



Max Planck Institute  
for Evolutionary Anthropology



# Scholarly Workbench – Ein Baustein : Linguistic Literature

Persönliche Dokumentensammlungen am  
Beispiel der Erschließung linguistischer  
Literatur aus dem Netz für die  
wissenschaftliche Arbeit

Gisela Lausberg, Kirstin Baumgarten, MPI-EVA  
Andreas Gros, MPDL



# Projektpartner



- MPI-EVA, Leipzig  
(Comrie, Cijssouw, Lausberg, Baumgarten)
- MPI Psycholinguistik, Nijmegen  
(Kastens)
- MPDL, Abt. R&D und Digital Collections  
(Dreyer, Gros)

# MPI für evolutionäre Anthropologie

- \*1997, Leipzig
- untersucht die Geschichte der Menschheit
- vergleichende Analysen von Genen, Kulturen, Sprachen, sozialen Systemen, kognitiven Fähigkeiten
- Populationen von Menschen und verwandten Primaten in Vergangenheit u. Gegenwart
- interdisziplinärer Ansatz: 5 Abteilungen

# MPI-EVA, [Abt. Linguistik](#)

- Leitung Bernard Comrie
- erforscht die Diversität menschlicher Sprachen
- Gemeinsamkeiten und Unterschiede einer Vielzahl von Sprachen und Sprachfamilien
- ‚Stammbaum‘ der Sprachen der Welt
- [WALS \(World Atlas of Language Structures\)](#) → [WALS online](#)

# Das Projekt : Linguistic Literature

- Sammlung von linguistischen Arbeiten, v. a. Beschreibungen seltener, entlegener Sprachen
- frei zugänglich im Internet aufgefunden, kein kommerzieller Wert
- meist etwas älter (5 Jahre +)  
i. d. R. im pdf-Format
- Flüchtigkeit: Webseiten der Autoren, Institutionen u. ä.

# Bearbeitung der Dateien - bisher

- Auswahl der elektronischen Volltexte durch Wissenschaftler der Abteilung Linguistik
- Bearbeitung in Bibliothek in drei Schritten:
  1. Aufarbeitung der Dateien
  2. Aufnahme in den Bibliothekskatalog
  3. Einpflege ins Netzwerk

# 1. Aufarbeitung der Dateien

- Datei herunterladen, ggf. in PDF-Format umwandeln, speichern
- Teildokumente zu einer PDF-Datei zusammenfügen
- mit automatischer Texterkennung (OCR) durchsuchbar machen
- Seitenzählung anpassen
- doppelseitig gescannte Dokumente in einseitige umwandeln

### QUADRO I

| Ortografia Prática | Transcrição Fonêmica | Exemplos   |
|--------------------|----------------------|--|
| p                  | /p/                  | i <sup>3</sup> pi <sup>2</sup> - /i <sup>3</sup> pi <sup>2</sup> /<br>"terra"    |
| t                  | /t/                  | ta <sup>3</sup> we <sup>2</sup> - /ta <sup>3</sup> we <sup>2</sup> /<br>"macaco" |
| k                  | /k/                  | ku <sup>3</sup> - /ki <sup>3</sup> /<br>"roça"                                   |
| r                  | /r/                  | o <sup>2</sup> 'a <sup>3</sup> - /o <sup>2</sup> 'a <sup>3</sup> /<br>"machado"  |
| b                  | /b/                  | bi <sup>2</sup> o <sup>3</sup> - / bi <sup>2</sup> o <sup>3</sup> /<br>"anta"    |



### QUADRO I

| Ortografia Prática | Transcrição Fonêmica | Exemplos   |
|--------------------|----------------------|--|
| p                  | /p/                  | i <sup>3</sup> pi <sup>2</sup> - /i <sup>3</sup> pi <sup>2</sup> /<br>"terra"    |
| t                  | /t/                  | ta <sup>3</sup> we <sup>2</sup> - /ta <sup>3</sup> we <sup>2</sup> /<br>"macaco" |
| k                  | /k/                  | ku <sup>3</sup> - /ki <sup>3</sup> /<br>"roça"                                   |
| ʔ                  | /ʔ/                  | o <sup>2</sup> ʔa <sup>3</sup> - /o <sup>2</sup> ʔa <sup>3</sup> /<br>"machado"  |
| b                  | /b/                  | bi <sup>2</sup> o <sup>3</sup> - /bi <sup>2</sup> o <sup>3</sup> /<br>"          |

current expectations or presuppositions. In such a situation the speaker may perform a speech act which attempts to override the hearer's supposed expectations.

The types of information occurring in PM position are listed in (5) (Payne 1992:145):

- (5) Information occurs in preverbal PM position when it is:
- Pragmatically marked: this includes information which is contrastive, questioned, and other similar statuses.
  - Nonidentifiable and the hearer is instructed to open a new active discourse file for it: the speaker may instruct the hearer to open an active discourse file for certain information because he or she anticipates that the information is potentially important.
  - Some nonreferential mentions also occur preverbally.

