

# Explaining the lack of formal differentiation between WHO? and WHAT?<sup>1</sup>

**Dmitry Idiatov**

**Center for Grammar, Cognition and Typology**

**University of Antwerp**

[dmitry.idiatov@ua.ac.be](mailto:dmitry.idiatov@ua.ac.be)

Association for Linguistic Typology, 6th Biennial Meeting, 21-25 July 2005,  
Padang, West Sumatra, Indonesia

## 1. State of affairs

- It is generally assumed that the distinction between WHO? and WHAT? is basic to natural languages.

All languages have words for WHO and WHAT, and can distinguish lexically between the questions “What is this?” and “Who is this?”. The distinction between ‘who’ and ‘what’, ‘someone’ and ‘something’, ‘person’ and ‘thing’ provides the most fundamental form of human categorization... Natural languages differ in this respect sharply from artificial languages relying on the abstract notion of “referential indices”.

(Wierzbicka 1996:38-39)

- However, examples of the languages where one form means both WHO? and WHAT? have been reported here and there in the typological literature. E.g.,

- Latvian, Lithuanian [Indo-European]; Khasi [Austro-Asiatic; India]

(1) Latvian [Indo-European] (Nau 1999:146)

*kas tas ir?*  
who/what this is  
‘Who/what is this?’

- Lindström (1995) adds Mataco [= Wichí Lhamtés Nocten] (Bolivia/Argentina)

- Cysouw (2004) gives 18 languages

- (2)
- |                    |  |
|--------------------|--|
| – INDO-EUROPEAN:   | Latvian, Lithuanian  |
| – AUSTRO-ASIATIC:  | Khasi (India)  |
| – KHOISAN:         | !Xóǀ (Namibia/Botswana)  |
| – TUPI:            | Mekens Sakirabiat (Brazil)   |
| – ARAUAN:          | Paumarí (Brazil)   |
| – CHIBCHAN:        | Ika (Colombia)   |
| – ARAWAKAN:        | Achagua (Colombia), Apurinã (Brazil), Asheninca<br>Campa [= Ashéninka Pichis?] (Peru), Baré (Venezuela),<br>Nomatsiguenga (Peru), Terêna (Brazil), Warekena<br>(Venezuela) |
| – MATACO-GUAICURU: | Macá (Paraguay), Mataco [= Wichí Lhamtés Vejoz]<br>(Argentina), Mocoví (Argentina), Toba (Argentina)   |

- In a world-wide convenience sample of some 500 languages, I found around 60 languages that do not oblige their speakers to distinguish WHO? from WHAT?. (This number does not

---

<sup>1</sup> This paper results from research conducted as a part of the GOA (Geconcerteerde Onderzoeksactie) “Mood and Modality” project of the University of Antwerp. I wish to thank the Research Council of the University of Antwerp for its financial support.

include the well-known cases of partial formal overlap of the paradigms of WHO? and WHAT? of the type found in Latin or Tsez).

## 2. Aims (as far as time permits...)

- Provide a typology of WHO?/WHAT? identity/overlap:
  - kinds of identity/overlap
  - areal/genetic patterns
  - correlation with any grammatical/lexical features
  - diachrony
  - explanation + analysis of a few interesting cases
- Find answers to the following questions:
  - Why do (most) languages oblige their speakers to distinguish WHO? from WHAT?
  - Why do (some) languages not oblige their speakers to distinguish WHO? from WHAT?

## 3. Sample

A world-wide convenience sample of some 500 languages (see Appendix 1), which includes:

- 151 WALS-languages: 81 languages from the WALS-100 list and 68 from the additional WALS-200 list
- 317 languages from a weighted 500-language sample, which was made up on the basis of Rijkhoff, Bakker, Hengeveld & Kahrel (1993)'s maximal genetic diversification algorithm<sup>2</sup>

## 4. Some general notes on interrogative pro-words

- The formal side:
  - part-of-speech categories the interrogative pro-words stand for matter most, because these words are *pro*-words
  - therefore, one should first of all speak about INTERROGATIVE PRO-NOUNS, INTERROGATIVE PRO-ADVERBS, INTERROGATIVE PRO-ADJECTIVES, etc.<sup>3</sup>
- The semantic side:
  - prototypical referents of NOUNS (part-of-speech category) are ENTITIES (ontological category). Thus, a language should in principle be perfectly happy with only one interrogative pro-noun. After all, in a situation where, for instance, a person hears some noise behind a dense bush and wants to know what real-world ENTITY produces this noise, this person is actually supposed to be ignorant about whether this noise-making ENTITY is a person, thing, animal, etc.
  - subdivision of NOUNS as to how they and the ENTITIES they refer to are related:
    - (3) – PROPER NAMES
    - SHIFTERS (indexical words)
    - COMMON NOUNS

<sup>2</sup> I would like to thank Dik Bakker for kindly providing such a 500-language sample.

