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

(see Dunn et al 2005, Dunn et al 2007, Dunn et al 2008, Reesink et al 2009, Reesink & Dunn 2012)

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
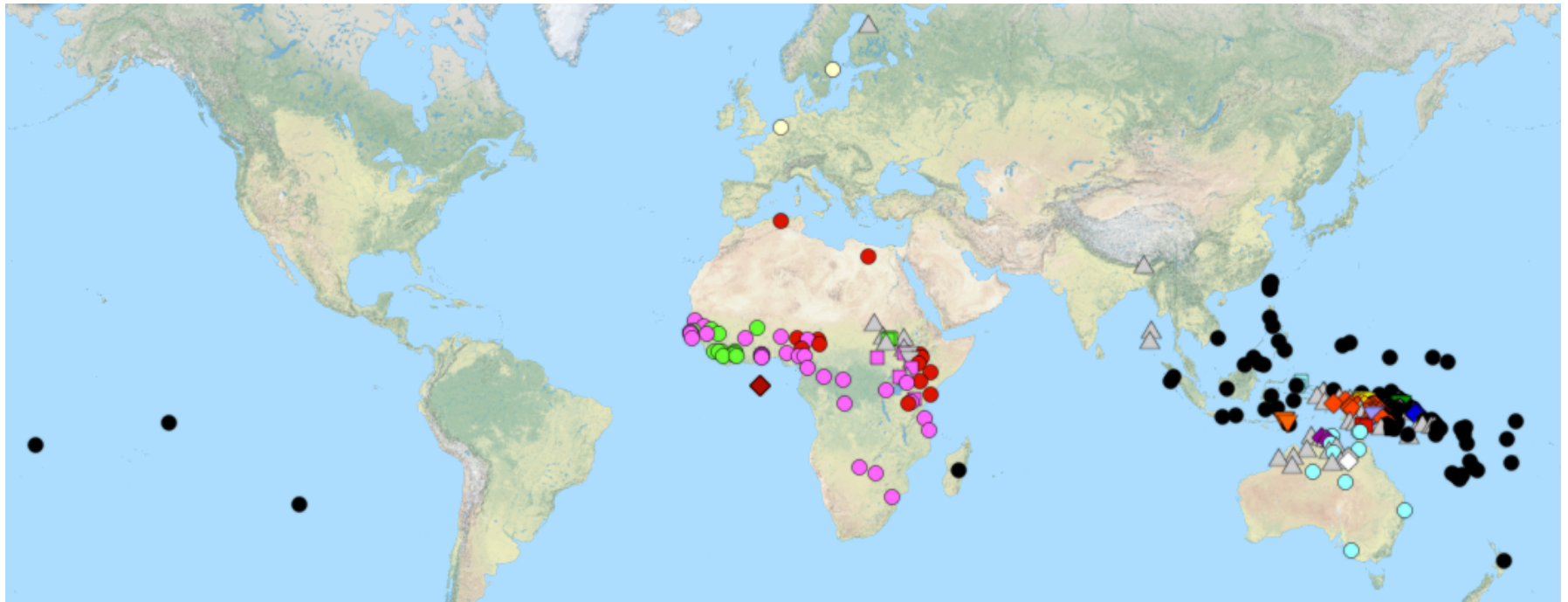
The NTS-sample: languages

- 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

Papuan	=	165
Australian	=	25
Africa	=	70
Eurasia	=	11

The NTS-sample: features

Features	=	329
from Reesink et al (2009)	=	204
new features	=	125
Feat coded for African languages	=	266
from Reesink et al (2009)	=	141
new features	=	125
complementing previous	=	64
inspired by WALS	=	32
inspired by Di Garbo (2014)	=	19
Africa-specific	=	2
other	=	8

