DOCUMENTARY LINGUISTICS I
Prof. **Tomasz Wicherkiewicz**
UAM winter semester 2017/2018

Thirteenth and Fourteenth lectures
16 and 23 anuary 2018

WHAT KIND OF LINGUISTIC DATA DO WE RECORD, COLLECT, WORK WITH LATER?

❖ events / **texts** / non-texts (words, isolated sentences)

❖ **elicited** (primarily produced for documentation) / **observed** (produced for other purposes) / in between

❖ grade of preparation, **planning** (spontaneous / planned)

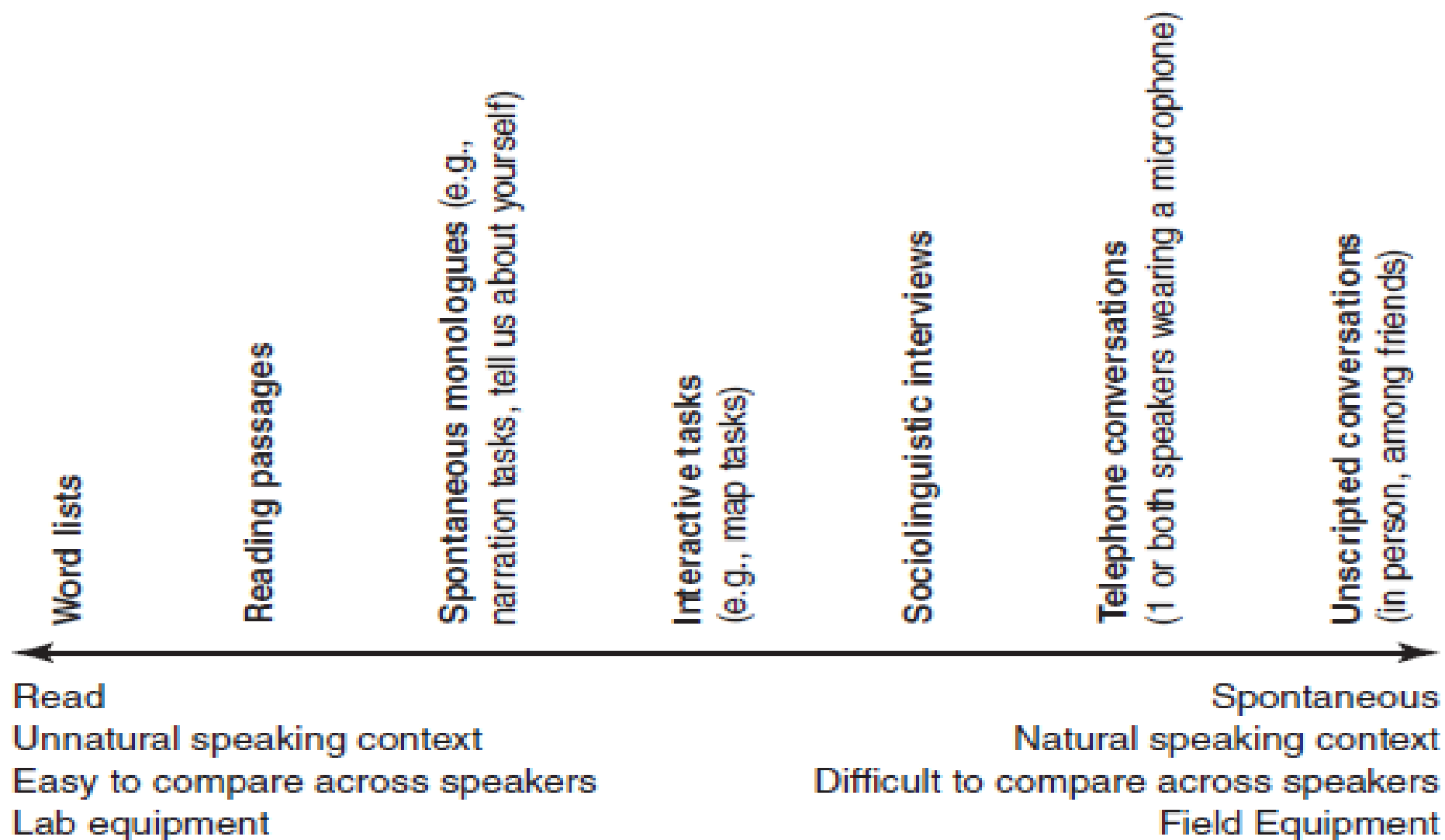❖ **mode(s)**: spoken / written / signed / multimodal

Figure 9.4. *Range of data collection scenarios*

WORD LISTS

All word lists should contain, on their first page, basic information.

You will need to include the METADATA about the list and the researcher:

- date

- language name = LINGUONYM

- alternate language names

- location of the language (including the wider geopolitical names, e.g. district, county)

- origin of the word list (this may not be the place where it's elicited)

- researcher's name

When writing down the location of the language group and the place where the list is being collected, be as detailed and informative as possible.

Include the country, the province or region, and all relevant lower level political divisions as well as the name of the specific village.

information about the participant/s

bear in mind that you may be required to protect the identity of your participant/s.

If this is the case for any reason, you should list the following details not on the front page of the word list but in a separate notebook and to assign a code for each participant or group of participants. On the word list front page, write the code to identify which participant/s the data relates to. Include the following information about the participant/s:

name

age

sex

place of birth

present residence

travel history (i.e. time spent away from the speech community)

learn all of the variations of the names for the languages, dialects, and peoples you will be studying that are identified in Ethnologue.org **and other sources**.

look for previous research or publications in this particular language or language variety, dialect cluster, etc.

search according to variations of the language names discovered in the first step and keep track of any new names that come up. Check in-house for publications your organisation might have which might not be publically available.

visit relevant libraries and look for previous studies of the language you're researching.

copy maps that you find in reading and write in the margins any additional information or possible contradictions that you notice.

record all available information about the sound systems of the [lect] in question and the larger language family to which they belong.

summarize this information and review before starting the survey.

note the names of any speakers mentioned who live outside the language area and the names of outsiders or organizations who work in or frequently visit the area. If possible, contact them for more information.

good reasons to exclude certain words from your list

Non-local concepts

- Don't include words for concepts that are unlikely to be part of the language.
- For example, there may be no words for "snow" or "apple."

Multiple items with the same roots

- For example, the pairs 'bark'- 'skin', 'hair'- 'feather', and 'blood'- 'red'.
- In each set, you might find that the local words are the same, or have the same root with different modifiers.
- Eliminate all but one from each set because eliciting more than one will not add any new data.

One-to-many word mappings

- One word list item might represent only one word but in the local language, that item uses many words.
- Such words will be difficult to elicit consistently unless you make the word list item more specific.
- For example, if you ask for the word for "to carry", you might have a problem because many languages have a number of words for this action and no single generic word for it.
- They might have "to carry on one's back", "to carry on one's shoulders", etc.

- Pronoun paradigm.
- one may ask for their word for "I", but the local language might have many different words for this depending on age, gender, status, etc.
- one might get the male pronoun in one language and the female pronoun in another. You want to be basing lexical similarity judgments on words that were elicited for the same concept. Alternatively, you might have more than one word on your word list that are commonly referred to by the same word in the local languages. For example, "woman" and "wife". All but one of these words should be excluded from your word list.

Semantic range differences

- For example, what you call "blue" might be what they call "green".
- Their word for "arm" might include all the parts of the body from the shoulder to the end of the fingers (whereas in English, there are the words "arm" and "hand").
- There might "semantic shifts" where the meaning of a word in some varieties has changed.
- In such cases, the truly cognate word pairs would not be elicited by the same gloss.
- A different word could have been adapted (or borrowed) for the old meaning.
- Thus, there will be a seeming non-similarity caused by semantic shift.

Compound words

- Eliminate words that in the local languages are typically compound words which include words already occurring elsewhere in the list.

- Such words do not add any new information.

- For example, the local word for "branch" might be "tree" + "arm". Some nouns might just be a verb + nominalizer (or a verb might be a noun + verbalizer).

- If you already have the root word on the list and languages in this area typically use a nominalizer/verbalizer for that word, then it will not add any new information.

Taboo/Embarrassing words

- Some words are not allowed to be spoken by certain people.
- Other words are spoken, but are embarrassing to say to a stranger.
- You do not want to make the subject feel uncomfortable.

Onomatopoeia

If words are derived from sounds, for example, if the word for 'cat' is just the sound a cat makes, then you could end up with words that are lexically similar but not because of any linguistic relatedness.

Elicitation difficulty

Eliminate any other words which you conclude are difficult to elicit consistently across the languages of the region.

One possible reason is if the word in the language of elicitation is not well known and/or the concept is not easy to demonstrate in any other way (e.g. by a picture, or by acting).

You may need to adopt a frame to elicit specific types of items such as numbers and colors which might not ever be said in isolation.

For example, if numbers are usually used with a noun, then you can use "one person" as the elicitation prompt rather than just "one."

You might also need a frame for eliciting tones.

Often, tones are more comparable when they occur in the same environment.

If you want to exclude words based on the principles in this section before you collect data, then you will need to look at previous research and pilot test the word list. Problematic words that you did not know about until after the fieldwork could also be dropped in the analysis stage.

What you want to research is a **language variety**

But in fact, what you do is interview people.

Therefore, it is crucial to make sure that your participants really represent the language variety you are interested in.

When you meet a potential word list participant, ask some SCREENING QUESTIONS.

In general, a word list participant should be representative of their L1, the language variety you are studying.

What this actually looks like will vary.

it might be that there has been a lot of migration due to civil unrest, or that everyone marries someone from somewhere else.

typically, the following criteria will work to ensure that the person knows and uses the local variety:

Born in that village

Grew up in that village

If they have lived elsewhere, it is not a significant amount of recent time because, if so, this influences loan words and fluency.