<sup>3</sup> See Idiatov & van der Auwera (2004) for a discussion of (im)possible part-of-speech categories of interrogative pro-words.

- The “most fundamental form of human categorization” of ENTITIES is, as Wierzbicka correctly points out, between PERSON and THING, or maybe better between PERSON and NON-PERSON, with THING being the prototypical case of the latter category. Other cut-off points, e.g., based on ANIMACY, are secondary. Non-human animates belong to a transitory zone, which can be categorized in various ways.
  - PERSONS are prototypical holders (referents) of PROPER NAMES. Personal PROPER NAMES seem to be the best representatives of PERSON-denoting nouns. It is not a coincidence that PERSON-denoting nouns generally tend to be more specific than THING-denoting nouns.
- (4) [The speaker is with person B. They see X (PERSON/THING). B seems to know X. The speaker asks B about the identity of X:]
- a. – *Who is this?*  
– *It’s John/ my brother/ a friend of mine/ the doctor/ my doctor* vs. *It’s a doctor.*
  - b. – *What is this?*  
– *It’s an apple* vs. *It’s the apple.*
- Questions asking for a choice from the known range of alternatives and suggesting a free choice should be distinguished.
  - Indexical lemmas, such as *this (one)* or *that (one)*, seem to represent typical answers in situations when the choice is restricted to a closed set of referents.
  - This gives us the following basic categories of interrogative pro-nouns:

(5) INTERROGATIVE PRO-NOUNS	CATEGORY	ANSWER
Free choice:		
– WHO/WHAT?	[ENTITY]	...
– WHO?	[PERSON]	<i>John, my brother, a friend of mine</i>
– WHAT?	[THING]	<i>a table, an apple</i>
Restricted choice (= SELECTION, IDENTITY, etc.):		
– WHICH.ONE?	[ENTITY]	<i>this one</i>
– WHICH.PERSON?	[PERSON]	<i>this man</i>
– WHICH.THING?	[THING]	<i>this apple</i>

## 5. A typology of WHO?/WHAT? identity/overlap

### 5.1 Types

**TYPE A: A special WHO/WHAT?-word** (see Map 2)

**TYPE A.1:** One or (rarely) more words WHO/WHAT? + no other word meaning WHO? or WHAT?

**TYPE A.1.1:** One word WHO/WHAT?

- |                   |  |
|-------------------|--|
| (6) AFRO-ASIATIC: | Tamazight Berber (Ayt Ndhir dialect; Morocco),<br>Tahaggart Tamahaq Berber (Algeria)   |
| ARAUAN:           | Paumarí (Brazil)   |
| ARAWAKAN:         | Achagua (Colombia), Baure (Bolivia), Cabiyaí<br>(Colombia), Kaixana (Brazil), Machinere (at least one<br>dialect; Brazil), Maipure (Venezuela), Mandahuaca<br>(Venezuela), Piapoco (Colombia), Terêna (Brazil),<br>Yucuna (Colombia) |

AUSTRO-ASIATIC:	Sapuan (Laos)
CHAPACURA-WANHAM:	Itene/Moré (Bolivia/Brazil)
CHIBCHAN:	Arhuaco/Ika (Colombia)
GARAWAN:	Eastern Garawa (Australia)
GUNWINGGUAN:	Ngandi (Australia)
INDO-EUROPEAN, BALTIC:	Latvian, Lithuanian
ISOLATE:	Urarina (Peru)
MATACO-GUAICURU:	Kadiwéu (Brazil), Macá (Paraguay), Wichí Lhamtés Nocten (=Mataco; Bolivia/Argentina), Wichí Lhamtés Vejoz (=Mataco; Argentina), Toba (Argentina)
MAYAN:	Kaqchikel (most dialects; Guatemala), Northern Mam (San Pedro Necta dialect; Guatemala), Tzutujil (Guatemala)
NYULNYULAN:	Djawi (Australia), Dyaberdyaber (Australia), Dyugun (Australia), Nimanbur (Australia), Nyigina (Australia), Nyulnyul (Australia), Warrwa (Australia), Yawuru (Australia)
PAMA-NYUNGAN:	Walmajarri (Australia), Warumungu (Australia)
TANGIC (?PAMA-NYUNGAN):	Kayardild (Australia), Yukulta/Ganggalida (Australia)
TUPI:	Sakirabiat Mekens (Brazil)

## (7) Tzutujil [Mayan; Guatemala]

- a. *naq npit chwaaq*  
 WHO/AT 3SG.A-come tomorrow  
 ‘Who/what is coming tomorrow?’ (Dayley 1985:332)
- b. *naq neechoyb’eej ja q’aayiis*  
 WHO/AT 3SG.O-2PL.A-cut-with the weed  
 ‘What do you all cut the weed with?’ (Dayley 1985:335)

**TYPE A.1.2:** Two words WHO/AT?

## (8) GARAWAN: Western Garawa (Australia)

**TYPE A.2:** One (or more?) word WHO/WHAT? + other words meaning either WHO? or WHAT?**TYPE A.2.1:** WHO/WHAT? + WHAT?

- (9) AFRO-ASIATIC: Egyptian/ Coptic  
 WEST BARKLY: Djingili (Australia)  
 YANOMAM: Sanumá (Venezuela)?

