Areality in colexification patterns

Maria Koptjevskaja-Tamm, ¹ Volker Gast²

¹Stockholm, ²Jena & Olomouc

28 June 2018
General context: Areality in lexical typology (e.g. Koptjevskaja-Tamm and Liljegren 2017)\(^1\)

Large-scale convergence patterns in the organization of the lexicon, with a focus on colexification.\(^2\)

Using data from two major lexical databases,

- The Crosslinguistic Database of Colexifications (CLICS),\(^3\) and
- the Automated Similarity Judgment Programme (ASJP).\(^4\)


Overview

1. Introduction

2. Methodology: Identifying areal clusters of colexification patterns

3. Some results: ASJP
   - Nature
   - Metaphorical extensions of body parts

4. Some results: CLICS
   - Cultural artifacts
   - <language, x>
   - <easy, difficult>

5. Some questions
Areal lexico-semantic patterns

- Types of areal lexico-semantic patterns:
  - lexico-semantic parallels (colexification patterns, lexico-constructional patterns),
  - shared formulaic expressions,
  - area-specific lexicalizations.
  
  (cf. Koptjevskaja-Tamm and Liljegren 2017):^5

- Colexification pairs as the ‘atomic’ concept of (lexical) semantic maps.

---

The data

- **CLICS**: 92,805 data points
- **ASJP**: 2,060,856 data points

Each data point is a quadruple of the form \(<C_1, C_2, L, B>\)

- \(C\): concept,
- \(L\): language,
- \(B\): Boolean, TRUE or FALSE

Example: \(<'arm', 'hand', 'russ1263', TRUE>\)

Some negative evidence could be inferred.

Note that the concept of a binary distinction between colexification and differentiation is of course simplifying (e.g. ‘arm’, ‘hand’ in Russian).
The starting point

Figure 1: Colexification (black squares) vs. differentiation (red circles) of ‘feather’ and ‘hair’ in the ASJP-data
Introduction
Methodology: Identifying areal clusters of colexification patterns
Some results: ASJP
Some results: CLICS
Some questions

Identifying geographical clusters

Figure 3: Dendrogram for the colexification pair <‘feather’, ‘hair’> (meso-clusters)
Identifying geographical clusters

Figure 4: Hypothesized meso-clusters for the colexification pattern <‘feather’, ‘hair’>
Testing the clusters

- For each cluster, a (Bayesian) mixed effects model was fitted (cf. Hadfield 2010, Levshina 2010).
- Language family was treated as a random effect.
- Moreover, a genealogical diversity index was calculated for each cluster (entropy).

---


Two clusters

Figure 5: Cluster Areas 1 and 2 for the colexification pair <"feather", "hair">
Genealogical diversity in the clusters

Figure 6: Clusters 1 and 5 for <‘feather’, ‘hair’>
Some results: ASJP

- We identified the clusters that were ‘strongest’ in terms of the regression statistics and genealogical diversity.
- Strongest areal clusters for the ASJP-data:
  - <fire, tree>
  - <mountain, stone>
  - <ear, leaf>
  - <bark, skin>
Introduction
Methodology: Identifying areal clusters of colexification patterns
Some results: ASJP
Some results: CLICS
Some questions

Nature
Metaphorical extensions of body parts

<fire, tree>

Figure 8: Significant meso-cluster for <‘fire’, ‘tree’>
<fire, tree>: Discussion

Lexical-semantic relations between ‘fire’ and ‘tree’ is a well-known feature of the languages of Australia, New Guinea and the surrounding islands, cf. Schapper et al. (2016), Östling (2016).

---


<fire, tree> in Sahul

- Schapper et al. (2016)
- ASJP
- Östling (2016)
Schapper et al. (2016): Frequencies

Most, but not all, languages colexifying ‘tree’ and ‘fire’ also exhibit colexification of both concepts with ‘firewood’ (31/38) (which is not on the Swadesh list).

