The Nijmegen Typological Survey (NTS)

Harald Hammarström & Hedvig Skirgård

2014-11-25
Nijmegen Typological Survey

Original questionnaire designed for studies of the phylogeny of languages of Sahul and Melansia by Ger Reesink, Michael Dunn et al


Extensions, clarifications and new coding of African lgs was later added by Harald Hammarström, Suzanne van der Meer, Jeremy Collins and Hedvig Skirgård in 2013-2014

Part of Cross-Linguistic Linked Data-project (CLLD), so it will be freely available online etc
The NTS-sample: languages

- follows the ISO 639-3 of language names
- a language in NTS ≠ doculect (Cysouw & Good 2013)
- several sources and one coder per language

Languages by area

<table>
<thead>
<tr>
<th>Region</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papuan</td>
<td>165</td>
</tr>
<tr>
<td>Australian</td>
<td>25</td>
</tr>
<tr>
<td>Africa</td>
<td>70</td>
</tr>
<tr>
<td>Eurasia</td>
<td>11</td>
</tr>
</tbody>
</table>
### The NTS-sample: features

<table>
<thead>
<tr>
<th>Feature Description</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Features from Reesink et al (2009)</td>
<td>204</td>
</tr>
<tr>
<td>New features</td>
<td>125</td>
</tr>
<tr>
<td>Total features</td>
<td>329</td>
</tr>
<tr>
<td>Features for African languages from Reesink et al (2009)</td>
<td>141</td>
</tr>
<tr>
<td>New features</td>
<td>125</td>
</tr>
<tr>
<td>Total features for African languages</td>
<td>266</td>
</tr>
<tr>
<td>Complementing previous features</td>
<td>64</td>
</tr>
<tr>
<td>Inspired by WALS</td>
<td>32</td>
</tr>
<tr>
<td>Inspired by Di Garbo (2014)</td>
<td>19</td>
</tr>
<tr>
<td>Africa-specific</td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td>8</td>
</tr>
</tbody>
</table>
The African set

Afro-asiatic
Atlantic-Congo
Dizoid
Heiban
Ijoid
Khoe-kwadi
Koman
Kuliak
Mande
Nilotic
Nubian
Songhay
Design of the questionnaire

• remember
  – several sources and one coder per language
  – in NTS ≠ doculect (Cysouw & Good 2013)

• Working with a survey of this kind involves a lot of discussions on definitions and the criteria of categories etc.

<table>
<thead>
<tr>
<th>Values</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>blank</td>
<td>No information</td>
</tr>
<tr>
<td>?</td>
<td>Not enough information to code</td>
</tr>
<tr>
<td>n/a</td>
<td>Depending on other feature value which is 0</td>
</tr>
<tr>
<td>0/1/2/3/4/5</td>
<td>Enough information</td>
</tr>
</tbody>
</table>

• of all features of the NTS, only five are multivalue
• comparative work is difficult because of different traditions of analysis and terminology
• consistency > “true” categories
Finding answers

• consulting already existing descriptions
  (overview based Glottolog.org & conversing with experts)

• problems
  • conflicting analyses
  • certain features take longer time than others
  • absence of evidence ≠ evidence of absence
  • finding only what linguists thought to look for

• solution (in part) = consulting language-specific researchers and/or speakers
  • more details on conflicting descriptions
  • advice on “good” and “bad” sources
  • sometimes reliable evidence of absence
  • confirmation of potentially unusual pattern
Thanks to