If you feel this subject represents a village other than the one you are in, note that. You could dismiss the subject, or go ahead and collect a word list if you are interested in that variety and may not be able to go there.

L1 was their first language

L1 is their best language

Both parents are L1 people from that village

Both parents spoke L1 to the subject as a child

Spouse is an L1 person from that village

Is the right age and gender for the population you want to sample. It is good to choose a specific age and gender combination and try to find an participant to match that combination. In this way, your word lists will be more comparable. Men and women tend to have different pitches and languages do change over time. The old might use a form that the young no longer use. Which you want, again, depends on your purpose. Also, young people might not know some words in their language due to lack of experience. You could always try to get a 40-50 year old man, for example.

Educated to the level you require for your sample.

empowerment models
(- one of previous lessons)

that community members become **co-researchers with outsiders,** not passive participants that we do research on, but increasingly active people who the research is done by.

This will require training and skills transfer to develop capacity locally, however it is important to remember that training is specialized activity and tile fact that a researcher has certain skills does not mean necessarily that they are able to teach them or train others.

We may need to learn how to train.

outcomes of 'mobilization'
— one of previous lessons

Communities want the following from information and communication technology aspects of documentation projects, especially in the context of multimedia products:

• the sound of spoken language

• product development processes that respect people's 'ownership' of language

• products that represent tile community's relationship to the language by implementing meaningful pathways between information providers and users

• a range of diverse and adaptable products from comprehensive linguistic and cultural multimedia documentations ... to learning resources, songs, games, and even spelling checkers

• products that are easy to use.

[Nathan, David & Eva A. Csató 2006. Multimedia: A community-oriented information and communication technology. In: A. Saxena & L. Borin (eds.) *Lesser-known languages of South Asia. Status and policies, case studies and applications of information technology*. Mouton de Gruyter, 257-277]

classwork / or homework

1. Choose an ethnolinguistic domain, where your L1 language, or language you know quite well (L2) have a different lexical representation and distribution than English (might be KINSHIP)

2. Find 5-10 words in the domain for which there is no word-to-word dictionary translation into English

3. What words may be more difficult to describe (*define*) than others?

DOCUMENTING «THE LEXICON» OF A LANGUAGE

Word inventories and word meaning

Interdisciplinary cooperation

Linguistic and encyclopedic knowledge

Dictionaries and language documentation

Importance and use of dictionaries

making an inventory of words

describing the meaning of words

describing the use of words

describing relations between words

what? how to?

why? who?

Field Linguistics

A Beginner's Guide

TERRY CROWLEY

Johnson, Samuel 1755.

Dictionary of the English Language

- the grammatical system of a language is said to be more circumscribed than the vocabulary of a language.
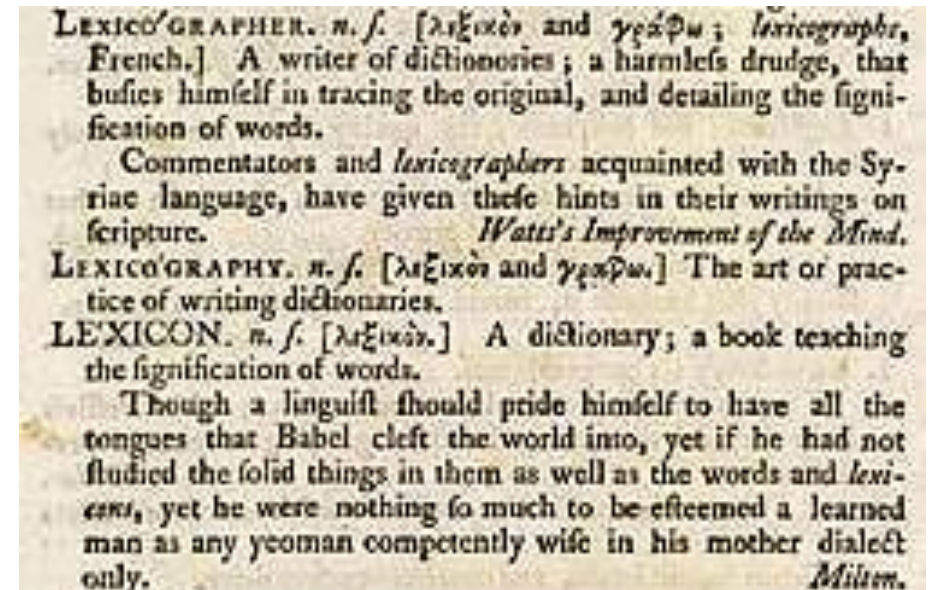
While writing a grammar can be challenging enough, as well as time-consuming, it can be much more difficult to compile a substantial dictionary of a language.

- the difference is reflected in the number of published grammars of languages in comparison to published dictionaries of the same languages.

- some of the most widely cited published grammars have no corresponding published dictionaries.

- some linguists saw the lexicographer's work as beneath their dignity

- there is simply far too much work involved if the job is to be done at all well

LEXICO'GRAPHER. *n. f.* [λεξικὸν and γράφω; *lexicographe,* French.] A writer of dictionaries; a harmless drudge, that busies himself in tracing the original, and detailing the signification of words.

Commentators and *lexicographers* acquainted with the Syriac language, have given these hints in their writings on scripture. *Watts's Improvement of the Mind.*

LEXICO'GRAPHY. *n. f.* [λεξικὸν and γράφω.] The art or practice of writing dictionaries.

LEXICON. *n. f.* [λεξικὸν.] A dictionary; a book teaching the signification of words.

Though a linguist should pride himself to have all the tongues that Babel cleft the world into, yet if he had not studied the solid things in them as well as the words and *lexicons,* yet he were nothing so much to be esteemed a learned man as any yeoman competently wife in his mother dialect only. *Milton.*

lexicographer, 'a writer of dictionaries; a harmless drudge, that busies himself in tracing the original, and detailing the signification of words.'

WHAT IS «THE LEXICON»?

"The lexicon, in modern linguistics, has come to mean a repository of otherwise anarchic facts, an inventory of arbitrary pairings of pronunciations with bundles of features. It is where language stores its idiosyncrasies and irregularities." (Haviland 2006)

• Haviland, John 2006. "Documenting lexical knowledge". In: *Essentials of language documentation*; 129-162.

Glottopedia.org

"The lexicon is the component of a linguistic system which can be regarded as a list or network of words or lexical entries (also lexical items, lexemes). It contains information about

(a) the pronunciation,

(b) the meaning,

(c) morphological properties, and

(d) syntactic properties of its entries.

Furthermore, the lexicon must contain at least the idiosyncratic information about its entries."

"A lexicon is the knowledge that a native speaker has about a language. This includes information about:

• the form and meanings of words and phrases

• lexical categorization

• the appropriate usage of words and phrases

• relationships between words and phrases, and

• categories of words and phrases.

Phonological and grammatical rules are not considered part of the lexicon."

(SIL linguistic glossary:

http://www-01.sil.org/linguistics/glossaryoflinguisticterms/whatisalexicon.htm)

WHY MAKE A DICTIONARY OR A WORD LIST?

Haviland (2006: 132):

colonial tradition

ethnographic tradition

pedagogical tradition

"In what I call the Colonial tradition, collecting vocabularies was always a vocation of imperialists, often an accidental byproduct of exploration and conquest. Explorers collected flora and fauna, and often they also collected words.

Somewhat less innocent were the wordlists created explicitly to aid in conversion, conquest, and control.

The friars' dictionaries of Indian languages in the New World, or vernacular vocabularies destined for colonial bureaucrats in Africa and India, represented unabashedly instrumental "documentation," often of languages whose eventual endangerment was a byproduct of colonial expansion in the first place. Such wordlists were plainly not made "for" the speakers of the languages so documented".

what **WAS** for the „speakers"

PODOBENSTVÍ O HŘIVNÁCH - Matouš 25. kapitola
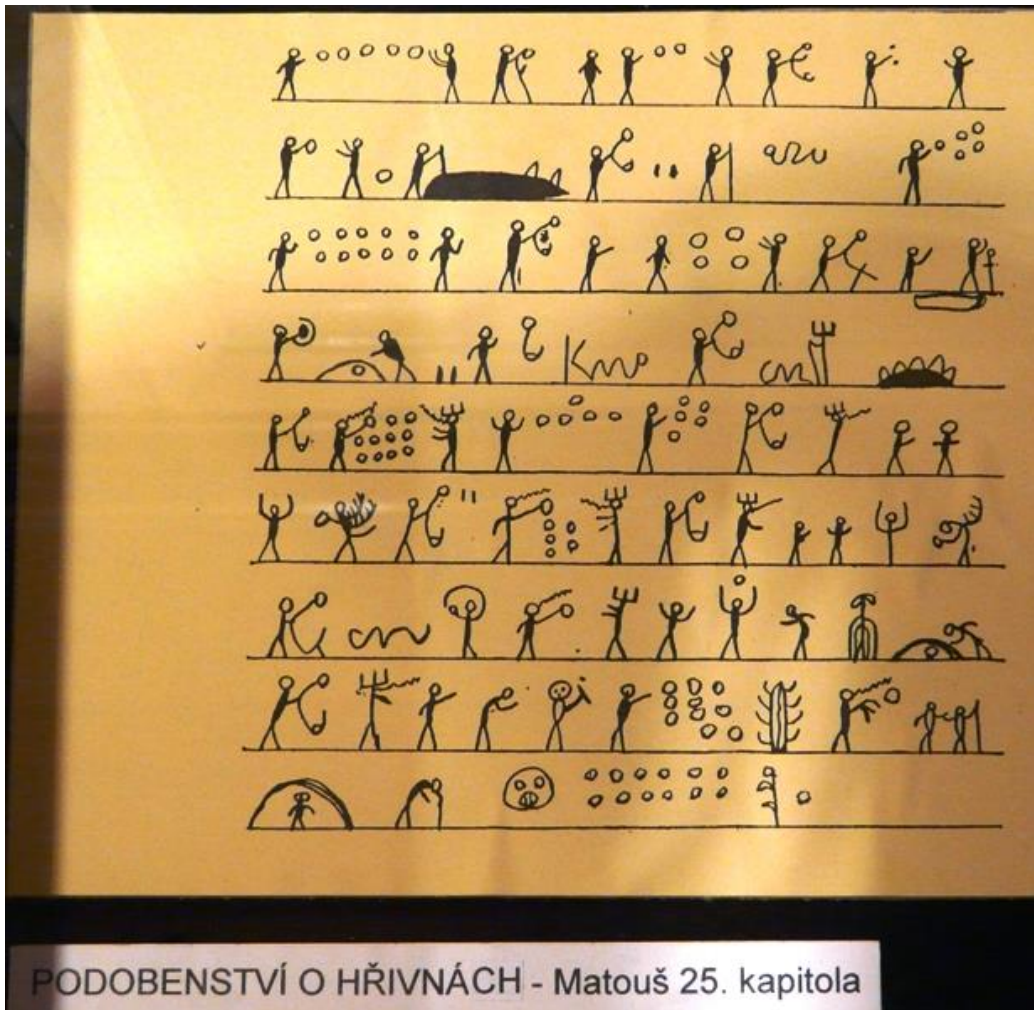
Parable of the talents   *Origin & development*