## (10) Djingili [West Barkly; Australia]

- a. *Aji-rni-mbili ya-miki jama-rni-ma*  
 WHO/AT-?-LOC 3SG-came that:M-FOC-EMPH  
 ‘Who/what did he bring?’ (Pensalfini 2003:144)
- b. *Nyamba jiminiki-ri? Jiminiki-rni darrangku ngaba-nya-ju*  
 WHAT this:N-FOC this:N-FOC tree have-2SG-do  
 ‘What’s that? That’s a tree you have there’ (Pensalfini 2003:142)

**TYPE A.2.2:** WHO/WHAT? + WHO?

- (11) ARAWAKAN: Baré (Venezuela/Brazil), Warekena (Venezuela/Brazil)  
 NYULNYULAN: Bardi (Australia)?

(12) Baré [Arawakan, Venezuela/Brazil] (Aikhenvald 1995:25)

*ne* 'who?, what?'  
*abadi* 'who?, which one?'

**TYPE A.2.3:** WHO/WHAT? + WHO? + WHAT?

(13) Tamazight Berber (Aït Seghrouchen of Oum Jeniba dialect) [Afro-Asiatic; Morocco] (Bentolila 1981:110)

*may* 'who?, what?'  
*wi* 'who?'  
*maTa* 'what?'

**Unattributed:** Arawak [Arawakan; Suriname/ French Guiana/ Guyana/ Venezuela] (the sources are somewhat contradictory), Abipon [Mataco-Guaicuru; Argentina] and Kwini [Wororan; Australia] (maybe WHO/AT? + WHAT?, but the descriptions are not clear), Sabanês [Nambiquaran; Brazil] (the description is not clear), Mocoví [Mataco-Guaicuru; Argentina], !Xóǀ [Khoisan; Namibia/Botswana].

**Special cases:** (a) Ashéninka Campa [Arawakan; Peru], Nomatsiguenga [Arawakan; Peru], Apurinã [Arawakan; Brazil], Khasi [Austro-Asiatic; India], Wambaya [West Barkly; Australia] (b) Mongo-Nkundo [Niger-Congo; Democratic Republic of Congo] (see Section 8)

**TYPE B: No special WHO/WHAT?-word**

**TYPE B.1:** Functional overlap (at least around 30 languages; see Map 3)

**TYPE B.1.1:** NAME: 'WHO is X's name?' (X is PERSON, THING, PLACE...)

(14) PERSON

Paama [Austronesian; Vanuatu]

a. *Ise-n* *isei*  
 name-3SG WHO  
 'What is his/her name? (lit.: Who is his name?)' (Crowley 1982:242)

b. *Ise-n* *asaa*  
 name-3SG WHAT  
 'What is its name? (lit.: What is its name?)' (Crowley 1982:242)

(15) THING

Rapanui [Austronesian; Chile]

*Ko-ai te ava'e ko tara hao hai vanga tire*  
 FOC-WHO SPC month FOC *tara hao* INST language Chile  
 'What is *tara hao* [= January] called in Spanish?' (Du Feu 1996:22)

(16) PLACE

Pitjatjantjara [Pama-Nyungan; Australia]

*Ngura nyangatja ini ngana-nya*  
 place name this WHO-ABS  
 'What's the name of the place here?' (Eckert & Hudson 1994:119)

**TYPE B.1.2.1:** KIND: 'WHAT is (PERSON) X? (a man or a woman)' → 'What kind of person is person X? (a man or a woman)'

(17) Eton [Niger-Congo, Bantu; Cameroon]

*Tômɔ̀ ànè jé? dè mìnngá, dè pâm?*

Tomo is WHAT YNQ woman YNQ man

‘What is Tomo? A man or a woman?’ (Mark Van de Velde, p.c.)

(18) Russian [Indo-European]

*Tomo kto? Muzhchina ili zhenschina?*

Tomo WHO man or woman

‘What is Tomo? A man or a woman?’

**TYPE B.1.2.2:** OCCUPATION: ‘WHAT is (PERSON) X? (a doctor or a bus-driver)’ → ‘What is person X’s occupation? (a doctor or a bus-driver)’

(19) *What is he? A doctor or a bus-driver?*

(20) Russian [Indo-European]

*On kto? Doktor ili voditel’ avtobusa?*

he WHO doctor or driver of.bus

‘What is he? A doctor or a bus-driver?’

