

Minimal systems of Interrogative Words

Michael Cysouw

Max Planck Institute for Evolutionary Anthropology

cysouw@eva.mpg.de

Workshop on Questions, Leipzig

Friday, 15 October 2004

1. Preliminaries

Wer, wie, was?

Der, die, das!

Wieso, weshalb, warum?

Wer nicht fragt, bleibt dumm!

(German Sesame Street)

This talk is a progress report on an investigation into the semantic distinctions that are made in questionwords. This investigation can be seen as a cross-linguistic *Wortfeld* analysis (Trier 1931) of a closed class of items.

The subject of this talk will be the minimal number of questionwords a language might have ('how low can you go?'). Do all languages have questionwords? Which categories are distinguished within the set of questionwords? Which distinctions are minimally made? For a preliminary report of the diversity of possible categories asked for by questionwords among the world's linguistic diversity, see Cysouw (2004a)

(1) Kinds of interrogative 'words'

- uninflected free form
- inflected form (inflecting like nominal, verb, or adjective)
- bound forms (always clitic-like?)

(2) Typology of morphological structure of interrogative words

- 1a) unanalyzable morpheme (*how*, but also *who*)
- 1b) diachronically analyzable (*why*, *weshalb*)
- 2a) analyzable but nonsense semantically (*wofür*)
- 2b) partly analyzable ('raspberry'-type)
- 3a) transparently derived from another questionword (*how much*)
- 3b) identical to another questionword (*how long?* vs. *how did you do it?*)

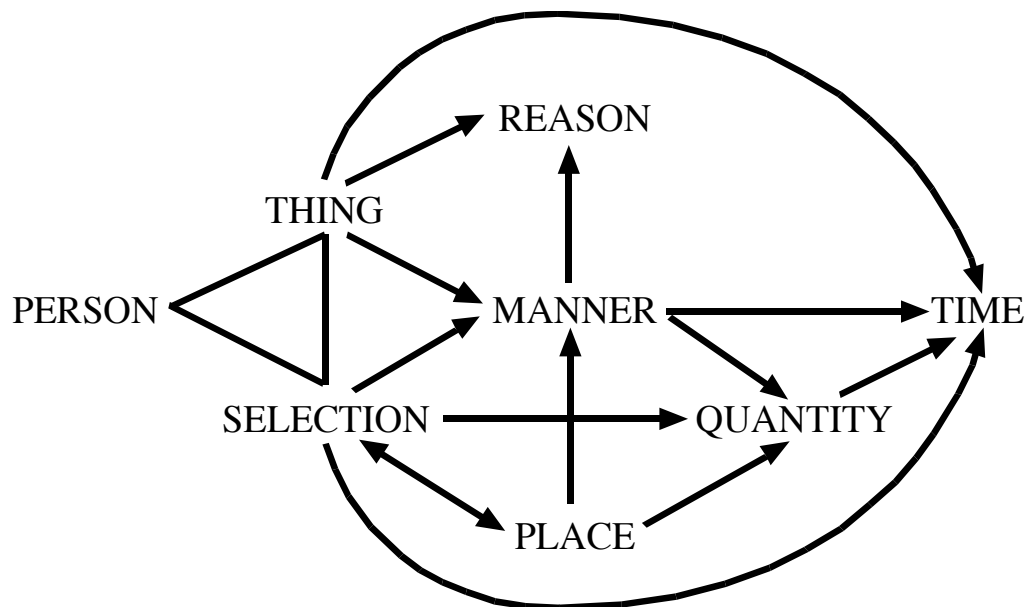
2. Transparency

- Some semantic domains are more commonly expressed by an unanalyzable morpheme than others.
- These empirically major categories overlap, but are not identical to most proposed ontologies (e.g. Aristotle,¹ Heine *et al.* 1991a: 159; Jackendoff 1983: 52-3)
- In general, there appears to be a high innovation rate for interrogative pronouns, also resulting in many near synonyms (e.g. German *wieso*, *weshalb*, *warum*).