POČÁTEK A VÝVOJ TLUMOČENÍ EVANGELIA NA ALJAŠCE
od roku 1890
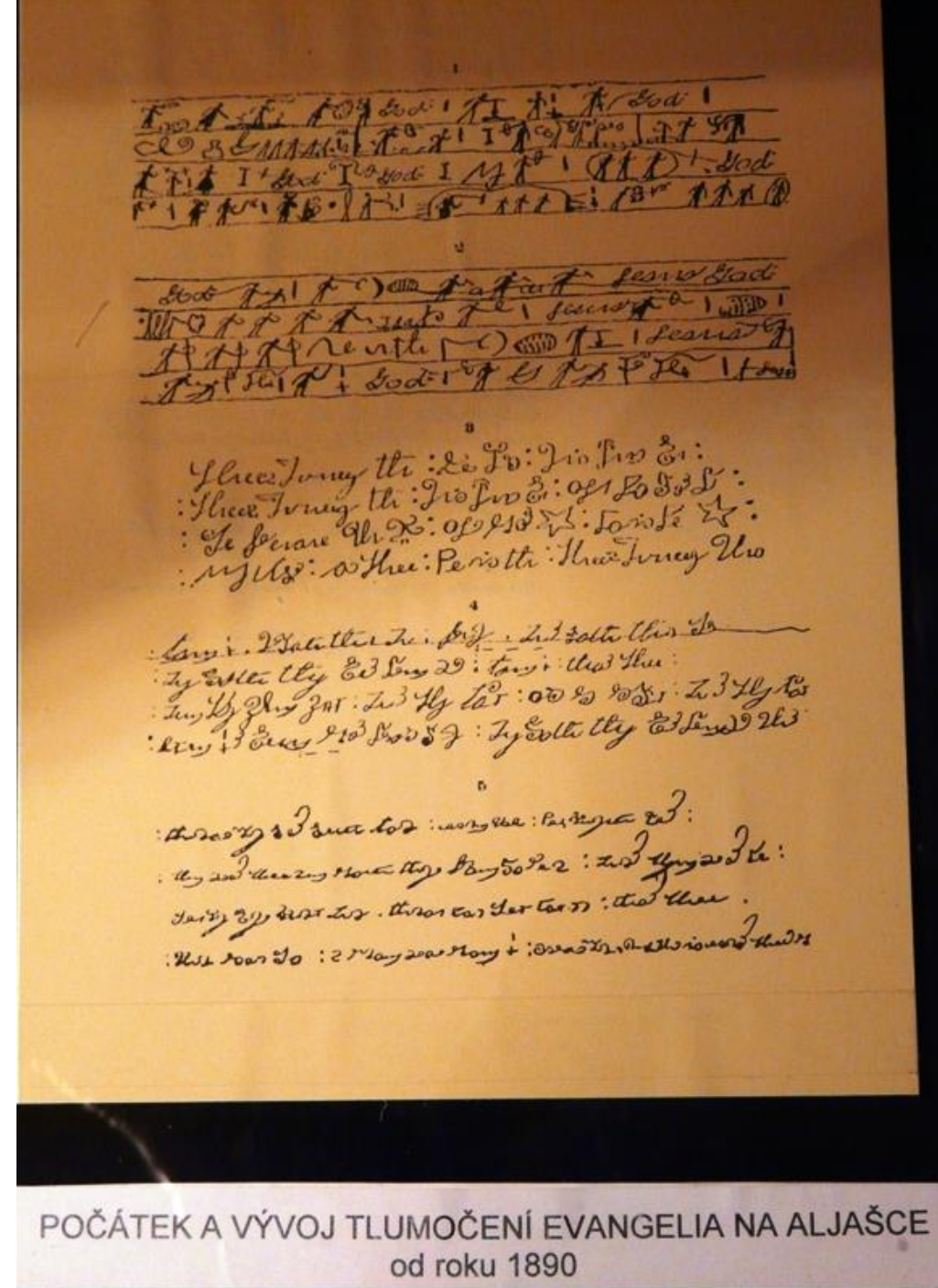
Účelem výpravy Bratří do Ameriky bylo jednak hlásat evangelium. Indiánům, ale také poohlédnout se po nových sídlech, neboť v Evropě byly neustále potíže

The purpose of the Brethren 's expedition to America was to proclaim the Gospel. To the Indians, but also to look after the new settlements, because they were constantly in trouble in Europe

translation of the entry in Aulie and Aulie (1978) for the Chol word *ajaw ajaw*, reflex of a root which means "lord, master, God" in other Mayan languages.

According to the Aulies, the Chol word means "*espíritu malo de la tierra*," and they go on to comment:

They call it *lak tat* 'our father.' It is believed that a person can make a pact with it. Such a person can make requests of the spirit for or against another. The person who establishes such relations with the *ajaw* is called a "sacristán." If a man or woman offends the sacristán, the latter appeals to the spirit to curse the other, and in a short time the other person will die.

mayatecum.com

AJAW

MEANING – SUN, LIGHT, LORD, KING, RULER
DIRECTION – SOUTH/YELLOW

HAAB
CHOL Q'IJ   WWW.STILLALIVE-2012.COM

… both the lexicographers' voice and its underlying ideological accent are plainly on display. Thus, for the Aulies there is no apparent dissonance between their proposed gloss, "evil spirit of the earth" and the alternate locution "our father" (with a first-person plural inclusive prefix). Furthermore, the 'they' of the comment is clearly someone other than the dictionary writers (though perhaps not different from the dictionary users). Note finally an interesting voicing contrast. Although the possibility of "making a pact" with ajaw is cited as something "believed" (presumably by 'them'), the consequences of the appeal on the part of the hypothetical sacristán (the term itself a Spanish loan introduced into Chol during the Catholic conversion of Chol speakers following the Conquest) are given a different epistemological status: "in a short time the other person will die." The dictionary thus incorporates different, perhaps mutually contradictory stances towards Chol beliefs and practices into the lexical entries themselves.

"ethnolinguistic" lexicographic tradition

immediate origins are in ethnographic research > traditional bilingual dictionaries

extensive glosses (in English) of Tzotzil words (over 35,000 Tzotzil to English entries), both derived and simple, and arranged under their putative underlying roots
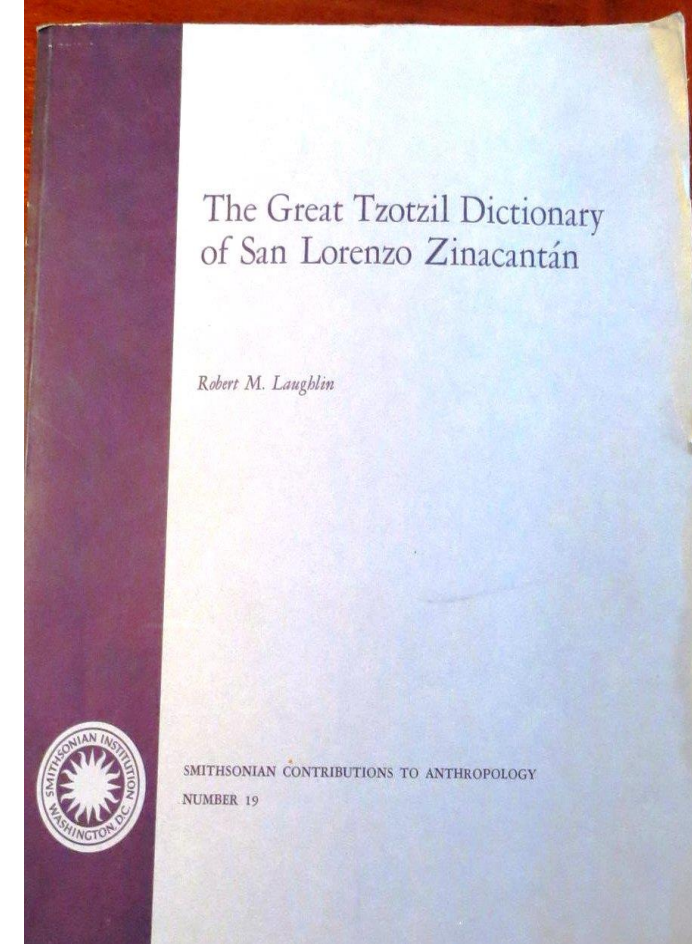
English index to the Tzotzil section.

one of the largest dictionaries of an indigenous language of the Americas.

However, it is a bilingual dictionary in Tzotzil and *English*, limiting its direct use to the handful of people who speak those two languages.