**TYPE B.1.2.3:** (primarily KINSHIP) RELATION: ‘WHAT is (PERSON) X to you? (your husband, your brother)’ → ‘What is person X’s RELATION to you? (husband-RELATION, brother-RELATION)’

(21) Bambara [Niger-Congo, Mande; Mali]

– *í mùn dòn? – ní dógóké’ dén’ dòn*

your WHAT is my younger.brother-ART child-ART is

‘– What is he to you? (lit.: ‘[He] is your what?) – It’s my younger brother’s child’ (Dumestre 2003:151)

(22) Russian [Indo-European]

– *On tebe kto? – Brat.*

he to.you WHO brother

‘– What is he to you? – It’s my brother.’

**TYPE B.1.3:** SELECTION: ‘WHICH THING?’ (lit.: WHO THING?)

(23) Assamese [Indo-European; India]

a. *kon mānuh*

WHO man

‘Which man?’ (Babakaev 1980:76)

b. *kon kitāp*

WHO book

‘Which book?’ (Babakaev 1980:76)

**TYPE B.2:** Formal overlap of the paradigms of WHO? and WHAT?

**TYPE B.2.1:** Within the case paradigms of WHO? and WHAT?

E.g., in Latin [Indo-European] or Tsez/Dido [North Caucasian; Russia]

**TYPE B.2.2:** Within the gender/number paradigms of WHO? and WHAT?

E.g., in Kuvi [Dravidian; India] (Reddy 1979), to a certain extent Swedish [Indo-European] and apparently most of the aforementioned Arawakan languages (see Type A) at a certain stage in their development

In Swedish, according to Holmes & Hinchliffe 2003:185-187, the plural form of ‘which (one)?’ *vilka* (from singular *vilken*) “also serves as the plural of *vem* [‘who?’] and under certain conditions as a plural for *vad* [‘what?’]”.

Most Arawakan languages have only two genders, masculine and feminine, which seem to have originally been restricted to animates in most languages. Inanimates, i.e. THINGS, are treated differently with the help of classifiers in some languages. However, most languages have generalized one of the animate genders to include inanimates. Most often it has been feminine, sometimes masculine, i.e. nowadays the opposition is often between feminine (animate) vs. non-feminine or masculine (animate) vs. non-masculine. Thus, in Nomatsiguenga and Asheninka [Arawakan; Peru] the neutralization has happened according to the masculine (animate) vs. non-masculine pattern (e.g., Shaver 1996:29). The word for WHO? typically distinguishes masculine and feminine gender. As a result, in these languages the meaning WHO? has become coded by the WHO/AT:M? form and the meaning WHAT? by the WHO/AT:F? form, as in (24).

(24) Nomatsiguenga [Arawakan; Peru]

- a. *paírí*            *játatsi*  
 WHO/AT:M    he:is:coming  
 ‘Who is coming?’ (Shaver 1996:37)
- b. *paíró*            *píjíta*  
 WHO/AT:F    you:are:called  
 ‘What’s your name? (lit.: ‘What are you called?’)’ (Shaver 1996:37)

In Asheninka Campa, the gender is not (no more?) marked on the interrogative pro-noun, but on the verb.

(25) Asheninka Campa [Arawakan; Peru]<sup>4</sup>

- a. *tsi-ka*        *i-pait-a-ka*            *pok-atsi-ri*            *iroñaaka*  
 QW-INT    3M-call-TAM-INT    come-TAM-REL    now  
 ‘Who is coming? (lit.: ‘What is he, who is coming now, called?’)’ (Cysouw 2004 citing Anderson 1985/1986)
- b. *tsi-ka*        *o-pait-a-ka*            *h-a-ake-ri*  
 QW-INT    3F-call-TAM-INT    3M-take-TAM-REL  
 ‘What did he take? (lit.: ‘What is she, who he took, called’ (Cysouw 2004 citing Givón 2001:304-5 who cites D. Payne, p.c.)
- c. *tsi-ka*        *p-a-ake-ro-ka*  
 QW-INT    2-get-TAM-3F-INT  
 ‘Where did you get **it**?’ (Cysouw 2004 citing Givón 2001:304-5 who cites D. Payne, p.c.)

## 5.2. Remarks

- OCCUPATION > KIND
- NAME:PLACE > NAME:PERSON and NAME:THING > NAME:PERSON
- There appears to be a certain tendency for the implication NAME:PLACE > NAME:THING
- Functional overlaps (Type B.1) can be explained in terms of:

<sup>4</sup> Cysouw (forthcoming) discusses in a very detailed and insightful way the interesting system of interrogative pro-word(s) of Asheninka Campa. However, he explicitly states that he has not “found any reason for the choice between (...) the third person masculine and feminine prefixes”.