(3) Major interrogative categories (estimates of cross-linguistic frequency based on a preliminary, not well-sampled set of languages, see Cysouw 2004a).

| | |
|--------------------------------|--------------------|
| – PERSON (<i>who</i>) | ± 95% unanalyzable |
| – THING (<i>what</i>) | ± 95% unanalyzable |
| – PLACE (<i>where</i>) | ± 90% unanalyzable |
| – SELECTION (<i>which</i>) | ± 60% unanalyzable |
| – QUANTITY (<i>how much</i>) | ± 60% unanalyzable |
| – MANNER (<i>how</i>) | ± 40% unanalyzable |
| – TIME (<i>when</i>) | ± 40% unanalyzable |
| – all others fall below 10% | |

(4) Major pathways of derivation of questionwords



¹ See Aristotle's *Categories*, part 4, as translated by E. M. Edghill – available at <http://classics.mit.edu/Aristotle/categories.1.1.html>

3. Extreme Transparency?

Also called ‘pure transparency’ (Muysken & Smith 1990). In such systems, there is only one questioning element, all others are derived from it.

(5) 18th Century Sranan (English based Creole, Surinam, Bruyn 1993)

| | | |
|--------------------|-------------------------|--|
| <i>hu</i> -NOUN | ‘which’, ‘what sort of’ | |
| <i>(hu)-summa</i> | ‘who’ | (<i>somma</i> ‘person, someone’) |
| <i>(hu)-sanni</i> | ‘what’ | (<i>sanni</i> ‘thing, something’) |
| <i>hu-dissi</i> | ‘which’ | (<i>dissi</i> ‘this’) |
| <i>hu-sortu</i> | ‘what sort of’ | (<i>sortu</i> ‘sort’) |
| <i>hu-peh</i> | ‘where’ | (<i>plehi</i> ‘place’) |
| <i>hu-sei</i> | ‘where’ | (<i>sei</i> ‘place’) |
| <i>hu-tem</i> | ‘when’ | (<i>tem</i> ‘time’) |
| <i>hu-menni</i> | ‘how much’ | (derived from Eng. ‘many’, but not Sranan) |
| <i>hu-fasi</i> | ‘how’ | (<i>fasi</i> ‘manner’) |
| <i>hu-fa</i> | ‘how, why’ | (<i>fasi</i> ‘manner’) |
| <i>va-hu-heddi</i> | ‘why’ | (<i>va</i> ‘for’, <i>heddi</i> ‘head’) |

- Such systems are rare overall (cf. Indo-Pakistani Sign Language, Zeshan 2000: 155-7; Zeshan 2004: 23; and Kenya Swahili Pidgin, Heine *et al.* 1991b: 57).
- Almost complete transparency is more common. Only one questionword is not transparent in e.g. Ewe (Westermann 1930: 164-5; Pasch 1995: 79-80), Nambikuara (Kroeker 2001: 17-9), and Pirahã (Everett 1991: 80-3).
- Widespread transparency is not confined to Pidgins/Creoles, though it might be more common in such languages than in others.
- Note that the English system does not count as transparent, because *wh-* is not a morpheme

(6) English

| | |
|--------|---------|
| wh-o | *th-o |
| wh-at | th-at |
| wh-ich | *th-ich |
| wh-ere | th-ere |
| wh-en | th-en |
| how | (th-us) |
| why | *th-y |

4. Extreme Ambiguity?

Givón claims there to be only one questionword in Asheninka, an Arawakan language from Peru. This questionword *tsika* appears to be used in the same contexts as all different English questionwords.

(7) Asheninka (Givón 2001: 304-5, citing D. Payne, p.c.)

- a. *tsika i-tim-i-ka iri_ ta*
 WH 3MASC-be-TAM-Q 3MASC-there
 ‘**Who** is he (that one)?’
- b. *tsika o-pait-a-ka h-a-ake-ri*
 WH 3FEM-name-TAM-Q 3MASC-take-TAM-REL
 ‘**What** did he take?’
- c. *tsika p-a-ake-ro-ka*
 WH 2-get-TAM-3FEM-Q
 ‘**Where** did you get it (fem)?’
- d. *tsika i-ka_ t-aaiti-ro-ka choclo*
 WH 3MASC-say-TAM-3FEM-Q corn
 ‘**How** do you say “corn”?’
- e. *tsika i-karat-i-ka iri-ka*
 WH 3MASC-be-TAM-Q 3MASC-here
 ‘**How much** is it?/How many are there?’