The African set
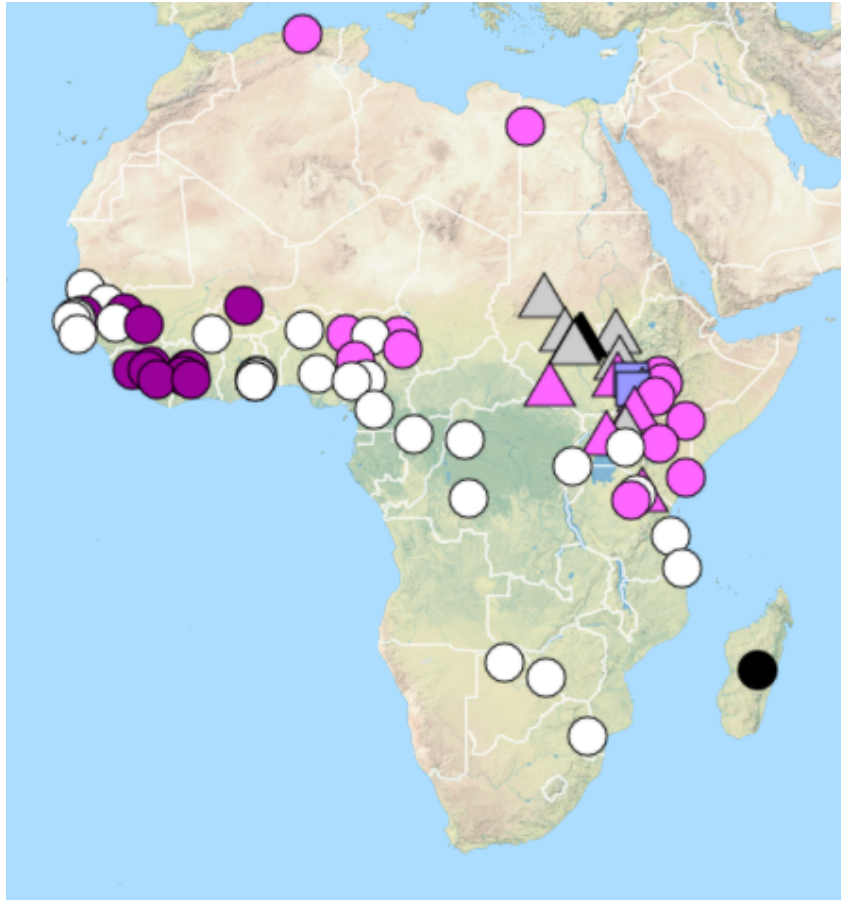
Afro-asiatic
Atlantic-Congo
Dizoid
Heiban
Ijoid
Khoe-kwadi
Koman
Kuliak
Mande
Nilotic
Nubian
Songhay

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.

Values	
blank	No information
?	Not enough information to code
n/a	Depending on other feature value which is 0
0/1/2/3/4/5	Enough information

- 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 Univeristy), 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

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

Discrete categories
of continua

CONSISTENCY

numeral one				indefinite/non-specific article
demonstrative				copula
lexical verb		auxiliary		affix
???		auxiliary		affix
body part		reflexive pronoun		middle voice
definiteness marker		noun class/gender	noun classifier	numeral classifier
phrasal verbs	"verb particles"	light verb constructions	co-verb constructions	copula

The feature set of NTS

Phonology

Word order

Polar questions

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

The feature set (cont.)

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

NTS <-

Phonology

Word order

Polar questions

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

NTS <-

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 mulitple 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

| particle	=	element that is invariable, i.e. does not inflect. Need not be unbound
morphologically marked on the verb	≠	only affixes or clitics
morphologically marked on the verb	=	affixes, clitics, suppletion and reduplication
morphologically marked on the verb	≠	serial verbs, verb compounding or clause-chaining
affixes & clitics	≠	only prefixes, suffixes, proclitics and enclitics
affixes & clitics	=	prefixes, suffixes, proclitics, enclitics, circumfixes, infixes, inclitics and circumclitics
tone	=	affix/clitic
auxiliary	=	phonologically independent marker of TAM

Systems of nominal classification

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

nominal root	class
banana	A
bird	B
house	C
water	D
human	E
telephone	F
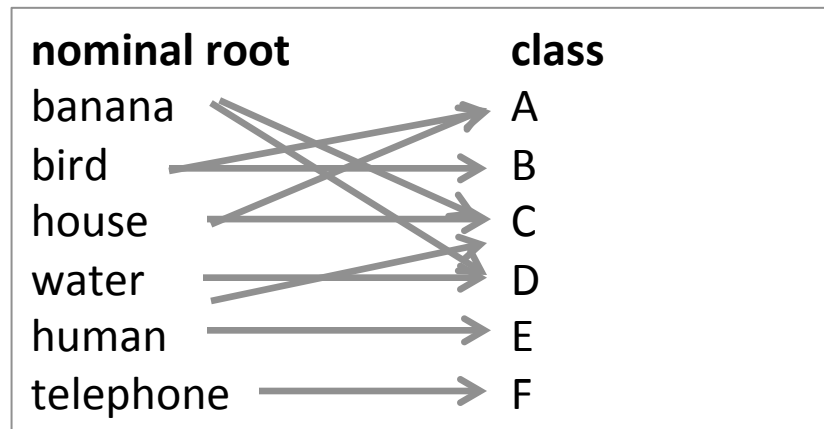
Dedication & productivity

- existential or locative predicator 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?