It is also a defiantly dialect-bound (and even gender-bound) dictionary, documenting the way middle-aged men spoke during the 1960s and 1970s in just the single municipality of Zinacantán, arguably a minority variant of what has since become a dominant Indian language in highland Chiapas with a much larger number of speakers from other dialects.

Thus, the choice of language variety in the dictionary reflects accidents of the background research rather than principled lexicographic or sociolinguistic design. Moreover, grouping entries by a theoretical underlying root (a form which does not occur in speech, having only psychological rather than surface "reality"), and stripping words of all affixes – i.e. **lemmatizing** them – makes locating a word in this dictionary something of an analytical challenge, again, a reflection of the intellectual priorities of its producers, but with possibly inconvenient consequences for many potential Tzotzil-speaking users.

The Great Tzotzil Dictionary
of San Lorenzo Zinacantán

Robert M. Laughlin

SMITHSONIAN CONTRIBUTIONS TO ANTHROPOLOGY
NUMBER 19

Lemmatisation   **Headword**

grouping together the inflected forms of a word so they can be analysed as a single item, identified by the word's lemma, or dictionary form …

headword, lemma, or **catchword** – the word under which a set of related dictionary entries appear.

Not to confuse with *head word* (HW; governor) in dependency grammar

a different variant of the ethnolinguistic wordlist illustrates Haviland's another aspect of the field lexicographer's dilemma - the now defunct Barrow Point language.

many linguists have documented Australian Aboriginal languages with very few remaining speakers, often not fully fluent.

wordlists reflect spontaneous opportunity more than systematic planning, and coverage is spotty, based on coincidence and luck.

nonetheless, even haphazardly assembled lists of words may be significant when political processes – – use linguistic evidence to establish links between land and Aboriginal culture and society

for example, "native title" claims to traditional Aboriginal territory

everything from a place name to a plant name may turn out to have unsuspected relevance.
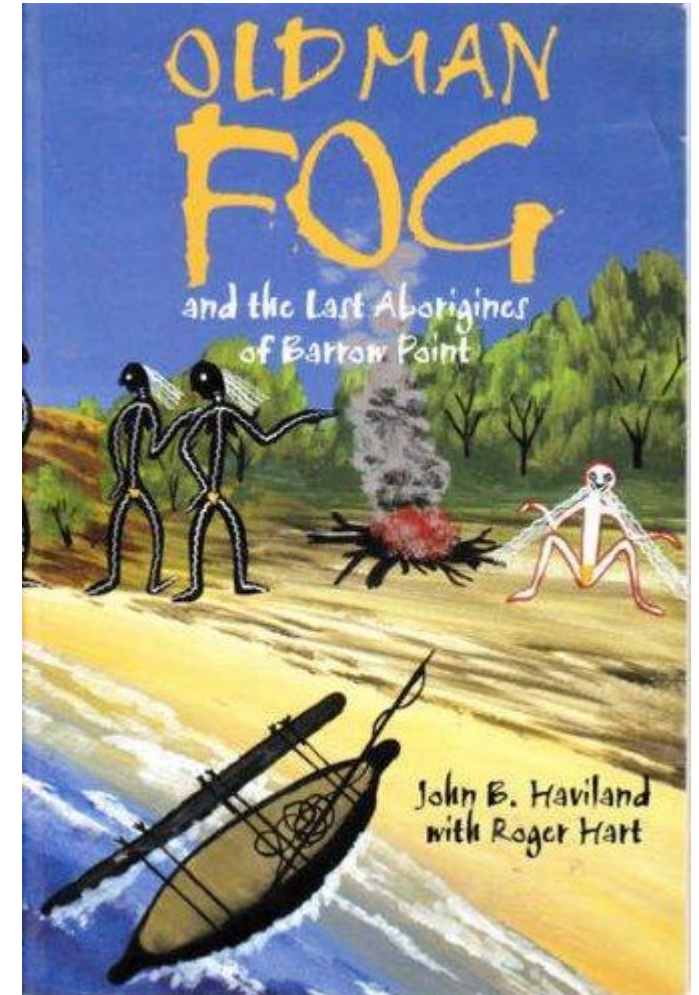
咼壺 (Wang & Pan 2005:287), ᔓᔓᔓ (Wang 1994:55)

tai²pjaːi²; tai¹pjaːi²

> *n* Sequence of Deaths, 丧排 (*Sāngpái*), in Chinese sources named as 重丧 (*Chóngsàng*) or 代排 (*Dàipái*)

> Name of a demon which causes a chain of deaths; if a person dies or a funeral is held on the day affected by its presence, the date must be ritually changed, otherwise people will die one after another (Lu 2010:14). Fragment of the "Sequence of Deaths" passage: 𐤥𐤥𐤥𐤥𐤥 𐤥𐤥𐤥𐤥𐤥𐤥~𐤥𐤥 𐤥𐤥~𐤥 çi³ ŋo² maːu⁴ ju² njen² haːm¹ ljok ⁸ʨu³ sup⁸ n̠i⁶ maːu⁴ ju⁴ pjaːi² çi³ ŋo² van¹ si² pjaːi² 子、午、卯、酉年三月六月九月十二月卯日酉日卯的丧排、子日、子时和午日、午时是丧排凶 During the third, sixth, ninth and twelfth months of the *zi*, *wu*, *mao* and *you* years, the *mao* and *you* days are affected by the Sequence of Deaths, the *zi* and *wu* days and times bring misfortune of the Sequence of Deaths (Wei 2007:33).

𐤥 𐤥, 𐤥, 𐤥, 𐤥, 𐤥 (Yang & Wei 2010:10), **中** (Liang 2009:16)

ta⁵; tum³ta⁵; ʨoŋ¹

> *n* centre; middle, 中 (*zhōng*)

> *prep* among; between, 中间 (*zhōngjiān*)

𐤥

ta⁵si²maːu⁴

> *n* middle of the period of a day from 5 a.m. to 7 a.m. which is considered to be very auspicious, 正卯时 (*zhèngmǎoshí*)

Adam Mickiewicz University
Faculty of Modern Languages and Literatures
Chair of Oriental Studies
Department of Languages of China

THE SCRIPT AND LANGUAGE
OF THE SUI IN SOUTHERN CHINA
DESCRIPTION, DICTIONARY, RELATIONS
WITH CHINESE LANGUAGE AND CULTURE

Kamil Burkiewicz

Ph.D. dissertation
written under the supervision of
Prof. Dr. Alfred F. Majewicz

Poznań 2015

Photo № 49: A carefully calligraphed and illustrated copy of a book, made from cotton paper, Wei Jiagui's Sui Culture Private Museum in Bangao Village, Sandu County, September 2010.
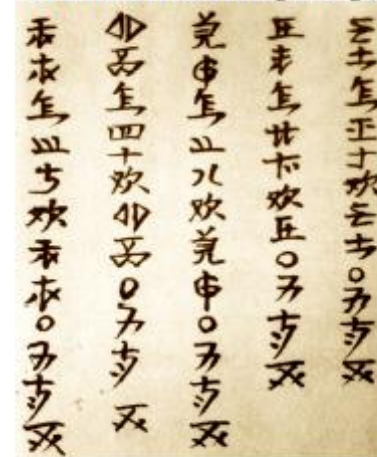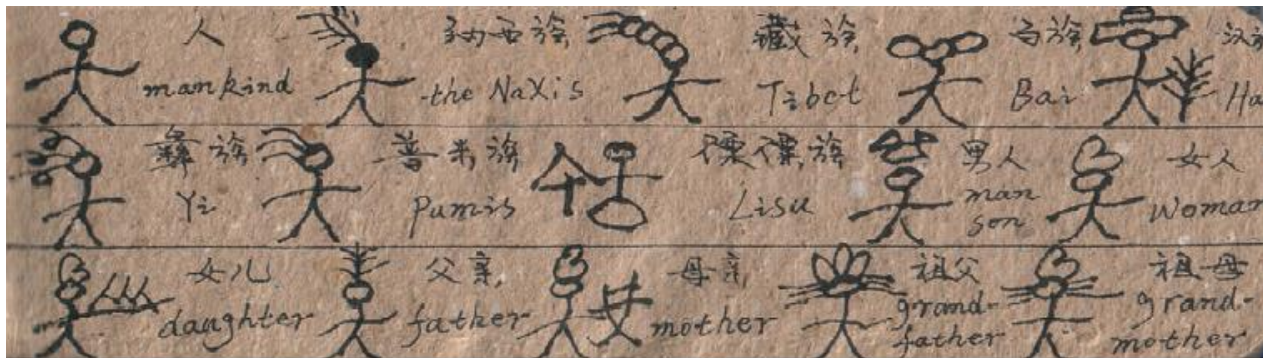

Photo № 50: Sui characters styled on Chinese calligraphy, Wei Jiagui's Sui Culture Private Museum in Bangao Village, Sandu County, September 2010.

Photo № 51: This author receiving a special gift from Professor Pan Chaolin – his own Sui script calligraphy piece, Guiyang, July 2012.

²bä–²ffŭ. Command you to do. 命令你做。
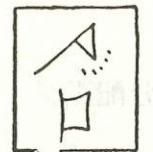
²bä–²gu. Brothers. 弟兄。

²bä–³ho. Allow me to do. 允许我去做。

²bä–³hu. It shall be. 它将是。
See SNL, p. 79. Both symbols serve as phonetics.
见 *See SNL*，第79页。两个符号皆用以标音。

²bä ³hu. May that be (done); may that be (fulfilled), expressing a wish. 可以被做；可以被完成，表示一种愿望。 Instead of the symbol ³hu = stomach, the symbol for gate ²k'u is used and then read ³hu. 用门(²k'u)的符号代替符号 ³hu(胃)，并读做 ³hu。

languagesindanger.eu

Pedagogical tradition in dictionary making

- source of the most common dictionaries: those used by students to look up unfamiliar words, or by tourists to translate menus.