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Description</th>
<th>Papuan languages</th>
<th>Australian languages</th>
<th>TOTAL</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>abc</td>
<td>full differentiation</td>
<td>92</td>
<td>10</td>
<td>102</td>
<td>34%</td>
</tr>
<tr>
<td>abb</td>
<td>firewood/fire colexification</td>
<td>70</td>
<td>64</td>
<td>134</td>
<td>45.5%</td>
</tr>
<tr>
<td>aab</td>
<td>tree/firewood colexification</td>
<td>26</td>
<td>0</td>
<td>26</td>
<td>8.5%</td>
</tr>
<tr>
<td>aba</td>
<td>firewood differentiation</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td>2%</td>
</tr>
<tr>
<td>aaa</td>
<td>full colexification</td>
<td>22</td>
<td>9</td>
<td>31</td>
<td>10%</td>
</tr>
</tbody>
</table>
Schapper et al. (2016): Frequencies

- ‘Subcolexification’: the ‘primary lexifier’ is the same, e.g. Daga *oma* ‘tree, fire’, *oma oaewa* ‘firewood’.
- The number of patterns is reduced.

<table>
<thead>
<tr>
<th></th>
<th>Papuan languages</th>
<th>Australian languages</th>
<th>Patterns of subcolexification</th>
</tr>
</thead>
<tbody>
<tr>
<td>abc</td>
<td>full differentiation</td>
<td>66</td>
<td>2</td>
</tr>
<tr>
<td>abb</td>
<td>firewood/fire colexification</td>
<td>80</td>
<td>52</td>
</tr>
<tr>
<td>aab</td>
<td>tree/firewood colexification</td>
<td>33</td>
<td>2</td>
</tr>
<tr>
<td>aaa</td>
<td>full colexification</td>
<td>38</td>
<td>7</td>
</tr>
</tbody>
</table>
<fire, tree>: Discussion

- Lexical-semantic relations:
  
  TREE – FIREWOOD – FIRE

- Where does ‘wood’ (e.g. as construction material) come in?

The vast majority of Sahul languages do not have a distinct lexeme for ‘wood’; this is colexified with the lexeme used for English ‘tree’ . . . In 168 of the Papuan languages surveyed ‘tree’ and ‘wood’ were identically colexified (for 37 languages no translation for ‘wood’ was found in the source used). In only 11 languages were they recorded as non-identical and in three of these there appeared to be a possible derivational relationship between the two lexemes. In the Australian languages only 17 of the 63 languages are recorded as having a word for ‘wood’ or ‘stick’ that is different from the word for ‘tree’ and in all but six of these there exists a colexified word for ‘wood’/‘tree’ as well. In 11 Australian languages no translation for ‘wood’ or ‘stick’ was available. (Schapper et al. 2016: 364/5, our emphasis)
Cohexification of ‘wood’ and ‘tree’
‘Wood’ and ‘tree’

- Wiktowski et al. (1981)
  - have claimed that words for ‘tree’ are originally derived from wood, rather than being generalizations over various tree species.
  - Put differently, generic words for ‘tree’ are primarily associated with the material (wood), rather than ‘tree’ as a plant species.

---

Is there a general tendency for languages (of a specific area) to encode the ‘raw material’ and the product made from it with the same word?

“Some – but by no means all – Australian languages take the principle of having a single term to describe some natural object, and also something that can be made from it, to the extreme of having a single lexeme covering both ‘tree, wood’ and ‘fire’.” (Dixon 1980: 103)\textsuperscript{11}

‘Actual’ and ‘potential’ polysemy?

What has been called “actual/potential polysemy” is ubiquitous in Australian languages (cf. Dixon 1980:102-103): a large number of languages exhibit such polysemes as ‘cloud/rain’; ‘firewood/fire’; ‘breast/milk’; ‘animal/meat’, ‘tree/wood/implement’, ‘bush/bushfire’, ‘hit/kill’, ‘sick/dead’, ‘dead/rotten’. Significantly, in most Australian languages the word for ‘make’ is the factitive of ‘good’ (e.g. Warlpiri *ngurrju* ‘good’, *ngurrjumani* ‘make’). Most of the physical world is regarded, with respect to the transformations that can be worked on it, in the same way that Michelangelo is said to have regarded blocks of marble: as already containing the form of David, etc. Successive phases of potential transformation are then named with the same term.