For example, the question word in information questions occurs in the sentence-initial PM position, as shown in (7).<sup>4</sup>

- (6) *Sa-ŋŋy-nū tā-too-qū ēlūjyāŋ buy(ɛ-sara)* Yagua  
 3sg-give-3sg one-CL bowl-one sweet/drink mix-n.NOM  
 'He gave him one bowl of prepared sweet drink.' (Payne 1990:106)

- (7) *C'hj jikhŋŋya qūm-qū?*  
 who 2sg-give-INAN money  
 'Who did you give the money to?' (Payne 1990:71)

Similarly, non-*wh* NPs that are pragmatically marked (topic or focus) occur in PM position. For example, a fragment of a narrative in (8) describes the situation where Mocyu and the wasp twins are making shelter for the rain to come. The clauses in (8c)–(8d) describe how leaves make a good shelter to guard against rain.

<sup>4</sup>Note that the verb does not immediately follow the PM element in (7) because Yagua places (one class of) clitics in second position.



but mud does not.<sup>5</sup> These clauses contain focal elements (in boldface) that are salient in the discourse in preverbal PM position: the preverbal NP 'mud' in (8d) is clearly contrastive with 'leaves' in the preceding utterance in (8c):

- (8) a. *Nanda-ɛjtŋy jŋta ndco:vaŋu:jŋy jŋy-ɛjŋŋy-day* Yagua  
 3PL-make shelter HIFA wasp-PL COB-place-at-DAY  
 'The two wasps made a shelter for themselves.'
- b. *Sa-ɛjtŋy jŋta nuu-nti-day, Mocyu-day.*  
 3sg-make shelter HIFA other-NP-DAY Mocyu-DAY  
 'The other one also made shelter, Mocyu.'
- c. ***nŋvŋ-ta*** *sa-ɛjtŋy-day*  
 leaves-INST 3sg-make shelter-DAY  
 'With leaves he made shelter.'
- d. ***Mūcodū-ta*** *nanda-ɛjtŋy-day*  
 mud-INST 3PL-make shelter-DAY  
 'With mud they two made shelter.' (Payne 1990:203)

In addition to information questions and focus of contrast shown above, Payne (1990) illustrates with a number of examples from Yagua that the PM position may contain a discourse referent that is (a part of) counter expectation, restatement, added detail/restatement, an answer to an information question, a threat, an assertion which is counter to cultural or situational expectations, and constituent negation – in other words, the element in the preverbal PM position is, in some sense, pragmatically salient.

Quakenbush (1992) makes similar observations about the preverbal position in the Philippine language Agutaynen. Although the basic word order in Philippine languages is verb-initial, SV order (or 'subject-like' elements in preverbal position) is apparently commonly observed in narrative discourse. In Agutaynen, the majority of

<sup>5</sup>The underlined words *ŋŋta* and *ŋŋy* are discourse particles. The former is referred to as a discourse structuring (mid)clitic by Payne (though written as an independent word).

## CHAPTER 2. DISCOURSE CONFIGURATIONALITY

23

but mud does not.<sup>7</sup> These clauses contain focal elements (in boldface) that are salient in the discourse in preverbal PM position: the preverbal NP ‘mud’ in (8d) is clearly contrastive with ‘leaves’ in the preceding utterance in (8c);

- (8) a. Naada-sútay jííta núcovaañu-júy jíy-ajíjy-day. Yagua  
 3DL-make.shelter JIITA wasp-DL COR-place.at-DAY  
 ‘The two wasps made a shelter for themselves.’
- b. Sa-sútay jííta nuu-nti-day, Mocáyuday.  
 3sg-make.shelter JIITA other-REP-DAY Macáyu-DAY  
 ‘The other one also made shelter, Mocayu.’
- c. **návii-ta** sa-sútay-day  
 leaves-INST 3sg-make.shelter-DAY  
 ‘**With leaves** he made shelter;’
- d. **Múcadíi-ta** naada-sútay-day.  
 mud-INST 3DL-make.shelter-DAY  
 ‘**With mud** they.two made shelter.’ (Payne 1990:203)

In addition to information questions and focus of contrast shown above. Payne

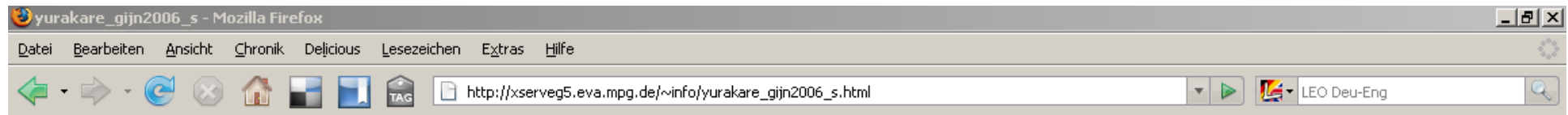
## 2. Aufnahme in den Bibliothekskatalog

