.8 0.3 0.8 0.6 0.53 0.33 0.3 0.4
| V_agreees_ | V_filenames_ | V_requi | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ | V_filenames_ |
|------------|--------------|----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 1          | 0            | 0        | 0            | 0            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            |
| 0          | 0            | 0        | 1            | 0            | 0            | 0            | 0            | 1            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            |
| 1          | 1            | 0        | 0            | 0            | 0            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            |
| 0          | 0            | 0        | 0            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            |
| 1          | 1            | 0        | 0            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            |
| 1          | 0            | 1        | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            |
| 1          | 1            | 0        | 0            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            |
| 0          | 0            | 0        | 0            | 0            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            |
| 1          | 0            | 1        | 1            | 1            | 0            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            |
| 0          | 0            | 0        | 0            | 0            | 0            | 0            | 0            | 1            | 1            | 0            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            |
| 1          | 1            | 0        | 0            | 0            | 0            | 0            | 0            | 1            | 1            | 1            | 0            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            |
| 0          | 1            | 1        | 0            | 1            | 1            | 1            | 1            | 1            | 1            | 0            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            | 1            |
| 0          | 0            | 1        | 1            | 0            | 1            | 1            | 1            | 1            | 0            | 1            | 0            | 0            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            |
| 0          | 0            | 0        | 0            | 0            | 0            | 0            | 0            | 0            | 1            | 1            | 0            | 0            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 1            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            |
| 0          | 0            | 0        | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 1            | 0            | 0            | 0            | 1            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            | 0            |

13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13 13
Trends of the data (admittedly subject to sampling bias)

❖ All languages
  ❖ (A15) encode Verb for Voice [1.0]
❖ Most languages
  ❖ (A16) encode Verb for Tense. [0.9]
  ❖ (A8) have Prepositions. [0.8]
  ❖ (A18) require Verb to agree with Subject. [0.8]
  ❖ (A6) employ Relative clause which follow head Noun. [0.8]
  ❖ (A22) derive Infinitive from Bare Verb. [0.8]
  ❖ (A3) encode Noun for Plurality. [0.73]
  ❖ (A19) encode Verb for Person. [0.7]
  ❖ (A20) encode Verb for Plurality. [0.67]

❖ Few languages
  ❖ (A14) require Subject. [0.2]
  ❖ (A23) require Verb to agree with Object.* [0.2]
  ❖ (A15) have Postpositions. [0.3]
  ❖ (A24) employ Tense Agreement. [0.3]
  ❖ (A6) require Adj to follow N. [0.3]
  ❖ (A5) encode Noun for Case. [0.33]
  ❖ (A10) require Adj agree with Noun-class. [0.33]
  ❖ (A21) encode Verb for Subject Class. [0.33]
  ❖ (A1) have definite articles. [0.4]
  ❖ (A2) Fewer have indefinite articles. [0.3]

*OV languages are under-represented.
Concept Explorer 1.3 at Work
available at http://conexp.sourceforge.net/
Results

What results were obtained under what conditions.
Note

- This equals to Fig. 2 in the paper
- Red lines indicate “collisions” that appear when inconsistencies are detected in FCA.
- This is a feature of Concept Explorer 1.3.
FCA 0
— enlarged
Idea for optimization

❖ Optimization is necessary.
❖ Unrestricted FCA doesn’t tell much about how trade-offs in grammar are resolved or “compromised.”

❖ 3 counteracting conditions for good FCA
❖ A Hesse diagram is good if
❖ **Condition 1)** objects are as much separated as possible, but

❖ **Condition 2)** there are as few empty nodes as possible, and
❖ **Condition 3)** the diagram is in a geometrically good shape.

❖ Caveat
❖ **Condition 3** is admittedly subjective and even esthetic, but it’s not bad in itself
❖ unless tools for FCA are provided with algorithms for optimization.
Monte Carlo procedure for optimization

- Procedure for optimal selection of attributes
  - Start with the state in which all attributes are unselected.
  - Select $n$ attributes randomly and check the result.
  - Roughly, $0 < n < 5$
  - If the result looks bad, undo the last selection to get a better result.

- If not, select the next $n$ attributes randomly, and check the result.
- Stop selection if any better result can be obtained.