---

<mountain, stone>

Figure 9: Significant macro-clusters for <‘mountain’, ‘stone’> in the ASJP-data

Nature
Metaphorical extensions of body parts
<mountain, stone> in Australia
Climatic zones of Africa
Aso Rock (near Abuja, the capital of Nigeria)
The Kalahari basin (Khoisan)
Alternative colexification partners for ‘mountain’

- A certain affinity between ‘mountain’ and ‘stone’ can also be observed in Indo-European languages, e.g. Goth. *hallis* ‘rock’, Old Norse *hallr* ‘large stone’, cf. Lat. *collis*; see Buck 1949: Sect. 1.22).\(^{13}\)

- Alternatively, ‘mountain’ is often colexified with ‘forest’, or ‘mountain forest’, e.g. (Latin American) Spanish *selva* (cf. also Urban 2012).\(^{14}\)

---


Introduction
Methodology: Identifying areal clusters of colexification patterns
Some results: ASJP
Some results: CLICS
Some questions

Nature
Metaphorical extensions of body parts

Catamarca (Argentina)
Introduction

Methodology: Identifying areal clusters of colexification patterns

Some results: ASJP
Some results: CLICS

Some questions

Nature

Metaphorical extensions of body parts

<ear, leaf>

Figure 10: Significant meso-clusters <‘ear’, ‘leaf’> in the ASJP-data
<ear, leaf> in the Nilotic-Surmic spread zone (Nilo-Saharan?)

<ear, leaf>: Vegetation (ensete/Ethiopian banana)

- Sheko (shek1245): *haay* ‘ear’, ‘leaf of ensete or yam’
  (Helltenthal 2010: 493)\(^\text{16}\)

<ear, leaf>: Vegetation (agave)
Introduction
Methodology: Identifying areal clusters of colexification patterns
Some results: ASJP
Some results: CLICS
Some questions

Nature
Metaphorical extensions of body parts

Figure 12: Macro-clusters for ‘bark’, ‘skin’ (ASJP-data)
<bark, skin>: Discussion

- The colexification of ‘bark’ and ‘skin’ has been discussed as a potential lexical trait of Mesoamerican languages (Smith-Stark 1994).\(^{17}\)
- The actual overlap between these concepts is actually more widespread in Mesoamerica, but many of the relevant cases are instances of ‘partial colexification’.
- Are such areal patterns reflexes of a more general tendency towards metaphorical transfer between humans, animals and plants?

\[\text{(1) Copainalá Zoque} \]
\[kuʔyu naka\]
\[tree skin\]
\[(Harrison et al. 1981: 280)\(^{18}\)]


Some results: CLICS

Some interesting clusters:

<feather, pen>
<money, x>
<language, x>
<easy, difficult>
Introduction
Methodology: Identifying areal clusters of colexification patterns
Some results: ASJP
Some results: CLICS
Some questions

Cultural artifacts
<language, x>
<easy, difficult>

<feather, pen>
Introduction
Methodology: Identifying areal clusters of colexification patterns
Some results: ASJP
Some results: CLICS
Some questions

Cultural artifacts
"language, x"
"easy, difficult"

"money, silver"
Introduction
Methodology: Identifying areal clusters of colexification patterns
Some results: ASJP
Some results: CLICS
Some questions

Cultural artifacts
<language, x>
<easy, difficult>

<money, coin>
Cultural artifacts: Discussion

Terms for cultural artifacts such as ‘pen’ and ‘money’ often seem to form relatively well defined clusters of limited size.

In comparison to other concepts, cultural artifacts are relatively young and were often borrowed or calqued.

Types of ‘lexical acculturation’ (Brown 1999):
- adoption of loan words,
- borrowing a lexico-constructional pattern,
- extension of native terms,
- coining of new expressions.

Colexification typically seems to result from the extension of native terms (‘money’ as ‘gold’/‘silver’, ‘shells’/‘cowry’).

In some cases colexification may be the result of real-world developments in cultural practices (‘pen’ as ‘feather’, ‘money’ as ‘silver’).

---

Introduction
Methodology: Identifying areal clusters of colexification patterns
Some results: ASJP
Some results: CLICS
Some questions

Cultural artifacts

<language, x>
<easy, difficult>

<language, tongue>
Introduction
Methodology: Identifying areal clusters of colexification patterns
Some results: ASJP
Some results: CLICS
Some questions

Cultural artifacts

<language, x>
<easy, difficult>

<language, word>
Introduction
Methodology: Identifying areal clusters of colexification patterns
Some results: ASJP
Some results: CLICS
Some questions
Cultural artifacts
<language, x>
<easy, difficult>

<language, voice>
Radden (2004): Terms for language as the result of metonymic change
speech organ > speaking > speech > language

Europe and the Caucasus: instrument-for-action metonymy

South America: action-for-result metonymy (also specific-for-generic)

Comparison to cultural artifacts: much larger clusters

---

The concept of ‘language’

- Goddard (2010): The term ‘language’ is highly polysemous.
- In expressions such as ‘the English language’ it has “ideological underpinnings” (Goddard 2010: 44) which are incompatible with certain communities (e.g. of Aboriginal Australia).
- If a community lacks the ideological concept of ‘language’, the action or the result of speaking (‘speech’, ‘word’) will be used as the closest translation equivalent.
- Do areal patterns in colexification reflect such ‘ideological’ differences?