Nicholas Rolle (Berkley), Anna Marie Diagne (Cheikh Diop Uni in Dakar), Doris Richter (Cologne University and Radboud University), Don Killian (Helsinki University), Viktoria Apel (Humboldt-Berlin), Jenia Gutova (Leiden Uni), Alexandra Vydrina (LLACAN), Amina Mettouchi (LLACAN), Daria Mishchenko (LLACAN), Elena Perekhvalskaya, (LLACAN), Guillaume Segerer (LLACAN), Maria Khachaturyan (LLACAN), Paulette Roulon-Doko (LLACAN), Tatiana Nikitina (LLACAN), Valentin Vydrin (LLACAN), Mark Dingemanse (MPI Nijmegen), Simeon Floyd (MPI Nijmegen), Rebecca Defina (MPI Nijmegen), Saskia van Putten (MPI Nijmegen), Ronald Schaefer (Southern Illinois University Edwardsville), Olga Kuznetsova (St petersburg, Russian academy of Sciences), Maria Konoshenko (St. Petersburg State University), Francesca di Garbo (Stockholm University), Yvonne Agbetsoamedo (Stockholm University), Becky Paterson (Univeristy of Oregon), Hugh Paterson (Univeristy of Oregon), Denis Creissels (University de Lyon), Frank Seidel (University of Florida), Francis Ekboghare (University of Ibadan), Serge Sagna (University of Surrey), Vera Wilhemsen (Uppsala University), Jean-Leopold Diouf, Desalegn Hagos Asfawwesen (Stockholms uni), Gerrit J. Dimmendaal (Cologne), Tom Güldemann (Humboldt-Berlin), Maarten Kossmann (Leiden Univeristy), Maarten Mous (Leiden Univeristy), Christian Rapold (Leiden/Regensburg), Martine Vanhove (LLACAN), Yvonne Treis (LLACAN), Felix Ameka (Leiden Uni), Gertie Hoymann (MPI Nijmegen), Liza Kushnir (Sholokhow Moscow state univeristy for the humanities), Gerard Dumestre (LLACAN), Solace Yankson (Radboud University), Martin Kohlberger (Leiden)
Some of these fantastic people
## A few typological databases

<table>
<thead>
<tr>
<th></th>
<th># languages</th>
<th># features</th>
<th>datapoints</th>
<th>features with over 200 lgs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic Structures of the World's Languages (2009)</td>
<td>237</td>
<td>93</td>
<td>14 440</td>
<td>?</td>
</tr>
<tr>
<td>World Atlas of Language Structures (2013)</td>
<td>2 679 (192)</td>
<td>165 (76 465)</td>
<td>69 590</td>
<td>129</td>
</tr>
<tr>
<td>Nijmegen Typological Survey (2014)</td>
<td>274</td>
<td>279</td>
<td>47629</td>
<td>145</td>
</tr>
<tr>
<td>Atlas of Pidgin and Creole Language Structures (2013)</td>
<td>76</td>
<td>130</td>
<td>20 624</td>
<td>N/A</td>
</tr>
<tr>
<td>Phonetics Information Base and Lexicon (2012)</td>
<td>1 010</td>
<td>1 680</td>
<td>75 386</td>
<td>N/A</td>
</tr>
<tr>
<td>South American Indigenous Language Structures (2014)</td>
<td>167</td>
<td>604</td>
<td>31 794</td>
<td>49</td>
</tr>
<tr>
<td>Automated Similarity Judgment Program (ASJP)</td>
<td>4424</td>
<td>N/A</td>
<td>“238 976”</td>
<td>N/A</td>
</tr>
</tbody>
</table>
## Discrete categories of continua

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numeral one</td>
<td>Indefinite/non-specific article</td>
<td></td>
</tr>
<tr>
<td>Demonstrative</td>
<td>Copula</td>
<td></td>
</tr>
<tr>
<td>Lexical verb</td>
<td>Auxiliary</td>
<td>Affix</td>
</tr>
<tr>
<td>???</td>
<td>Auxiliary</td>
<td>Affix</td>
</tr>
<tr>
<td>Body part</td>
<td>Reflexive pronoun</td>
<td>Middle voice</td>
</tr>
<tr>
<td>Definiteness marker</td>
<td>Noun class/gender</td>
<td>Noun classifier</td>
</tr>
<tr>
<td>Noun classifier</td>
<td>Numeral classifier</td>
<td></td>
</tr>
<tr>
<td>Phrasal verbs</td>
<td>“Verb particles”</td>
<td>Light verb constructions</td>
</tr>
<tr>
<td>Co-verb constructions</td>
<td>Copula</td>
<td></td>
</tr>
</tbody>
</table>
The feature set of NTS