$(56{:}1 \lor 199{:}1 \lor 283{:}1) \rightarrow 50{:}1$

$(83{:}0 \land 84{:}0) \rightarrow 85{:}0$
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 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)

- Collins, Chris and Richard Kayne (2009) Database of the Syntactic Structures of the World's Languages. ((Available online at: http://sswl.railsplayground.net/)
- Cysouw, Michael & Jeff Good. 2013. Languoid, Doculect, Glossonym: Formalizing the notion "language". *Language Documentation and Conservation* 7. 331-359. http://hdl.handle.net/10125/4606
- Di Garbo, Francesca (forth) The interaction between gender, number and evaluative markers. Ph.D thesis at Stockholm Univeristy
- Dryer, Matthew S (2013) Polar Questions. In:Dryer, Matthew S. & Haspelmath, Martin (eds.) The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dryer, Matthew S. & Haspelmath, Martin (eds.) (2013) The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dunn, Michael, Angela Terrill, Ger Reesink, Robert A. Foley & Stephen C. Levinson. 2005. Structural phylogenetics and the reconstruction of ancient language history. *Science* 309. 2072–2075.
- Dunn, Michael, Robert A. Foley, Stephen C. Levinson, Ger Reesink & Angela Terrill. 2007. Statistical reasoning in the evaluation of typological diversity in Island Melanesia. *Oceanic Linguistics* 46(2). 388-403.
- Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesink, & Angela Terrill. 2008. Structural phylogeny in historical linguistics: Methodological explorations applied in Island Melanesia. *Language* 84(4). 710-759
- Floyd, S. (2013). Do field linguists really ask the "wrong questions"? A reply to Haspelmath (2012a). Blog post on Diversity Linguistics Comment.
- Hartmann, Iren & Haspelmath, Martin & Taylor, Bradley (eds.) 2013. Valency Patterns Leipzig. Leipzig: Max Planck Institute for Evolutionary Anthropology.

References (2/2)

- Haspelmath, M. (2010a). Comparative concepts and descriptive categories and in cross-linguistic studies. Language, 86:663–687. Max Planck Institute for Evolutionary Anthropology.
- Haspelmath, M. (2010b). Framework-free grammatical theory. In Heine, B. and Narrog, H., editors, The Oxford Hhandbook of Linguistics Analysis. Oxford University Press, Oxford.
- Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.) 2013. Atlas of Pidgin and Creole Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Moran, Steven. 2012. Phonetics Information Base and Lexicon. PhD thesis. University of Washington.
- Reesink, G., Singer, R., & Dunn, M. (2009). Explaining the linguistic diversity of Sahul using population models. **PLoS Biology, 7(11), e1000241. doi:10.1371/journal.pbio.1000241**
- Round, E. (2013). 'Big data' typology and linguistic phylogenetics: Design principles for valid datasets. In 21st Manchester Phonology Meeting.
- Round, Erich (2014) What to report? — Bayesian clustering & 'researcher degrees of freedom'. TaSiL, Aarhus, Denmark
- Round, E. and Bonnin, C. (2013). How to design a dataset which doesn't undermine automated analysis. In Association for Linguistic Typology's 10th Biennial Conference (ALT 10).
- **Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.). 2014. *Ethnologue: Languages of the World, Seventeenth edition.* Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.**
- Reesink, Ger & Michael Dunn (2012) Systematic typological comparison as a tool for investigating language history. in Nicholas Evans and Marian Klamer (eds) Language Documentation & Conservation Special Publication No. 5 Melanesian Languages on the Edge of Asia: Challenges for the 21st Century. pp. 34–71
- Stassen, Leon (2013) Predicative Possession. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology.

The larger research question of linguistic typology

What is the role of structure, areal, cultural, communicative, and historical/evolutionary constraints in shaping diversity, disparity, and change?

Ethnologue