In fact, the structure of Ashinca turns out to be a special case of transparency. There is hardly any ambiguity possible. I investigated all 206 content questions that appeared in an Asheninka text collection (Anderson 1985/1986). It turns out that the verb following the questionword *tsika* plays a determining role in the meaning of the question (for details, see Cysouw 2004b)

(8) Asheninca content questions (from Cysouw 2004b)

| Structure | Meaning | No. of Cases |
|--|-------------------------|---------------------|
| <i>tsica</i> + NP | be where | 9 |
| <i>tsica</i> + V | where | 4 |
| <i>tsica</i> + <i>-saic-</i> | live where | 4 |
| <i>tsica</i> + <i>-iyaat-</i> | go where | 7 |
| <i>tsica</i> + <i>-a(g)-</i> | take from where | 8 |
| <i>tsica</i> + <i>-quen-</i> | go where | 5 |
| <i>tsica</i> + <i>-quen-</i> | will do how | 3 |
| <i>tsica</i> + <i>-quen-</i> | can do how | 5 |
| <i>tsica</i> + <i>-cara-</i> | how much/many | 3 |
| <i>tsica</i> + <i>-tzim-</i> + NP | be who | 2 |
| <i>tsica</i> + <i>-tzim-</i> + relative clause | which of them | 2 |
| <i>tsica</i> + <i>-cant-</i> | say what | 4 |
| <i>tsica</i> + <i>-cant-</i> | do what/what happened | 8 |
| <i>tsica</i> + <i>-cant-</i> + relative clause | how | 12 |
| <i>tsica</i> + <i>-cant-</i> + relative clause | why | 5 |
| <i>tsica</i> + <i>-pait-</i> | call how | 8 |
| <i>tsica</i> + <i>-pait-</i> + relative clause | who | 4 |
| <i>tsica</i> + <i>-pait-</i> + relative clause | what/which | 12 |
| <i>tsica</i> + <i>-pait-</i> + relative clause | why | 15 |
| <i>tsica</i> + <i>-pait-</i> + relative clause | how | 1 |
| <i>-pait-</i> + relative clause | who | 6 |
| <i>-pait-</i> + relative clause | what | 16 |
| <i>-pait-</i> + relative clause | why | 40 |
| <i>-pait-</i> + relative clause | how | 1 |
| <i>-pait-</i> + relative clause | how could (exclamation) | 4 |
| <i>ipaitaca</i> | how are you/what's up | 4 |
| <i>tsica</i> + relative clause | (various meanings) | 4 |
| Other interrogatives | | 2 |
| No translation | | 8 |
| | Total | 206 |

(9) Summary of interrogative constructions in Asheninca

| Construction | | Interrogative category |
|-------------------|--------------------------------|--------------------------------|
| <i>tsica</i> | | PLACE ‘where’ |
| <i>tsica</i> | - <i>tzim-</i> | PERSON ‘who’ |
| (<i>tsica</i>) | - <i>pait-</i> relative clause | VERB ARGUMENT ‘who/what/which’ |
| | | REASON ‘why’ |
| <i>tsica</i> | - <i>quen-</i> | MANNER ‘how’ |
| <i>tsica</i> | - <i>cant-</i> relative clause | MANNER ‘how’ |
| <i>tsica</i> | - <i>cara-</i> | QUANTITY ‘how much’ |
| <i>tsicapaita</i> | | TIME ‘when’ |

(10) Asheninca (Anderson 1985/1986: I-196)

tsica pi-pait-a-ca

WH 2-call-TAM-Q

‘¿Cómo te llamas?’ (How are you called?/What is your name?)

(11) Asheninca (Anderson 1985/1986: I-136, II-128)

a. *tsica ipaitaca poc-atsi-ri iroñaaca*

WH come-TAM-REL now

‘¿Quién viene?’ (**Who** is coming?)

b. *ipaitaca pi-caim-i-ri incaaran-qui*

WH 2-call-TAM-REL ?-LOC

‘¿A quién estabas llamando?’ (**Who** are you calling?)

(12) Asheninca (Anderson 1985/1986: III-116, III-12)

a. *paitaca p-amemana-tzi-ri-ca jaca no-yovite-qui*

WH 2-search-TAM-REL-Q here 1-casserole-LOC

‘¿Qué estabas buscando en mi olla de barro?’

(**What** are you looking for in my casserole?)

b. *tsica opaitaca ant-aque-mi-ri*

WH do-TAM-1-REL

‘¿Qué te ha pasado?’ (**What** happened to you?)