- (a) “multifunctional agreement patterns”, which implies that one agreement pattern may be used to mark different agreement features, such as GENDER, ANIMACY, HUMANNES, etc. (for more details see Van de Velde, forthcoming). E.g., Rapanui (NAME:THING type) and Pitjantjatjara (NAME:PLACE type), English and Eton (KIND & OCCUPATION types)
- (b) choices made between one of the other controller for the purposes of agreement. E.g., Paama and Eton (NAME:PERSON type)

## 6. Areal/genetic patterns

- There appears to be a certain areal bias on a world-wide scale among the languages that were found to have a WHO/WHAT? interrogative pro-noun: almost half of them are spoken in South America and almost a third in Australia (see Map 2).
  - In South America, this areal pattern can be further reduced to a region which can be broadly defined as the Amazonia-Andes border.
  - In Australia, primarily the northern part of the continent is involved.
- Some families are particularly prone to “misbehave”, often influencing the neighbours:
  - In South America, the Mataco-Guaicuru family and especially the Arawakan family seem to be responsible for the attested areal patterning
  - In Australia, the languages at issue are typically Non-Pama-Nyungan. Nyulnyulan languages are one of the “leaders” here.
  - Baltic languages
  - Berber and Egyptian branches of the Afro-Asiatic family
  - Quichean-Mamean branches of the Mayan family
- Some kinds of functional overlap (Type B.1) also show a certain areal (and sometimes genetic) bias on a world-wide scale:
  - Functional overlaps of the NAME, KIND, OCCUPATION and RELATION types are virtually non-reported for South America.
  - Functional overlaps of the NAME type seem to be extremely exceptional in Europe and most parts of Eurasia in general.
  - In Europe, the OCCUPATION and RELATION type seems to be rare outside of Germanic and certain Romance languages.
  - The languages with NAME:THING and especially NAME:PLACE kinds of functional overlap seem to be primarily concentrated in Australia and to a lesser extent among the Austronesian and Papuan languages. At the same time, KIND, OCCUPATION and RELATION types are rarely reported for these regions.

## 7. Correlations with other features

- A significantly high number of split ergative alignment (most often of an active-stative type; for some languages only in diachrony) among the languages with a WHO/WHAT? interrogative pro-noun.
- A large number of languages with gender, and typically with only two genders, one of which carries a lower functional load.
- A very strong tendency for a WHO/WHAT? interrogative pro-noun to have no additional indefinite meaning



## 8. Mongo-Nkundo [Niger-Congo, Bantu, C61; DRC]

- In Mongo the interrogative pro-noun *ná* ‘who?’, as in (27), has been extended to the meaning ‘what?’, as in (28). Subsequently, the interrogative pro-noun *é* ‘what?’ has been extended to the meaning ‘who?’, as in (29), but its use in the latter meaning seems to be rather limited.

(27) *ǒ-kel-aki* *ná?*  
 G1:REL-do-HODIERNAL.PAST WHO/AT  
 ‘Who has done it? (lit.: ‘Who is it who has done [it]?’)’ (Hulstaert 1965:144)

(28) *ó-kel-aki* *ná?*  
 2SG-do-HODIERNAL.PAST WHO/AT  
 ‘What have you done?’ (Hulstaert 1965:144)

(29) *i-lɔmbɛ* *y-ǎ* *é*  
 G7-house G7-CONN WHAT/O  
 ‘whose house?’ (Hulstaert 1938:78)

- It is really surprising to find Mongo among the languages with a WHO/AT? interrogative pro-noun because:

- it seems to be exceptional in Niger-Congo and in Sub-Saharan Africa in general (only a Khoisan language !Xóǒ has been reported as another exception yet)
- it has a typical Bantu multigender system
- it is nominative-accusative

- How did this happen?

- An important factor responsible for this change must have been the preceding extension of the original word meaning ‘what?’ *é* to the function of a polar question marker, as in (28).

(30) *a-ól-úndol-a* *lɔkendo* *é*  
 G1-HODIERNAL.RES-come.back.from-FINAL.VOWEL journey YNQ[=WHAT]  
 ‘Is he back from the journey?’ (Hulstaert 1938:79)

- The sentence-final position of interrogative pro-forms in content questions must have played a role.
- Some peculiarities of the concord system of Mongo. The word for WHO? triggers the agreement pattern of gender 1 in Bantu languages. In Mongo, common nouns in generic use also trigger the same agreement pattern on the verb as gender 1 (for a discussion of the Mongo agreement system, see Van de Velde, forthcoming), as in (29b). Recall the link between WHAT? and non-specificity highlighted in example (4). The fact that the interrogative pro-noun *ná* did not have to change its agreement pattern when expanding to the meaning WHAT?, as can be seen in (30-31), must have undoubtedly facilitated the expansion.