- the question of dimension is telling.

Dictionaries are often explicitly graded by size:

a small version meant for schoolchildren with several thousand "basic" words,

a larger intermediate version with more,

and so on.

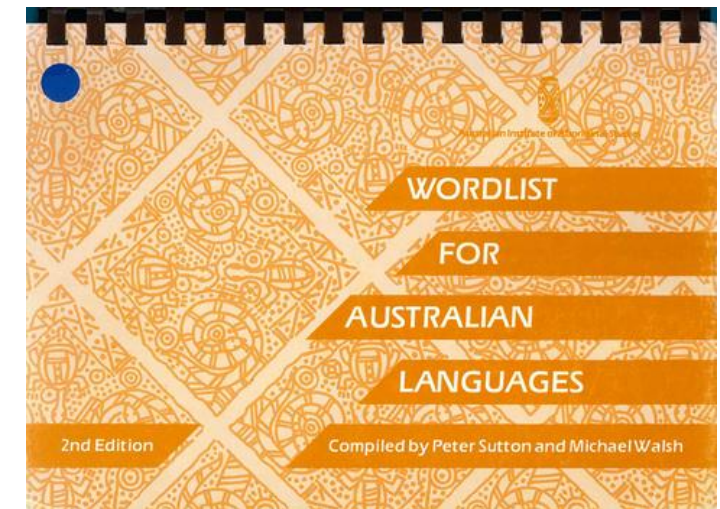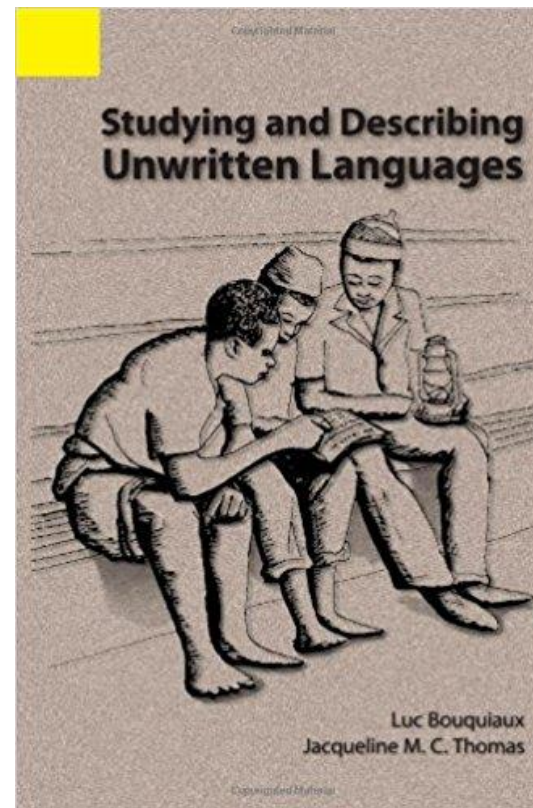in fact - also collonial dimension

Mexico, USSR, Yugoslavia, China

All celebrate Mexican Spanish, the most widely spoken variety of the language, but one relegated to a subsidiary status by the language academy of the colonial home country. The lexicon chosen and the facts of usage are drawn from a huge corpus of Mexican textual material, from letters, to newspaper articles, to popular songs. In Chiapas, the government has similarly commissioned a variety of *"diccionarios de bolsa"* or pocket dictionaries for the Indian languages of the state. These, along with a series of grammatical sketches, are meant as both pedagogical tools and political trophies, evidence of government concern for Indians in the wake of the Zapatista uprising of 1994. Of a similar design but with an opposite ideological thrust are the illustrated school primers, or basic wordlists, designed as literacy aids by Zapatista community schools which resist all government aid and standardized school materials.

How to gather words ?

Direct elicitation - directly eliciting vocabulary via another language

- vocabulary in semantic fields rather than running through a list organized alphabetically

- useful wordlists

Studying and Describing Unwritten Languages

Luc Bouquiaux
Jacqueline M. C. Thomas

WORDLIST FOR AUSTRALIAN LANGUAGES

2nd Edition

Compiled by Peter Sutton and Michael Walsh

It can sometimes be helpful to elicit vocabulary with small groups of people rather than in one-on-one sessions with a single language-helper.

It can be remarkably easy for somebody on their own to be unable to think of a translation equivalent for even a fairly basic vocabulary item, but another person will often be able to step in straight away with the appropriate word.

When one person says one word, this may also remind some of the other people present of other related words.

People in groups can sometimes also find additional senses to words in discussion which might not necessarily occur to somebody working singly with a linguist.

HOW TO COLLECT WORDS?

1. by word lists, dictionaries and translation

2. from texts, corpora

3. generating a list of possible wordforms

4. collecting by concepts, domains

COLLECTING BY DOMAINS
www.rapidwords.net

«Briefly the method involves gathering a group of native speakers of a language in a workshop. The speakers are trained in the method and then work approximately ten days to collect as many words as they can. They are provided with a copy of the list of domains. Each domain includes a few elicitation questions and sample words that help them understand the domain and the type of words that might be included in it. They then use their mental network to think of words in their language that belong to the domain. They write the words down on paper and a typist types them into the program. Approximately 200 language development projects have used the method to collect thousands of words. Most collect well over 10,000 words and one project collected over 20,000. By collecting words in semantic domains, the words are automatically classified. So the result is a classified word list (or what is more popularly called a thesaurus).»

Nettle, Daniel & Suzanne Romaine 2000.
Vanishing voices. The extinction of the world's languages.
Oxford University Press.

"The vocabulary of a language is an inventory of the items a culture talks about and has categorized in order to make sense of the world and to survive in a local ecosystem."

Pawley, Andrew 2011.
„What does it take to make an ethnographic dictionary? On the treatment of fish and tree names in dictionaries of Oceanic languages",
in: G. Haig et al. (eds) *Documenting endangered languages*. De Gruyter

languages of the Pacific islands may have hundreds of **fish names**

that their community has established.

Fishnames and preservation of biodiversity

Group A. Definitions that give 'fish, or 'fish, sea creature' without further definition

Arosi:	**i'a**, a fish.
Cheke Holo:	**sasa**, fish (generic).
Marshallese:	**ek**, fish.
Mota:	**iga**, a fish.
Owa:	**aiga**, fish, sea creature.
Rarotongan:	**ika**, fish.
Roviana:	**igana**, the generic name for fish.

Group C. Definitions that try to be comprehensive

Gela: **iga**, a creature of the sea, fish, mollusc, crayfish, whale, squid, sea anemone, etc.

Paamese: **mesau**, 1. fish. 2 any sea dweller (including also turtles, dolphins, shellfish, etc.).

Toqabaqita: **iqa**, 1. fish (generic term). 2. Also denotes a superordinate category that includes fish, whales, dolphins, turtles, dugongs.

Wayan: **ika**, 1. Typical fish, true fish, syn. **ika dū**. This category includes all gill-breathing fish with fins, including sharks, rays and eels. 2. Fish and certain fish-like creatures. A generic which includes all true fish (see sense 1) and dolphins. Most informants also regard turtles (**ikabula**) as **ika**. Some also include octopus (**sulua**) and squid (**suluanū**). Universally excluded are crustaceans (crabs, lobsters, etc.), molluscs with shells, sea cucumbers, sea urchins and jellyfish. (There follows a full list of names of **ika**.)

☐ ethnobotany; ethnozoology;

«environmental anthropology», «conservation biology»

☐ ethnogeography

☐ ethnoastronomy

☐ ethnography

Ethnobiology is the scientific study of dynamic relationships among peoples, biota, and environments (Salick et al. 2003).

This discipline has departed from being descriptive and now attempts to use the full spectrum of scientific methodologies and tools to understand and explain cultural differences and similarities in the knowledge and use of biota and environments (Bale´e 1994).

Mcclatchey, Will

in: Thieberger, Nicholas (ed.) 2011.
The Oxford handbook of linguistic fieldwork.
Oxford University Press.

Ungerer, Friedrich & Hans-Jörg Schmid 2006.
An Introduction to Cognitive Linguistics.
London: Longman.

Folk taxonomies, or experiential hierarchies, are taxonomies that are part of the common ground of a standard speech community. Prototype categories are essentially folk taxonomies.

Folk taxonomies are based on the speech community members' interaction with the surrounding world. Typically, the units that incude the most subordinate units are place in the middle level of the taxonomy, while the units at the superordinate level and subordinate level of a folk taxonomy subsume less units. Folk taxonomies typically focus on the middle level, because it is culturally salient and sometimes rooted in basic biological needs.

Folk taxonomies and language

Typically the forms that are used as labels for middle level units are morphologically simple in relation to those that serve as labels for superodrinate and subordinate units. Moreover, middle level units are readily available in communication, as they normally come first ti mind in categorizing or naming situations and experiences.

Franchetto, Bruna 2006.
"Ethnography in language documentation",
in: *Essentials of language documentation*

Like most other users of a lexical database, ethnographers will profit from the amount of detail provided in definitions and the care given to the wording of translations.

The search for ethnographically relevant information is facilitated by the definition of *thematic domains* in addition to the *semantic domains* used by many linguists and lexicographers.

Semantic domain refers to a set of features which define very general and inclusive fields of meaning, are often relevant to grammatical marking and which are associated with a large number of lexical entries.

The categories used may be created by the researcher or form part of native classifications.

Example features: ANIMATE, BODY PART, ACTION, or PROPERTY

Although some semantic domains may contain information useful for ethnographic purposes, the more narrowly defined thematic domains will be of greater interest in this regard.

ETHNOTAXONOMIES

- KINSHIP terminology

- BODY PARTS

- DISEASES

- COLOR terms

KINSHIP terminology

These terms, on the one hand, denote positions in a genealogical structure, but they are also inherently relational terms, associated with multiple denotata. The determination of kin relationships is influenced by many variables, such as genealogical distance or proximity, the calculations made through a third relation mediating between ego and the individual being addressed or referred to, as well as contextual and momentary variables, such as factional disputes, broken marriages, extraconjugal affairs, and so on.