- Conditions
  - In this case, all objects are trusted. If this is not the case, the same procedure needs to be applied to the selection of objects.
Conflations:
- None
- 5 empty nodes are allowed.
- Layout is symmetrical.
- equals to Fig. 3 in the paper

Used attributes:
- to be shown latter
16 attributes used in Optimization 1

- A1 has definite article
- A2 has indefinite article
- A3 N encodes plurality
- A4 N encodes class
- A6 Relative clause follows N
- A8 has prepositions
- A9 A agrees with N-plurality
- A10 A agrees with N-class
- A12 A follows N
- A14 requires Subject
- A15 V encodes Voice
- A16 V encodes Tense
- A18 V agrees with Subject
- A19 V encodes Person
- A20 V encodes Plurality
- A21 V encodes N-class
8 attributes discarded in Optimization 1

- The following 8 attributes turned out to be offensive.
  - A5 N encodes Case
  - A7 has Postpositions
  - A11 A agrees with N-case [missed in the paper]
  - A13 O must follow V
  - A17 V encodes Aspect
  - A22 V infinitive is derived
  - A23 V agrees with Object
  - A24 has Tense agreement
In my view, **Optimization 1 deserves the best** in the following reason, though the claim is admittedly debatable:

- While it contains 5 empty nodes (condition 2 violated),
- object classification is good enough (condition 1 well observed) and,
- layout is symmetrical enough (condition 3 well observed).

**Esthetics**

- I observed condition 1 strictly, and I ranked condition 3 higher than condition 2.
Under this hypothesis, the “convergent” and “divergent” classes of attributes were separated.

- the former comprises 16 attributes and the latter 8 attributes.

**Bonus**

- The optimization revealed 3 correlations among convergent attributes (to be shown later).
- The optimization revealed 7 implications among convergent attributes (to be shown later).
What FCA 1 tells us about the nature of grammar?
3 correlations among effective attributes

❖ Two attributes, \( A_4 \text{N} \) encodes Class and \( A_{10} \text{A} \) agrees with N-class, correlate, if not equivalent.

❖ Two attributes, \( A_{19} \text{V} \) encodes Person, and \( A_{20} \text{V} \) encodes Plurality, correlate, if not equivalent.

❖ Two attributes \( A_6 \) Relative clause follows N, and \( A_{18} \text{V} \) agrees with Subject, correlate, if not equivalent.
8 implications

1. A2 has Indefinite Article is a precondition for A14 requires Subject.

2. A1 has Definite Article is a precondition for A2 had Indefinite Article.

3. A9 A agrees with N-plurality is a precondition for A4 N encodes Class and A10 A agrees with N-class.

4. A20 V encodes Plurarily is a precondition for A4 N encodes Class, A9 A agrees with N-plurality, and A10 A agrees with N-class.

5. A19 V encodes Person and A3 N encodes Plurality are a precondition for A20 V encodes Plurality.

6. A8 has Prepositions is a precondition for A14 requires Subject, A9 A agrees with N-plurality, A12 A follows N, and A21 V encodes N-class.

7. A15 V encodes Voice and A6 Relative clause follows N are a precondition for A16 V encodes Tense, A3 N encodes Plurality, A12 A follows N, and A18 V agrees with Subject.

8. A16 V encodes Tense is a precondition for A19 V encodes Person and A3 N encodes Plurality.
The presented results have obvious bearings on Greenberg’s Language Universals.

But my results are more informative in that they give us something like geometry of possible grammars, thereby helping us to define grammar types.
FCA meets grammar typology at NLP 21

Comparison with other optimizations
FCA 2
Optimization 2

- **Note**
  - This equals to Fig. 4 in the paper

- **Conflations:**
  - None

- 4 empty nodes are allowed
  - at the expense of Finnish discriminability

- **Layout is fairly symmetrical.**

- **Difference from FCA 1:**
  - **A20** removed
FCA 3
Optimization 3

- **Note**
  - This equals to Fig. 5 in the paper
- **Conflations:**
  - None
- **3 empty nodes are allowed.**
- **Layout is fairly symmetrical.**
- **Difference from FCA 1:**
  - A1, A19, and A20 removed
FCA 4
Optimization 4

- **Note**
  - This equals to Fig. 6 in the paper

- **Conflations:**
  - \{Swahili, Russian, Czech\}, \{German, French\}

- 2 empty nodes are allowed.