---

Introduction
Methodology: Identifying areal clusters of colexification patterns
Some results: ASJP
Some results: CLICS
Some questions

Cultural artifacts
<language, x>
<easy, difficult>

<difficult, heavy>
Introduction
Methodology: Identifying areal clusters of colexification patterns
Some results: ASJP
Some results: CLICS
Some questions

Cultural artifacts
<language, x>
<easy, difficult>

<easy, light>
Colecification partners of ‘difficult’ and ‘easy’ (CLICS)
<table>
<thead>
<tr>
<th>Nr.</th>
<th>Language</th>
<th>ISO</th>
<th>Family</th>
<th>Source</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Avar (Zakataly)</td>
<td>ava</td>
<td>North Caucasian</td>
<td>IDS</td>
<td>бакла̀б</td>
</tr>
<tr>
<td>2</td>
<td>Botlikh</td>
<td>bph</td>
<td>North Caucasian</td>
<td>IDS</td>
<td>гъакъу̀ба</td>
</tr>
<tr>
<td>3</td>
<td>Bulgarian</td>
<td>bul</td>
<td>Indo-European</td>
<td>IDS</td>
<td>téžak</td>
</tr>
<tr>
<td>4</td>
<td>German, Standard</td>
<td>deu</td>
<td>Indo-European</td>
<td>IDS</td>
<td>schwer</td>
</tr>
<tr>
<td>5</td>
<td>Sorbian, Lower</td>
<td>dsb</td>
<td>Indo-European</td>
<td>WOLD</td>
<td>šęžki</td>
</tr>
<tr>
<td>6</td>
<td>Estonian</td>
<td>est</td>
<td>Uralic</td>
<td>IDS</td>
<td>řaske</td>
</tr>
<tr>
<td>7</td>
<td>Hungarian</td>
<td>hun</td>
<td>Uralic</td>
<td>IDS</td>
<td>neheez</td>
</tr>
<tr>
<td>8</td>
<td>Judeo-Tat</td>
<td>jdt</td>
<td>Indo-European</td>
<td>IDS</td>
<td>гурунг</td>
</tr>
<tr>
<td>9</td>
<td>Khanty</td>
<td>kca</td>
<td>Uralic</td>
<td>IDS</td>
<td>lawaft</td>
</tr>
<tr>
<td>10</td>
<td>Ket</td>
<td>ket</td>
<td>Yeniseian</td>
<td>WOLD</td>
<td>сээ</td>
</tr>
<tr>
<td>11</td>
<td>Khvarshi (Khvarshi)</td>
<td>khv</td>
<td>North Caucasian</td>
<td>IDS</td>
<td>луйт̱у́</td>
</tr>
<tr>
<td>12</td>
<td>Kilivila</td>
<td>kij</td>
<td>Austronesian</td>
<td>IDS</td>
<td>гра́н</td>
</tr>
<tr>
<td>13</td>
<td>Karata</td>
<td>kpt</td>
<td>North Caucasian</td>
<td>IDS</td>
<td>гъаркъибоб</td>
</tr>
</tbody>
</table>

Found 32 colexifications for "difficult" and "heavy". Note that the number of attested colexifications may differ from the number of languages in which the colexifications were attested.
**Colexification of** `<difficult, hard>`