(31) a. *mpulú é-tóngá jumbu*  
 [G9]bird G9-build nest  
 ‘The bird is building a nest.’ (Hulstaert 1966:17)

b. *mpulú á-tóngá jumbu*  
 [G9]bird G1-build nest  
 ‘Birds build nests (general truth).’ (Hulstaert 1966:17)

(32) [Ínyó betúbyaíso!] *ǒ-lekí* *ná?*  
 G1:REL-be.superior:PRES WHO/WHAT

[*mbeka nkína lotálio šyéngwéyá mbeka?*]

‘[Ye fools and blind:] for whether is greater, [the gift, or the altar that sanctifieth the gift]?’ (Matthew 23:19)

- (33) *ǒ-nko é?*  
 G1:REL-that WHAT/WHO  
 ‘What’s up?/ What is that?’ (Hulstaert 1938:78)

### Abbreviations:

A – agent, ABS – absolutive, CONN – connective, EMPH – emphasis, F – feminine, FOC – focus, G – gender, INST – instrumental, INT – interrogative, LOC – locative, M – masculine, N – neuter, O – object, PL – plural, PRES – present, QW – interrogative pro-word, REL – relative, SG – singular, SPC – specific, TAM – tense-aspect-modality, YNQ – polar question marker.

### References:

- Aikhenvald, Alexandra Y. (1995). *Bare*. München: LINCOM Europa. (LWM, 100)
- Anderson, Ronald Jaime (1985/1986). *Cuentos Folkloricos de los Asheninca*. Pucallpa: Instituto Lingüístico de Verano. (Comunidades y Culturas Peruanas, 18-20)
- Babakaev, Vasily D. (1980). *Ocherki morfologicheskoy struktury assamskogo jazyka* [Studies on the morphological structure of the Assamese language]. Moscow: Nauka.
- Bentolila, Fernand (1981). *Grammaire fonctionnelle d’un parler berbère: Aït Seghrouchen d’Oum Jeniba (Maroc)*. Paris: SELAF. (Langues et civilisations à tradition orale, 46)
- Crowley, Terry (1982). *The Paamese language of Vanuatu*. Canberra: ANU. (Pacific Linguistics, B-87)
- Cysouw, Michael (2004). *Minimal systems of interrogative words*. Paper read at the Workshop on Questions at MPI, Leipzig, October 15, 2004.
- Cysouw, Michael (forthcoming). *Interrogatives in Asheninka Campa (Arawakan)*.
- Dayley, Jon P. (1985). *Tzutujil grammar*. Berkeley: University of California. (UCPL, 107)
- Du Feu, Veronica (1996). *Rapanui*. London, New York: Routledge.
- Dumestre, Gérard (2003). *Grammaire fondamentale du bambara*. Paris: Karthala.
- Eckert, Paul & Joyce Hudson (1994). *Wangka Wiru: a handbook for the Pitjantjatjara language learner*. Underdale: University of South Australia.
- Givón, Talmy (2001). *Syntax: An Introduction*. Vol. 2. Amsterdam: Benjamins.
- Holmes, Philip & Ian Hinchliffe (2003). *Swedish: A comprehensive grammar* (2nd edition). London, New York: Routledge. (Routledge comprehensive grammars)
- Hulstaert, Gustaaf (1938). *Praktische grammatica van het Lonkundo (Lomongo) (Belgisch Kongo)*. Antwerp: De Sikkel.
- Hulstaert, Gustaaf (1965). *Grammaire du Lómóngó*, Vol. 2: *Morphologie*. Tervuren: Musée Royale de l’Afrique Centrale. (Annales sciences humaines, 57)
- Hulstaert, Gustaaf (1966). *Grammaire du Lómóngó*, Vol. 3: *Syntaxe*. Tervuren: Musée Royale de l’Afrique Centrale. (Annales sciences humaines, 58)
- Idiatov, Dmitry & Johan van der Auwera (2004). On interrogative pro-verbs. In Ileana Comorovski & Manfred Krifka (eds.), *Proceedings of the Workshop on the Syntax, Semantics and Pragmatics of Questions, ESSLLI 16, August 9-13, 2004, Nancy, France*, 17-23. Nancy.
- Nau, Nicole (1999). Was schlägt der Kasus? Zu Paradigmen und Formengebrauch von Interrogativpronomina. *Sprachtypologie und Universalienforschung* 52-2:130-150.
- Pensalfini, Rob (2003). *A grammar of Jingulu, an Aboriginal language of the Northern Territory*. Canberra: Australian National University.
- Reddy, J. (1979). *Kuvi grammar*. Mysore: Central Institute of Indian Languages.
- Rijkhoff, Jan, Dik Bakker, Kees Hengeveld & Peter Kahrel (1993). A method of language sampling. *Studies in Language* 17-1:169-203.
- Shaver, Harold (1996). *Diccionario nomatsiguenga-castellano, castellano-nomatsiguenga*. Yarinacocha: Instituto Lingüístico de Verano.
- Van de Velde, Mark (forthcoming). *Multifunctional agreement patterns in Bantu and the possibility of genderless nouns*.
- Wierzbicka, Anna (1996). *Semantics: Primes and Universals*. Oxford: Oxford University Press.