5. Attested Ambiguities

‘The number and kind of distinctions which QWs [Questions Words, MC] may or may not reflect in terms of those existing elsewhere in a give language vary considerably from language to language, but at least one constrast appears to be nearly universal: Q-pronouns show a human/nonhuman or, in a few cases, an animate/inanimate dichotomy.’ (Ultan 1978: 229)

(13) PERSON=THING ²

- Baltic:
 - Latvian (Nau 1998, 1999)
 - Lithuanian (Lindström 1995: 314)
- Khasi (Austro-Asiatic, India, Rabel 1961: 68-9)
- !Xõo (Tuu, Namibia/Botswana, Güldemann forthcoming)
- Mekens, Sak_rabiat dialect (Tupi, Brazil, Galucio 2001: 166-8)
- Paumari (Arawa, Brazil, Chapman & Derbyshire 1990: 203-16)
- Ika (Chibchan, Colombia, Frank 1990: 82-6)
- Arawakan:
 - Achagua (Colombia, Wilson 1992: 26, 125-7)
 - Apuriña (Brazil, Facundes 2000: 365)
 - Asheninca (Peru, Payne 1989: 151-153; Reed & Payne 1986: 328-330)
 - Bare (Venezuela, Aikhenvald 1995: 25)
 - Nomatsiguenga (Peru, Shaver 1996: 37, 40, 169)
 - Terêna (Brazil, Eastlack 1968: 7-8; Ekdahl & Butler 1979: 190-4)
 - Warekena (Venezuela, Aikhenvald 1998: 261, 325-6)
- Mataco-Guaicuruan:
 - Maká (Paraguay, Gerzenstein 1994: 178)
 - Mataco (Paraguay, Viñas Urquiza 1974: 106-7)
 - Mocoví (Argentina, Grondona 1998: 162)
 - Toba (Argentina, Klein 2001: 23)

(14) MANNER=QUANTITY

- Asmat (Central and South New Guinea, Irian Jaya, Voorhoeve 1965: 157-161)
- Nabak (Huon Finisterre, Papua New Guinea, Fabian *et al.* 1998: 32)
- Wardaman (Gunwingguan, Australia, Merlan 1994: 156, 590)
- Sanuma (Yanomam, Brazil, Borgman 1990: 66-72)
- Apalai (Carib, Brazil, Koehn & Koehn 1986: 56-60)
- Mekens (Tupi, Brazil, Galucio 2001: 166-180)
- Urubu-Kaapor (Tupí-Guaraní, Brazil, Kakumasu 1986: 353-5)

² Sango is quoted by Ultan (1978: 229), but this is not substantiated by the sources (Samarin 1967: 74-5, 217; Thornell 1997: 76).

(15) Asmat (Voorhoeve 1965: 160)

- a. *tirif ucim emamis*
 letters WH lie
 ‘**How many** letters are lying there?’
- b. *makik ucim ématamcém*
 fish line WH will you throw it
 ‘**How** will you throw out the fish line?’

(16) Urubu-Kaapor (Kakumasu 1986: 354)

- a. *myja ne tapi'ir ne ere-juka katu*
 WH 2SG.PRON tapir 2SG.PRON 2SG-kill well
 ‘**How** do you kill a tapir?’
- b. *myja ere-juka*
 WH 2SG-kill
 ‘**How many** did you kill?’

(17) THING=REASON

Migili (Niger-Congo/Platiod, Nigeria, Stofberg 1978: 131, 134)

- a. *á gbé (i)bõ_*
 3SG.PRON buy WH
 ‘**What** did he buy?’
- b. *ó bé (i)bõ_*
 2SG.PRON come WH
 ‘**What** have you come **for**?’

(18) THING=QUALITY

Manam (Austronesian/Oceanic, Papua New Guinea, Lichtenberk 1983: 403)

- a. *ra_ána stúa-lo _u-záza-i*
 WH store-LOC 2SG-buy-3SG
 ‘**What** did you buy in the store?’
- b. *ra_ána waríge áne-_a _u-wáur-i*
 WH rope INSTR-FOC 2SG-fasten-3SG
 ‘**What (kind of)** rope was it that you fastened it with?’

(19) Some other possible ambiguities

THING=MANNER

– Mocoví (Mataco-Guaicurán, Argentina, Grondona 1998: 162)

PLACE=TIME

– Terêna (Arawakan, Brazil, Ekdahl & Butler 1979: 190-4)

QUANTITY (MASS)=TIME

– Hup (Maku, Brazil; P. Epps, p.c.)