A systematic analysis of kinship terminology must include a precise indication of the positions covered by each term in a genealogical structure, using the vocabulary or abbreviations currently used in anthropology.

In addition to kinship terms proper, related general and specific terms are also part of this thematic field.

E.g.: Is there a general term for 'kin'?

Possibly no equivalent of general terms such as 'kin' is found in the language under study, but we may find collective terms in ego's generation, such as terms for male relatives of the same generation (i.e. a cover term for 'brothers' and 'cousins'), and this is a relevant source of ethnographical information.

Kinship Terminologies

kin terms are fundamentally arbitrary categories, different cultures can potentially group their relatives into a widely varying, indefinite number of classifications.

Curiously, anthropologists have observed that almost every culture has constructed a system of terms that conforms to one of six widely occurring basic patterns.
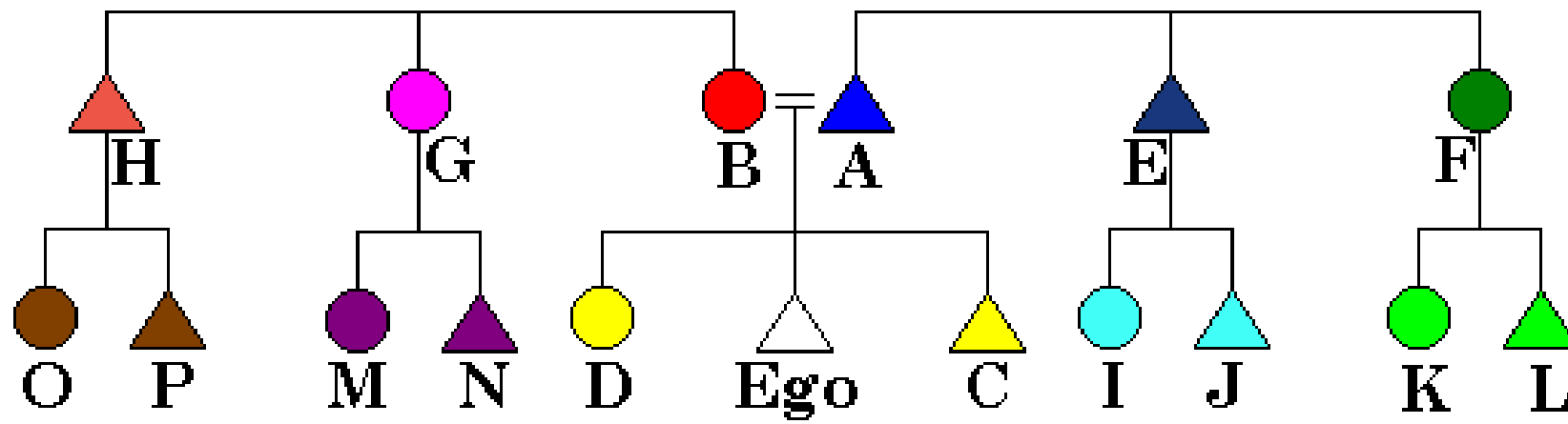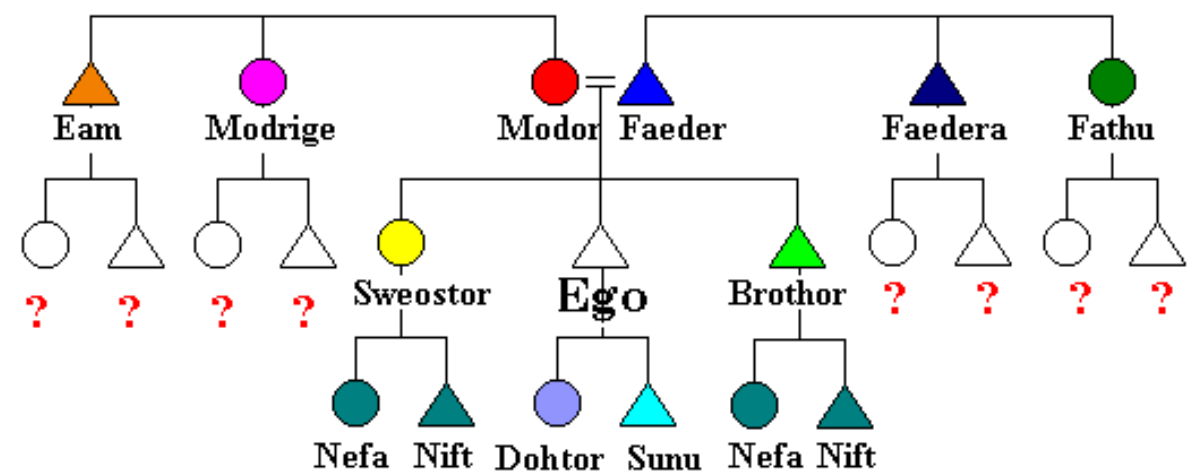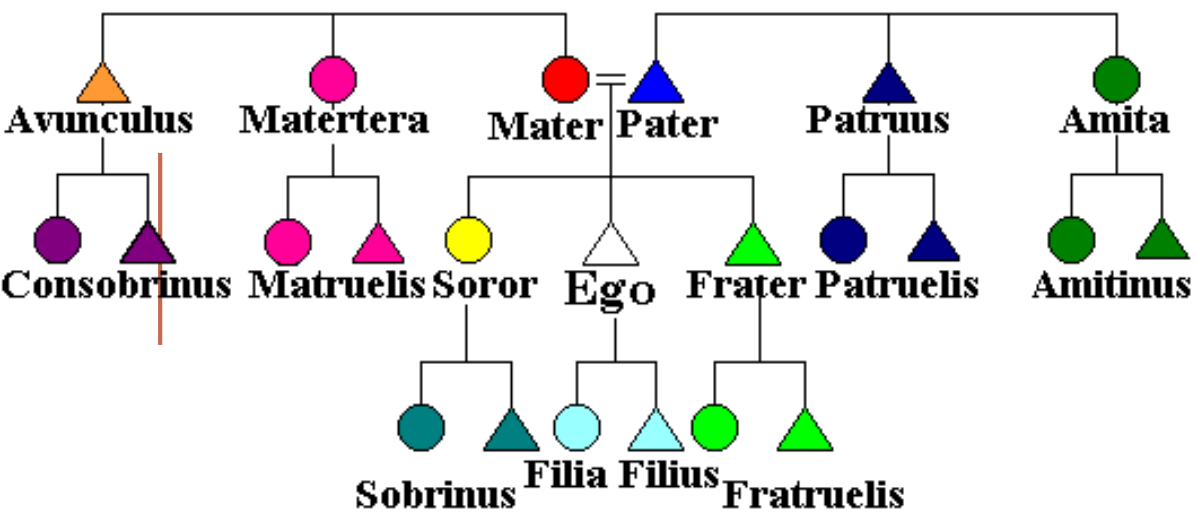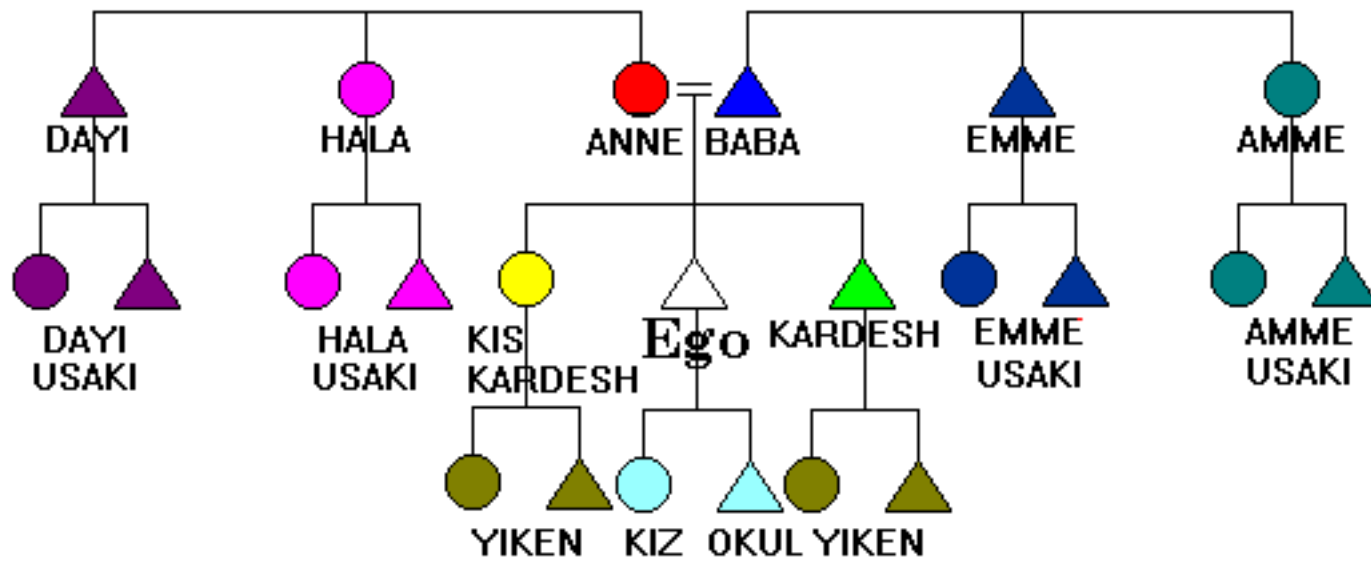
Sudanese

Hawaiian

Eskimo

Iroquois

Omaha

Crow

Sudanese

Kin Term	Kin Type
Pater	F
Frater	B
Filius	S
Patruus	FB
Avunculus	MB
Patruelis	FBS
Amitinus	FZS
Consobrinus	MBS
Matruelis	MZS
Fratuelis	BS
Sobrinus	ZS