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Language</th>
<th>ISO</th>
<th>Family</th>
<th>Source</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Negerhollands</td>
<td>dcr</td>
<td>Creole</td>
<td>IDS</td>
<td>tai</td>
</tr>
<tr>
<td>2</td>
<td>Emberá, Northern</td>
<td>emp</td>
<td>Choco</td>
<td>IDS</td>
<td>za’rea</td>
</tr>
<tr>
<td>3</td>
<td>Iraqw</td>
<td>irk</td>
<td>Afro-Asiatic</td>
<td>WOLD</td>
<td>gawid</td>
</tr>
<tr>
<td>4</td>
<td>Marathi</td>
<td>mar</td>
<td>Indo-European</td>
<td>SPRÅKBANKEN</td>
<td>katʰi:ŋ</td>
</tr>
<tr>
<td>5</td>
<td>Macushi</td>
<td>mbc</td>
<td>Carib</td>
<td>IDS</td>
<td>sa?me</td>
</tr>
<tr>
<td>6</td>
<td>Quichua, Imbabura Highland</td>
<td>qvi</td>
<td>Quechuan</td>
<td>WOLD</td>
<td>sinchi</td>
</tr>
<tr>
<td>7</td>
<td>Tarifit</td>
<td>rif</td>
<td>Afro-Asiatic</td>
<td>WOLD</td>
<td>qṣāḥ</td>
</tr>
<tr>
<td>8</td>
<td>Saramaccan</td>
<td>srm</td>
<td>Creole</td>
<td>WOLD</td>
<td>taáŋa</td>
</tr>
<tr>
<td>9</td>
<td>Sirionó</td>
<td>srq</td>
<td>Tupi</td>
<td>IDS</td>
<td>eātā</td>
</tr>
<tr>
<td>10</td>
<td>Swahili</td>
<td>swh</td>
<td>Niger-Congo</td>
<td>WOLD</td>
<td>gumu</td>
</tr>
<tr>
<td>11</td>
<td>Waurá</td>
<td>wau</td>
<td>Arawakan</td>
<td>IDS</td>
<td>kehežuka</td>
</tr>
</tbody>
</table>
The <difficult, heavy> colexification is also found in the Middle East and (North-)East Africa, e.g. (cf. Segerer and Flavier 2011–2018):²²

- Hebrew kaved/kveda
- Arabic taqīl
- Amharic kābbad
- Sheko ints

Some languages of West Africa (Senegal) seem to have it, too, e.g.:
- Soninke (Mande)
- Joola-Manjaku, Joola-Fonyi (Atlantic-Congo)

The metaphorical extension of ‘heavy’ to ‘difficult’ is regarded as a ‘primary metaphor’ in cognitive linguistics, as is the extension from ‘hard’ to ‘difficult’ (assumed universality, cf. Grady 2005). But then, the distribution of <hard, difficult> seems to exhibit heavy areal (as well as genealogical) biases.

Hypothesis: There are various (universal) options of conceptualizing ‘difficulty’, and there is no obvious motivation for preferring any of them (e.g. ‘heavy’ vs. ‘hard’).

Note that metaphorical extensions of this type may, to some extent, be externally motivated, cf. ‘warm’ as a positive emotional feeling.

Does the (cultural) ‘arbitrariness’ of metaphorical transfer of this type give language contact more weight?

---

Limitations of our approach

- Adjectives such as ‘difficult’ are complex semantically speaking and often imply metonymic shifts, e.g.:
  - This is a difficult task.
  - He’s a difficult person.
  - Those are difficult times.
  - etc.

- Some words, such as Russ. *tâželyj*, can only be used with the meaning ‘difficult’ in specific contexts.
‘Difficult’ and ‘easy’ in German and Czech

- The adjectives of Czech are distributed differently from those of German, insofar as *těžký* shows a clear affinity to temporal head nouns (in the data investigated).

(2) ...I think we have a person who has already shown in those very difficult days in Pakistan how responsibly he has been taking on these things ...

(3) ...máme myslím člověka, který již během těchto velmi těžkých dní v Pákistánu ukázal, jak zodpovědně k těmto věcem přistupuje ...
Convergence areas and types of lexical meaning

- What motivates colexification patterns in an areal perspective?
- What general tendencies can we observe with respect to types of lexical meanings and convergence areas (also in terms of size)?
- Cultural artifacts seem to form relatively clear-cut areal clusters, perhaps because of a relatively shallow time depth.
- Some colexification patterns seem to be related to (physical) properties of the environment.
- Are there general (areal) tendency towards specific types of metaphorical transfer, and perhaps “actual/potential polysemy” (Evans 1992)?
- Abstract concepts are sometimes associated with larger cluster areas.
- Problem: Words as elicitation stimuli/terra comparationis may actually be misleading (‘language’).
Thanks for listening!

- Any suggestions?