6. Conclusions:

- synchronically, transparency of questionwords is ubiquitous
- diachronically, there is a high ‘turn-over rate’ of questionwords
- extreme transparency is rare
- extreme ambiguity does not seem to exist
- ambiguities among the major categories are rare
- the ambiguity *who* = *what* appears to be the most common among them

7. References

- Aikhenvald, Alexandra A. Y. (1995). *Bare*. (Languages of the World/Materials; 100). Munich: Lincom.
- (1998). Warekena. In: Desmond C. Derbyshire & Geoffrey K. Pullum (eds.) *Handbook of Amazonian Languages*. Vol. 4, pp. 225-440. Berlin: Mouton de Gruyter.
- Anderson, Ronald Jaime (1985/1986). *Cuentos Folkloricos de los Asheninca*. (Comunidades y Culturas Peruanas; 18-20). Pucallpa: Instituto Lingüístico de Verano.
- Borgman, Donald M. (1990). Sanuma. In: Desmond C. Derbyshire & Geoffrey K. Pullum (eds.) *Handbook of Amazonian Languages*. Vol. 2, pp. 15-248. Berlin: Mouton de Gruyter.
- Bruyn, Adrienne (1993). Question words in 18th-century and 20th-century Sranan. In: Jaap van Marle (ed.) *Historical Linguistics 1991*, pp. 31-47. (Current Issues in Linguistic Theory; 107). Amsterdam: Benjamins.
- Chapman, Shirley & Desmond C. Derbyshire (1990). Paumari. In: Desmond C. Derbyshire & Geoffrey K. Pullum (eds.) *Handbook of Amazonian Languages*. Vol. 3, pp. 161-354. Berlin: Mouton de Gruyter.
- Cysouw, Michael (2004a). Interrogative words: an exercise in lexical typology. Presentation at “Bantu grammar: description and theory workshop 3”, ZAS Berlin, 13 Februari 2004. Handout available at <http://email.eva.mpg.de/~cysouw/pdf/cysouwQUESTION.pdf>
- (2004b). Interrogatives in Asheninca Campa (Arawakan). Unpublished manuscript available at <http://email.eva.mpg.de/~cysouw/pdf/cysouwASHENINCA.pdf>
- Eastlack, Charles L. (1968). Terena (Arawakan) Pronouns. *International Journal of American Linguistics* 34 (1): 1-8.
- Ekdahl, Elizabeth Muriel & Nancy Evelyn Butler (1979). *Aprenda Terêna*. Brasília: Summer Institute of Linguistics.
- Everett, Daniel L. (1991). *A Língua Pirahã e a Teoria da Sintaxe: Descrição, Perspectivas e Teoria*. (Série Línguas Indígenas). Campinas: Editora da Unicamp.
- Fabian, Grace, Edmund Fabian & Bruce Waters (1998). *Morphology, Syntax and Cohesion in Nabak, Papua New Guinea*. (Pacific Linguistics; C 144). Canberra: Australian National University.
- Facundes, Sidney (2000). *The language of the Apurinã people of Brazil (Maipure/Arawak)*. Ph.D. dissertation, University of New York at Buffalo.
- Frank, Paul S. (1990). *Ika Syntax*. (Studies in the languages of Colombia; 1). Arlington, Tx.: Summer Institute of Linguistics.
- Galucio, Ana Vilacy (2001). *The morphosyntax of Mekens (Tupi)*. Ph.D. dissertation, University of Chicago.
- Gerzenstein, Ana (1994). *Lengua Maká: Estudio descriptivo*. (Archivo de Lenguas Indoamericanas). Buenos Aires: Universidad de Buenos Aires.
- Givón, T. (2001). *Syntax: An Introduction*. Vol. 2. Amsterdam: Benjamins.