Modern English	Kin Type	Old English
Father	F	Faeder
Mother	M	Modor
Uncle	FB	Faedera
	MB	Eam
Aunt	FS	Fathu
	MS	Modrige

parent	ebeveyn / annevebaba
father	baba
mother	anne / ana
child	çocuk
son	oğul
daughter	kız
spouse	eş
husband	koca
wife	karı
sibling	kardeş
brother	erkek kardeş / abi (older)
sister	kız kardeş / abla
uncle	amca (paternal) / dayı (maternal)
uncle's wife	yenge
aunt	hala (paternal) / teyze (maternal)
aunt's husband	enişte
cousin	kuzen
nephew	yeğen
niece	yeğen
grandfather	dede / büyükbaba
grandmother	nine / büyükanne / anneanne (maternal) / babaanne (paternal)
grandchild	torun
great-grandfather	büyük dede / büyük büyükbaba
great-grandmother	büyük nine / büyük büyükanne
great-grandchild	torununun çocuğu
great-great-grandchild	torununun torunu
father-in-law	kaynata / kayınbaba / kayınpeder
mother-in-law	kaynana / kayınvalide
brother-in-law	kayın / kayınbirader
sister-in-law	görümce (according to women) / baldız (according to men)
ancestor	ata
progeny	kuşak
stepmother	üvey anne
stepchild	üvey çocuk
adopted child	evlatlık çocuk

Hawaiian

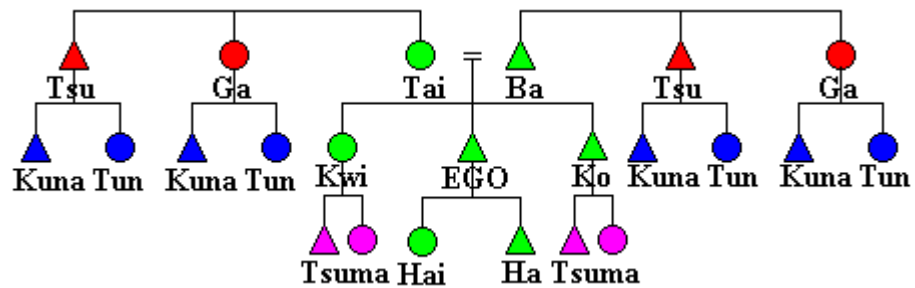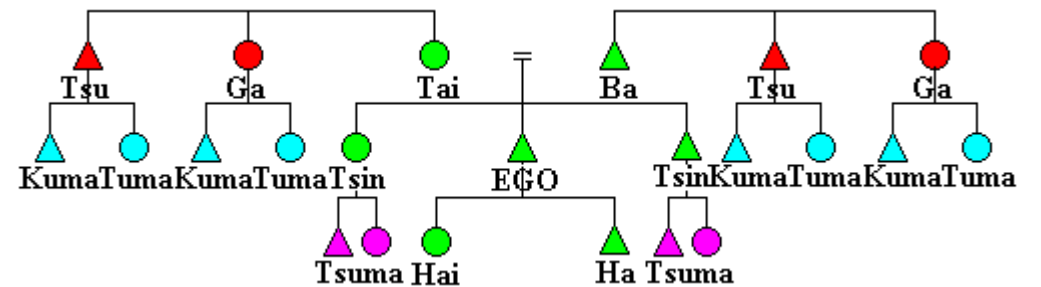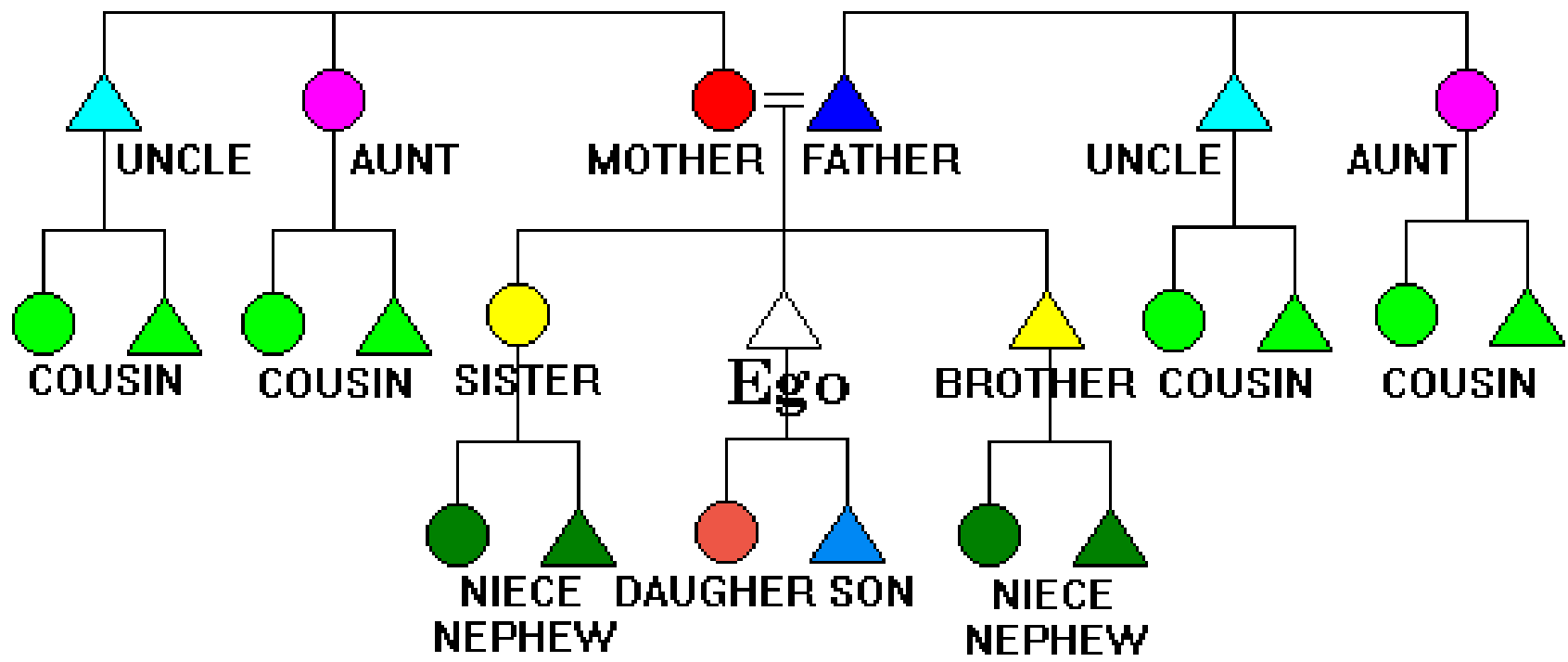| Kin term	Kin types		English terms
Makuakane	F FB MB		**Father**
			Uncle
Makuahini	M MZ FZ		**Mother**
			Aunt
Kaikua'ana	**Male Ego:** • B • FBS • FZS • MBS • MZS	**Female Ego:** • Z • FBD • FZD • MBD • MZD	**Brother/Sister**
			Cousin
Kaikuane	**Female Ego:** • B • FBS • FZS • MBS • MZS		**Brother**
			Cousin
Kaikuahine	**Male Ego** • Z • FBD • FZD • MBD • MZD		**Sister**
			Cousin
Keikikane	S BS ZS		**Son**
			Nephew
Kaikamahini	D BD ZD		**Daughter**
			Niece

Eskimo

Southeastern !Xuun

older cousins
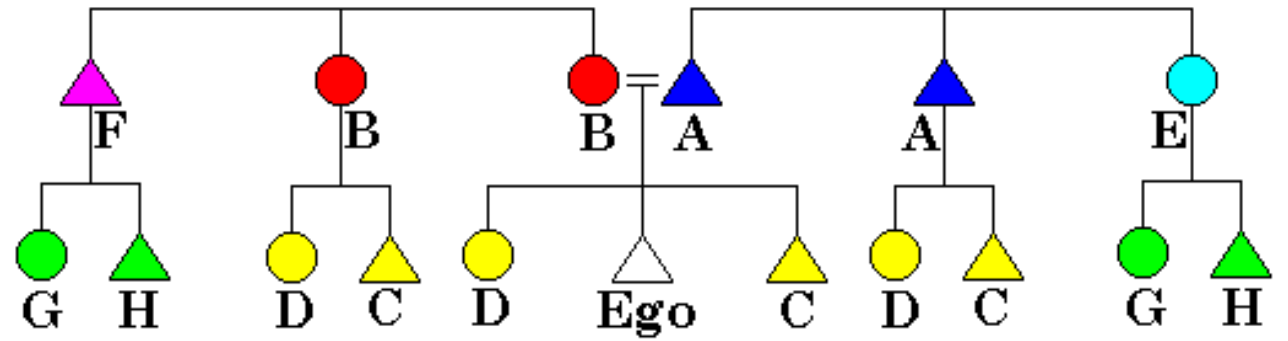
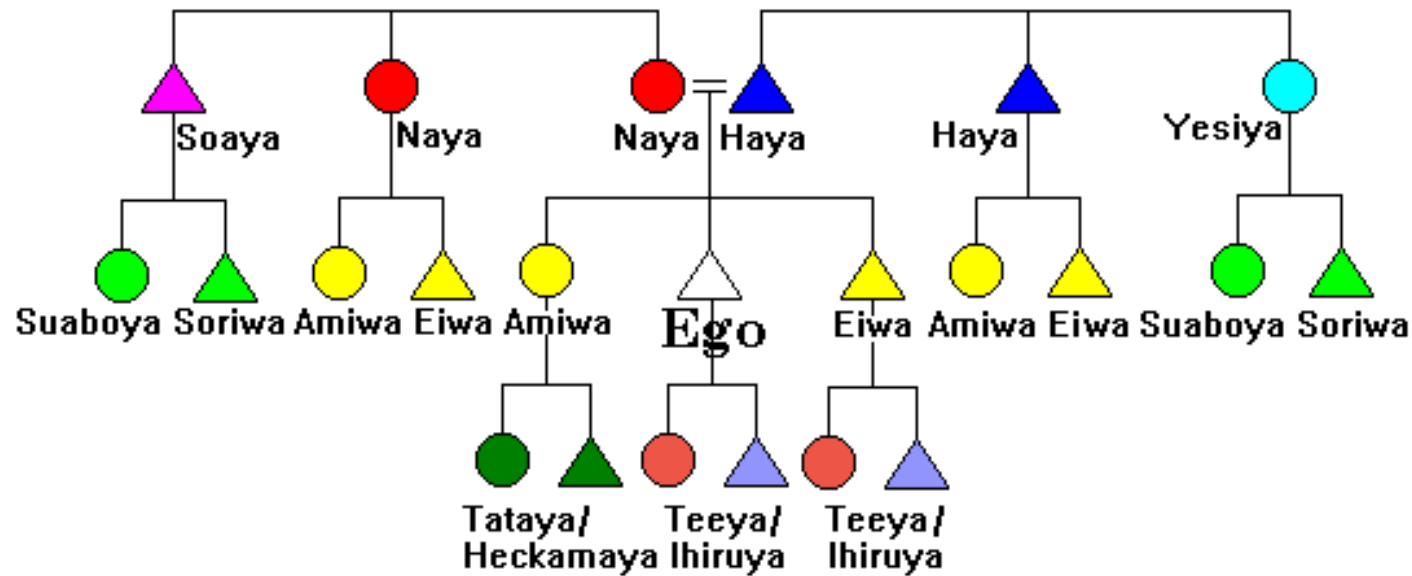younger cousins