- Layout is less symmetrical.

- **Difference from FCA 1**
  - A1, A9, A12, and A20 removed
Note
- No presentation was made in the paper.

Conflations:
- \{Swahili, Hebrew, Bulgarian\}, \{Latin, German\}

1 empty node is allowed.

Layout is less symmetrical.

Difference from FCA 1:
- A3, A4, A5, A6, A7, A8, A9, A10, A11, A15, A18, A19, and A20 removed
FCA 6
Optimization 6

- **Note**
  - This equals to Fig. 7 in the paper

- **Conflations:**
  - \{Russian, Latin, German, Czech\}, \{Swahili, Hebrew, French, Bulgarian\}

- **No empty node is allowed.**

- **Layout is less symmetrical.**

- **Difference from FCA 1**
  - A3, A4, A5, A6, A7, A8, A9, A10, A11, A15, A16, A18, A19, and A20 removed
Which language is most similar to Japanese in terms of grammar?

- The obvious but uninteresting answer:
  - Korean
    - which can be reached without moving around.

- More interesting answers:
  - Hungarian and Finnish
    - which can be reached without very deep descending.
  - Chinese
    - which can be reached without descending.
FCA meets grammar typology at NLP 21

Discussion
Relativized learnability index

- We can reasonably predict that, other things being equal, descending the Hasse diagram poses more difficulty in learning. This defines relativized learnability index for grammar.

- Examples
  - If a learner speaks a language without person-agreement on verbs and plurality-encoding on nouns, it would pose a handicap in his or her learning.
  - In general, learners will face more difficulty if their mother tongue is one of the agreement-free languages.
A vision for more effectively English instruction

Question

What is the most serious handicap for those who speak Japanese natively?

Answer

Japanese is a language that lacks two dominant attributes A3 N encodes Plurality and A19 V encodes Person, which are shared by a large portion of languages investigated.

In more detail, A3 N encodes Plurality is a precondition for A20 V encodes Plurality, which makes a precondition for A19 V encodes Person.
Suggestion

I contend that the lack of A3 and A19 forms the greatest barrier that blocks access to learning a wide range of languages.

Differently understood, however, drastic improvement in English education for the Japanese can be possible (only) if learning methods are developed to help the Japanese to acquire the two attributes effectively.
Caveat on the nature of representation

- Grammar types are represented, forcefully, as discrete objects, but we are strongly discouraged to take this at its face value.
- Grammar types are best understood as “attractors” in a dynamical system, in analogy with “niches” over a “fitness” landscape, on the assumption that what the Hasse diagrams represent needs to be understood in terms of probability.
  - Categories like N, V and A are abstractions. In reality, each of them subsumes a group of words that behave differently.
  - The operational definition Case is problematic, to say the least.
  - It is not clear how far the notion Noun class should cover.
- In terms of game theory, grammar types are Nash equilibria in the game of cost-benefit trade-off between speaker and hearer.
Why divergent attributes?

Two different sources of disturbance need to be recognized:

- involvement of definitional/phenomenological problems
- involvement of architectural/systematic problems (leading to conflicts, or trade-offs)

Reasons for the former:

- After a number of experiments, it turned out that attributes mentioning Case and Postposition are offensive and tend to generate inconsistencies.

(Possible) reasons for the latter

- (Grammar of a) language is very likely to be a “system of trade-offs” that involves counterbalancing a large number of costs and benefits.
Future directions

❖ Scale up, scale up, scale up!
❖ A set of 15 language is too small.
❖ In one estimation, 6,000 languages exist.
❖ But how?
❖ Use World Atlas of Language Structure (WALS)
❖ http://wals.info
❖ and automate the setup?
Summary

❖ Data and Analysis

❖ 15 languages are selected and manually encoded against 24 grammatical/morphological features.

❖ Formal Concept Analysis (FCA) was performed against a formal context with the 15 languages as objects and the 23 features as attributes.

❖ Results

❖ A series of experiments suggested a few optimal results, one of which I expect is informative enough to define relativized learnability index.

❖ Comparison between optimal and suboptimal FCA’s was revealing in typological studies of language.
Thank you for your attention