## Appendix 1: The languages of the sample

Aari, Abidji, Abipon, Abkhaz, Abun, Aceh, Achagua, Acholi, Achuar, Acoma/Keres (Western), Ainu, Akan, Alabama, Alablak, Albanian, Aleut (Western), Amarakaeri/Harakmbet, Ambei (West)/Opa, Ambonese (Malay), Amele, Amharic, Amo/Ba, Amuesha/Yanesha', Andoque/Andoke, Aneityum, Anuak/Anywa, Apinayé, Apurinã, Arabic (Egyptian), Arabic (Gulf), Araona, Arawak/Lokono, Arhuaco/Ika, Armenian (Eastern), Ashéninka (Campa=?Pichis), Asmat (Central), Assamese, Athpariya/Athpare, Au, Awa Pit/Awa-Cuaiquer, Aymara, Bambara, Banda-Linda, Baniva/Baniwa of Guanía, Baniwa, Barasana, Bardi, Baré, Barí, Bashkir, Basque, Batak (Karo), Baure, Bawm, Berber (Kabyle), Berber (Mauritanian)/Zenaga, Berber (Tamahaq Tahaggart), Berber (Tamazight, Aït Seghrouchen of Oum Jeniba), Berber (Tamazight, Ayt Nahir), Berber (Tamazight, Figuig), Berbice Dutch Creole, Bidyogo/Bijogo, Biri, Bislama, Bissa, Blackfoot, Bobo Fing/Bobo Madaré, Borôro/Bororo, Brahui, Breton, Brokskat, Bukiyip, Bulgarian, Burushaski, Cabiyari, Cahuilla, Camling, Carapana, Carib/Galibi, Carijona, Cashinahua, Cayapa/Chachi, Cayuvava/Cayubaba, Cebuano, Chácobo, Chemehuevi/Ute/Paiute (Southern), Chinantec (Comaltepec), Chinantec (Lealao), Chinantec (Tepetotutla), Chinese (Mandarin), Chinese (Yue)/Cantonese, Chrau, Chukchi/Chukot, Cogui/Kogui, Colorado, Coptic, Cubeo, Cuiba (Maibén), Culina/Kulina, Dagaare, Dan, Dehu/Drehu, Dení, Desano, Dhuwal/Djapu, Diegueño/ Kumiai/Jamul Tiipay, Djaabugay/Dyaabugay, Djawi, Djingili/Jingulu, Dogon (Tommo So), Dong, Dumbea, Dumi, Dutch, Dyaberdyaber, Dyirbal, Dyugun, Emberá (Northern, Napipi River), Emberá-Saija/Epéna Pedée, Engenni, English, Ese Ejja/Essejja, Eton, Ewe, Ewondo, Finnish, Fon, French, Fula/Pulaar, Fyam/Fyem, Garawa/Garwa (Eastern), Garawa/Garwa (Western), Garo, Georgian, German, Gikuyu, Gonja, Gooniyandi, Grebo, Greek (Modern), Greenlandic (West)/ Inuktitut (Greenlandic)/Kalaallisut, Guahibo/Sikuani, Guambiano, Guarani, Guayabero, Guaymi, Guguyimidjir, Gujarati, Gujari, Gumbainggir/Kumbaingar, Gunwinggu/Mayali, Guro, Gwari/Gbagyi, Halkomelem (Upriver), Hatam/Mansim, Hausa, Hawaiian, Haya, Hdi, Hebrew (Modern), Huave (San Mateo del Mar), Hungarian, Hunzib, Hupde/Hupda, Idoma, Igbo, Ignaciano, Ijo/Izon/Kolokuma, Ik, Imonda, Indonesian, Inga, Ingush, Iquito, Iraqw, Irish, Itene/More, Itonama, Iu Mien, Jakalteko (Western)/Popti', Japanese, Jaru/Djaru, Javanese, Jola-Fonyi/Diola-Fogny, Jowulu, Ju/hoan, Jula, Kadiwéu, Kaixana, Kalenjin/Nandi, Kalkutung/Kalkatungu, Kambara, Kammu/Khmu, Kanjobal (Western)/Akateko, Kannada, Kanoé/Kanoé/Kapishana, Kaqchikel, Karajá, Karok, Kashmiri, Kawaiisu, Kayah (Eastern)/Red Karen, Kayapó, Kayardild, Kenuzi-Dongola/Dongolawi, Kera, Ket, Khalkha/Halh Mongolian, Khana/Kana, Khasi, Khinalug, K'iche', Kilivila/Kiriwina, Kilmeri, Kiowa, Kissi/Kisi, Klamath-Modoc, Koasati, Kobon, Kongo (San Salvador), Korana, Korean, Koreguaje, Koromfe, Kpelle, Krongo, Kuku-Yalanji, Kumak/Nelemwa, Kuna (Border)/Cuna, Kunama, Kunimaipa, Kurdish (Central), Kurdish (Northern)/Kurmanji, Kuvi, Kwamera, Kwini, Ladakhi, Lahu, Lakota, Lango, Latvian, Lavukaleve, Laz, Lele, Lepcha, Leti, Lezgian, Lilloet, Lingala, Lithuanian, Lobi, Loniu, Luo, Luvale, Maba, Macá, Machinere/Mantinera, Macuna, Macushi, Ma'di, Maipure, Maithili, Maldivian/Dhivehi, Male/Maale, Maltese, Mam (Northern, San Pedro Necta), Manam, Mandahuaca/Mandawaka, Mandinka, Manikion/Sougb, Mapuche/Mapudungun, Marghi, Maricopa, Martuthunira/Martuyhunira, Masalit, Mataco (Wichí Lhamtés Nocten), Mataco (Wichí Lhamtés Vezoz), Matis, Maya (Itza), Maya (Yucatán), Maybrat/Mai Brat, Mbum, Meitei/Meithei/Manipuri, Mende, Mikir/Karbi, Mixteco (Chalcatongo), Miya, Mocoví, Mokilese, Mon, Mongonkundo, Monumbo, Motlav/Mwotlap, Movima, Mpur, Muinane, Muna, Mundang, Mundari, Murle, Mutsun/Ohlone (Southern), Mwaghavul/Mupun, Nadéb, Naga (Ao)/Ao, Naga (Sumi)/Sema, Naga (Tangkhal), Naga Pidgin, Nahuatl (Telcelingo), Nama (Khoekhoe), Namia, Nangikurrunggurr/Ngankikurrunggurr, Nateni, Ndyuka, Ngalakan, Ngandi, Ngbaka, Ngiti, Ngiyambaa, Nimanbur, Nomatsiguenga, Noon, Nubi, Nukak Makú, Nung, Nung, Nungubuyu, Nyigina, Nyulnyul, Ona/Selknam, Oneida, Oromo (Haron), Otomí (Mezquital), Ottawa/Nishnaabemwin, Paama/Paamese, Páez, Paiwan, Palauan, Palu'e, Panjabi/ Punjabi, Panyjima/Panytyima, Papiamentu, Paulohi, Paumarí, Pech, Pero, Persian, Piapoco, Piaroa, Pilagá, Pima Bajo, Pipil, Pirahã, Piro/Yine, Pitjantjatjara, Polish, Pomo (Southeastern), Poqom (Poqomam, Poqomchi'), Purik/Purki, Q'anjob'al (Eastern), Qawasqar/Kawasqar, Q'eqchi'/Kekchi, Quechua (Ancash), Quechua (Huallaga), Quechua (Imbabura), Quechua (North Junín), Rama, Rapanui, Rawa, Resígaro, Retuarã, Romani (Vlax/Vlach), Rundi, Russian, Rwanda, Saami (Northern), Sabanês, Sakirabiá/Mekens Sakirabiat, Sáliba, Samo (Southern), Sandawe, Sango, Sanumá, Sapuan, Saweru, Secoya, Semelai, Shilluk, Shipibo-Conibo, Shona, Shoshone, Sinaugoro, Sipakapense, Siriono (Yuqui/Jorá), Slave, Somali, Songhay (Koyra Chiini; Koyraboro Senni), Soninke, Soo, Sorbian (Upper), Sotho (Northern), Sotho (Southern), Spanish, Squamish, Suena, Supyire, Susu, Swahili, Swati, Sye/Sie/Erromanga, Taba/Makian (East), Tacana, Taliwang, Tamil, Tariano, Tatar, Tauya, Telugu, Tepehuna (Southeastern), Terêna, Teribe, Thai, Themne, Ticuna, Tigre, Tiri/Tinrin, Tiwi, Toba, Tol/Jicaque, Toura, Trió, Trique (Copala), Trumaí, Tshangla, Tsimané/Mosetén, Tsimshian (Coast), Tswana, Tucano, Tukang Besi, Turkana, Turkish, Tuvaluan, Tuyuca, Tzotzil, Tzutujil, Udihe, Una, Ura, Urarina, Urdu, Urubú-Kaapor, Usan, Vai, Vaimo, Vengo, Vietnamese, Waiwai/Wai Wai, Walapai, Walmajarri, Wambaya, Waorani, Wappo, Warao, Wardaman, Warekena/Guarequena, Wari/Pakaásnovos, Warrgamay, Warrwa, Warumungu, Wayampi, Wayuu/Guajiro, Welsh, Wichita, Witoto (Murui/Bue), Wiyot, Wolof (Dakar)/Waro-Waro, Woun Meu/Waunana, Wuzlam, Xhosa, Yabem, Yagua, Yámana/Yahgán, Yamphu, Yanyuwa, Yapese, Yaqui, Yawuru, Yele, Yidiny, Yokuts (Wikchamni), Yoruba, Yuchi, Yucuna, Yukaghir, Yukpa, Yukulta/Ganggalida, Yup'ik, Yurok, Zoque (Chimalapa), Zulu, Zuni, !Xóǀ