Phonology

Word order

Polar questions

Non-verbal/stative predication
  attributive property predication
  nominal predication
  possessive predication
  existential/locative predication
  (adpositions)
  comparative predication

Negation

Valency
  trans -> intrans
  intrans-> trans
  causatives
  ditransatives
  voice

TAM (tense-aspect-mood)
  reflexive marking
  reciprocity marking
  instrumental marking
  benefactive marking

Other features relating to verbs
  other prefixes/suffixes
  control/volition
  simultaneity vs. sequentaility
  conjugation classes
  serial verbs
  verb compounding
  incorporation into verbs

Other features relating to larger units
  clause-chaining
  ‘and’ different from ‘with’
  tail-head-linkage

Reduplication
The feature set (cont.)

<table>
<thead>
<tr>
<th>Nominal modifiers</th>
<th>Number marking on noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>article</td>
<td>agreement on noun</td>
</tr>
<tr>
<td>attributive demonstratives</td>
<td>agreement within NP</td>
</tr>
<tr>
<td>attributive property-words (“adjectives”)</td>
<td>agreement on verb</td>
</tr>
<tr>
<td>attributive possession</td>
<td>semantics of assignment</td>
</tr>
<tr>
<td>quantifiers</td>
<td>verbal classifiers</td>
</tr>
<tr>
<td>numeral</td>
<td></td>
</tr>
<tr>
<td>diminutive and augmentive</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nominalizations</th>
<th>Pronominal system</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Argument marking</td>
</tr>
<tr>
<td></td>
<td>case</td>
</tr>
<tr>
<td></td>
<td>adpositions</td>
</tr>
<tr>
<td></td>
<td>verbal agreement</td>
</tr>
<tr>
<td></td>
<td>marking on verb</td>
</tr>
<tr>
<td></td>
<td>argument alignment</td>
</tr>
</tbody>
</table>

Gender/noun classes
Phonology

Word order

Polar questions

Non-verbal/stative predication
  attributive property predication
  nominal predication
  possessive predication
  existential/locative predication (adpositions)
  comparative predication

Negation

Valency
  trans -> intrans
  intrans-> trans
  causatives
  ditransatives
  voice

reflexive marking
reciprocity marking
instrumental marking
benefactive marking

TAM (tense-aspect-mood)

Other features relating to verbs
  other prefixes/suffixes
  control/volition
  simultaneity vs. sequentiality
  conjugation classes
  serial verbs
  verb compounding
  incorporation into verbs

Other features relating to larger units
  clause-chaining
  ‘and’ different from ‘with’
  tail-head-linkage

Reduplication
Nominal modifiers
- article
- attributive demonstratives
- attributive property-words ("adjectives")
- attributive possession
- quantifiers
- numeral
- diminutive and augmentative

Number
- number marking on nouns
- agreement on verb
- obligatoriness
- relation to gender/noun class
- associative plural

marking on noun
- agreement within NP
- agreement on verb
- semantics of assignment
- verbal classifiers

Nominalizations

Pronominal system

Argument marking
- case
- adpositions
- verbal agreement
- marking on verb
- argument alignment

Gender/noun classes
Relation to other databases and analyses

• one team working together with descriptions and experts for all features and the same set of languages,
  • as opposed to one/few researcher per feature with different sets of languages from each other (WALS)
  • as opposed to few researchers for all features and experts per language (APiCS)

• overlapping features with WALS: comparative, predicative possession, polar questions, word order, demonstrative, TAM, gender, alignment
  • however, important differences between WALS and NTS:
    • lgs can be coded for multiple strategies in NTS where WALS only allows for one (though not in percentages as in APiCS)
    • more detailed definition on certain features
    • better coverage per language

• overlapping features with Di Garbo (forth): marking and conflation of evaluation, gender and number
• overlapping features with SAILS: large selection of structural features (604)
• more micro-features to allow for alternative analyses
  • gender/noun class broken down to different agreement targets
  • polar question marking differentiated between intonation and tone
## Distinctions in formal expression