- gedruckte Version in Bibliothek vorhanden:  
Links werden in Titelaufnahme des gedruckten Mediums eingefügt
- keine gedruckte Version vorhanden:  
Titelaufnahme als ‚Elektronische Ressource‘ nach RAK-NBM

### 3. Einpflege ins Netzwerk

- Dateien werden auf Server im internen Netz verschoben
- HTML-Seite wird erstellt, die Link zu der Datei enthält

WebOPAC: Grammar of Yurakaré



This material is provided for personal use only for members of the MPI for Evolutionary Anthropology. Unauthorised use or dissemination of this information is expressly forbidden. If you have any questions about this material, or if you want to report a concern, please contact our library at [library@eva.mpg.de](mailto:library@eva.mpg.de).

[yurakare\\_gijn2006\\_s.pdf](#)

When possible, the page numbering of the PDF has been made identical to the page numbering of the original book. In addition, filenames are marked for the searchability characteristics of the PDF:

- the ending `_s` indicates that the PDF was originally created as a searchable PDF.
- the ending `_o` indicates that the PDF has been made partially searchable by optical character recognition.

## Online-Katalog

### Catalogue Display

Back

Simple Search

Advanced  
Search

Clear Search  
Sets

Clear Search  
Term History

Refine Search  
Results

New Items List

Help

German

English (V)

Opening Hours

Exit Webopac

### \*\* Full Display

| Field name                     | Details   |
|--------------------------------|---|
| Document type                  | CF  |
| language used                  | eng   |
| Name 1                         | Gijn, Erik van  |
| Title                          | ↯A↯ grammar of Yurakaré   |
| Author statement               | door Erik van Gijn  |
| year of publ.                  | 2006  |
| extent                         | XI, 361 S.<br>Kt.   |
| thesis                         | Nijmegen, Radboud Univ., Diss., 2006<br>pdf, 2,45 MB  |
| Language(s)<br>described       | yuz   |
| Language(s)<br>described Chain | yuz   |
| subject headings               | Yuracare language   |
| subject headings<br>Chain      | Yuracare language   |
| external link                  | <a href="http://webdoc.ubn.ru.nl/mono/g/gijn_e_van/gramofyu.pdf">http://webdoc.ubn.ru.nl/mono/g/gijn_e_van/gramofyu.pdf</a>       |
| internal link                  | <a href="http://xserveg5.eva.mpg.de/~info/yurakare_gijn2006_s.html">http://xserveg5.eva.mpg.de/~info/yurakare_gijn2006_s.html</a> |

ISBD  
Display

Tag  
Display

Marc  
XML

Add Title  
to  
Basket

Item Information



# Vorhaben



- Einspielen der Daten in PubMan
- weitere formale u. inhaltliche Erschließung:  
Sprachcodes, geographische Informationen,  
Persistent Identifier
- Datenabgleich mit dem lokalen Katalog



# Vorhaben / 2

- Klärung, Dokumentation und Einholen von Nutzungsrechten
- abgestufte Zugangsrechte
- Volltextrecherche über alle Dokumente
- Referenzierung von Stellen, Dokumenten und Sammlungen (Pers. Id.)

# Projektziele

- bessere Nutzungsmöglichkeiten der Dokumente für die wissenschaftliche Arbeit
- Erhaltung der Dokumente
- Entwicklung von nachnutzbaren Arbeitsverfahren
- Vorteile für die Autoren: Sichtbarkeit und Erhaltung ihrer Arbeit in einem fachspezifischen Kontext



# Weitere Entwicklung



- fachliche Erweiterung der Sammlung
- internationale Kooperation



Max Planck Institute  
for Evolutionary Anthropology



# Vielen Dank

Kontakt

[lausberg@eva.mpg.de](mailto:lausberg@eva.mpg.de)



# Linguistic Literature The PubMan View

Andreas Gros (MPDL), Gisela Lausberg,  
Kirstin Baumgarten (MPI-EVA)



## Collection Overview

- Sprachrelevante Texte in PDF-Format
- Geographische Informationen und Sprach-Codes
- Im Volltext durchsuchbare PubMan-Collection



## Geplante PubMan-Erweiterungen

- Persönliche Dokumentsammlungen
- Transparente Zugriffssteuerung für Wissenschaftler
- Geographische Informationen und Sprach-Codes zu Publikationen



## Outlook

- Persönliche Dokumentsammlungen
- Transparente Zugriffssteuerung für Wissenschaftler
- Vorteil für die MPG:
  - Weiterer Baustein zur „Scholarly Workbench“
  - Ermöglichung wissenschaftlicher Workflows