- Grondona, Verónica Maria (1998). *A grammar of Mocoví*. Ph.D. dissertation, University of Pittsburgh.
- Güldemann, Tom (forthcoming). !Xõo of Lone Tree: Structure of the unmarked clause. In: Rainer Vossen (ed.) *The Khoisan Languages*. London: Curzon.
- Heine, Bernd, Ulrike Claudi & Friederike Hünemeyer (1991a). From cognition to grammar: Evidence from African languages. In: Elizabeth Closs Traugott & Bernd Heine (eds.) *Approaches to Grammatikalization*. Vol. 1, pp. 149-88. (Typological Studies in Language; 19). Amsterdam: Benjamins.
- (1991b). *Grammaticalization: A Conceptual Framework*. Chicago: University of Chicago Press.
- Jackendoff, Ray (1983). *Semantics and Cognition*. (Current Studies in Linguistics; 8). Cambridge, Mass.: MIT Press.
- Kakumasu, James (1986). Urubu-Kaapor. In: Desmond C. Derbyshire & Geoffrey K. Pullum (eds.) *Handbook of Amazonian Languages*. Vol. 1, pp. 326-406. Berlin: Mouton de Gruyter.
- Klein, Harriet (2001). *Toba*. (Languages of the World/Materials; 179). Munich: Lincom.
- Koehn, Edward & Sally Koehn (1986). Apalai. In: Desmond C. Derbyshire & Geoffrey K. Pullum (eds.) *Handbook of Amazonian Languages*. Vol. 1, pp. 33-127. Berlin: Mouton de Gruyter.
- Kroeker, Menno (2001). A Descriptive Grammar of Nambikuara. *International Journal of American Linguistics* 67 (1): 1-87.
- Lichtenberk, Frantisek (1983). *A Grammar of Manam*. (Oceanic Linguistics special publication; 18) University of Hawai'i Press.
- Lindström, Eva (1995). Animacy in interrogative pronouns. In: Inger Moen, Hanne Gram Simonsen & Helge Lødrup (eds.) *Papers from the 15th Scandinavian Conference of Linguistics*, pp. 307-15. Oslo: University of Oslo.
- Merlan, Francesca C. (1994). *A grammar of Wardaman: a language of the Northern territory of Australia*. (Mouton Grammar Library; 11). Berlin: Mouton de Gruyter.
- Muysken, Pieter & Norval Smith (1990). Question words in Pidgin and Creole languages. *Linguistics* 28: 889-903.
- Nau, Nicole (1998). *Latvian*. (Languages of the World/Materials; 217). Munich: Lincom.
- (1999). Was schlägt der Kasus. Zu Paradigmen und Formengebrauch von Interrogativpronomina. *Sprachtypologie und Universalienforschung* 52: 130-50.
- Pasch, Helma (1995). *Kurzgrammatik des Ewe*. (Grammatische Analysen Afrikanischer Sprachen; 5). Köln: Köppe.
- Payne, Judith (1989). *Lecciones para el Aprendizaje del Idioma Asheninca*. (Serie Lingüística Peruana; 28). Yarinacocha: Instituto Lingüístico de Verano.
- Rabel, Lili (1961). *Khasi, a Language of Assam*. (Humanities Series; 10). Baton Rouge: Louisiana State University Press.

- Reed, Judy & David L. Payne (1986). Asheninca (Campa) pronominals. In: Ursula Wiesemann (ed.) *Pronominal Systems*, pp. 323-31. (Continuum; 5). Tübingen: Narr.
- Samarin, William J. (1967). *A Grammar of Sango*. (Janua Linguarum, Series Practica; 38). The Hague: Mouton.
- Shaver, Harold (1996). *Diccionario Nomatsiguenga-Castellano, Castellano-Nomatsiguenga*. (Serie Lingüística Peruana; 41). Yarinacocha: Instituto Lingüístico de Verano.
- Stofberg, Yvonne F. (1978). *Migili Grammar*. (Language Data, Africa Series; 12). Dallas, Tx.: Summer Institute of Linguistics.
- Thornell, Christina (1997). *The Sango Language and its Lexicon*. (Travaux de l'Institut de Linguistique de Lund; 32). Lund: Lund University Press.
- Trier, Jost (1931). *Der deutsche Wortschatz im Sinnbezirk des Verstandes*. (Germanische Bibliothek; 31). Heidelberg: Winter.
- Ultan, Russell (1978). Some general characteristics of interrogative systems. In: Joseph H. Greenberg (ed.) *Universals of Human Language*. Vol. 4: Syntax, pp. 211-48. Stanford: Stanford University Press.
- Viñas Urquiza, Maria Teresa (1974). *Lengua Mataca*. (Archivo de Lenguas Precolombinas; 2). Buenos Aires: Universidad de Buenos Aires.
- Voorhoeve, C. L. (1965). *The Flamingo Bay dialect of the Asmat language*. Ph.D. dissertation, Universiteit Leiden.
- Westermann, D. (1930). *A Study of the Ewe Language*. London: Oxford University Press.
- Wilson, Peter J. (1992). *Una Descripción Preliminar de la Gramática del Achagua (Arawak)*. Bogotá: Instituto Lingüístico de Verano.
- Zeshan, Ulrike (2000). *Gebärdensprachen des indischen Subkontinents*. (Linguistic Research Forum; 4). Munich: Lincom.
- (2004). Interrogative constructions in signed languages: crosslinguistic perspectives. *Language* 80 (1): 7-39.