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>particle</td>
<td>Element that is invariable, i.e. does not inflect. Need not be unbound</td>
</tr>
<tr>
<td>morphologically marked on the verb</td>
<td>≠ only affixes or clitics</td>
</tr>
<tr>
<td>morphologically marked on the verb</td>
<td>= affixes, clitics, suppletion and reduplication</td>
</tr>
<tr>
<td>morphologically marked on the verb</td>
<td>≠ serial verbs, verb compounding or clause-chaining</td>
</tr>
<tr>
<td>affixes &amp; clitics</td>
<td>≠ only prefixes, suffixes, proclitics and enclitics</td>
</tr>
<tr>
<td>affixes &amp; clitics</td>
<td>= prefixes, suffixes, proclitics, enclitics, circumfixes, infixes, inclitics and circumclitics</td>
</tr>
<tr>
<td>tone</td>
<td>= affix/clitic</td>
</tr>
<tr>
<td>auxiliary</td>
<td>= phonologically independent marker of TAM</td>
</tr>
</tbody>
</table>
gender/noun class = every noun belongs to one class (occasional more). Overt marking, either on the noun itself, other elements in the NP or on the verb.

noun classifier = nouns can belong to more than one, the classifier introduces a change in the semantics of the root. Not necessarily that every noun belongs to a classifier.

some noun classes also function as markers of oblique case, most often locative

<table>
<thead>
<tr>
<th>nominal root</th>
<th>class</th>
</tr>
</thead>
<tbody>
<tr>
<td>banana</td>
<td>A</td>
</tr>
<tr>
<td>bird</td>
<td>B</td>
</tr>
<tr>
<td>house</td>
<td>C</td>
</tr>
<tr>
<td>water</td>
<td>D</td>
</tr>
<tr>
<td>human</td>
<td>E</td>
</tr>
<tr>
<td>telephone</td>
<td>F</td>
</tr>
</tbody>
</table>
Dedication & productivity

- existential or locative predicative cannot mark attributive or equative copula
- markers of simultaneity cannot be general imperfective/progressive markers
- distinction visible/nonvisible in demonstrative cannot be strictly correlated with distance
Dependencies

• 80 features that are dependent on another feature in the current set of 266 features coded for Africa

• Round (2013), Round and Bonnin (2013) and Round (2014)

F266 Can comparative constructions be construed with a locative comparative?
F267 Can comparative constructions be construed with a from-comparative?
F268 Can comparative constructions be construed with a to-comparative?
F269 Can comparative constructions be construed with a benefactive comparative?
F277 Can comparative constructions be construed with a at-comparative?

(56:1 v 199:1 v 283:1) -> 50:1

(83:0 ^ 84:0) -> 85:0
F83 Is there past tense regularly morphologically marked on the verb?
F84 Is there future tense regularly morphologically marked on the verb?

F85 Are there multiple past or future tenses, distinguishing distance from Time of Reference, marked on the verb?
Work in progress

- coding new and/or going over old coding
- double checking with experts before finalizing
- syncing with Di Garbo and SAILS
- adding to the documentation of features
  - comparing to definitions of GOLD, ISOcat etc
  - NB there are grammar or grammar sketches of 2,421 languages of the world
Applications of NTS

- published online in user-friendly interface, as all CLLD (in part bilingual French-English)
- study clusters of language in our data and compare to genealogies, known contact areas, archeological findings and genetic data
- study what features tend to be stable, direction of change and which features tend to be coupled with other features etc
- study distribution of functional load/complexity measurements
- what are the constraints on languages in the logically possible design space?
Applications of NTS

!DEMO TIME!
Dank u wel

Hedvig Skirgård &
Suzanne van der Meer

hedvig.skirgard@gmail.com
suzannevdmeer@gmail.com
References (1/2)

• Di Garbo, Francesca (forth) The interaction between gender, number and evaluative markers. Ph.D thesis at Stockholm Univeristy
The larger research question of linguistic typology

What are the cognitive, areal, cultural, historical/evolutionary constraints on diversity, disparity, and change?