English
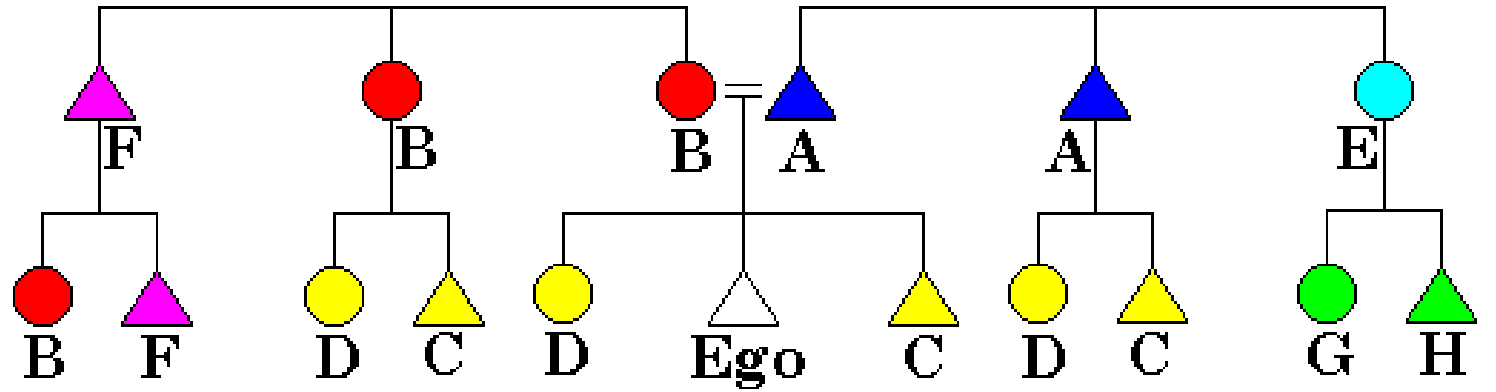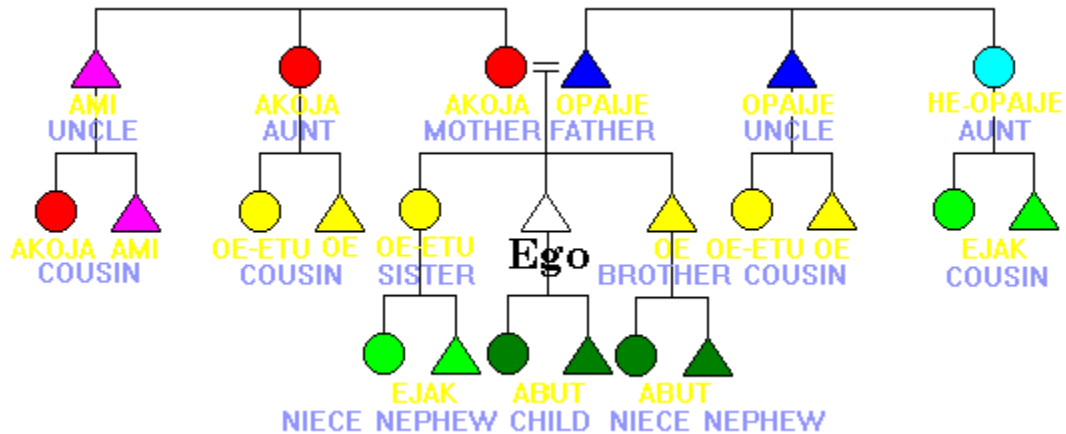
Iroquois

Yanomamo

Omaha
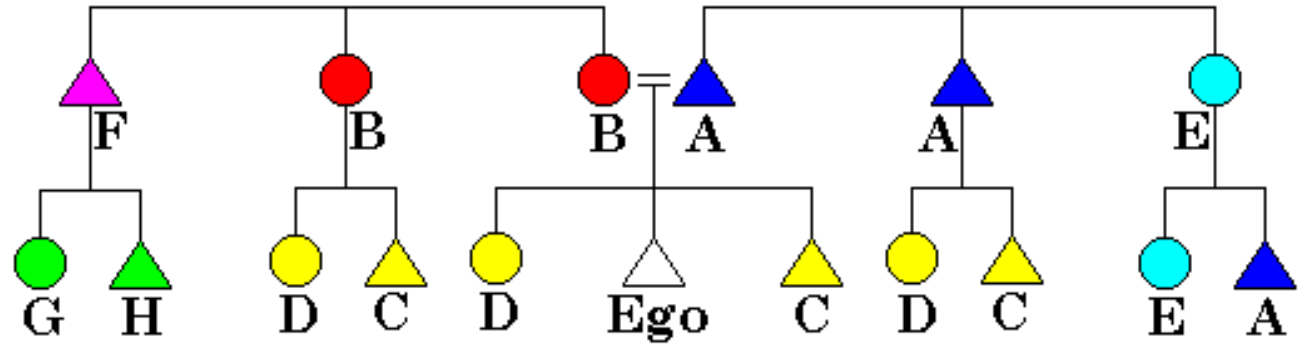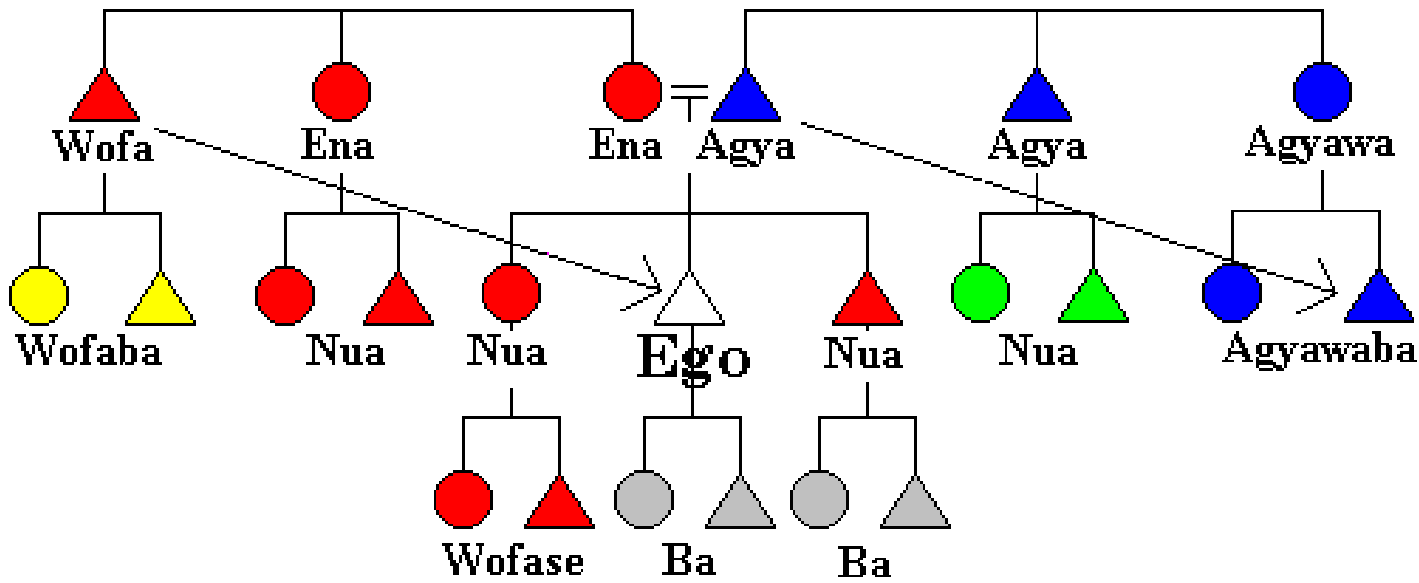
Dani / PNG

English Term	Kin Type	Dani Term
Father	F	Opaije
Uncle	FB	
	MB	Ami
Cousin	MBS	
	FBS	Oe
	MZS	
Brother	B	
Cousin	FZS	Ejak
	ZS	
Nephew	BS	Abut
Son	S	

Crow

Akan / Ghana

Franchetto & Kuikuro

Table 2. Kuikuro consanguineal kin terms (female ego)

Term	Denotata	English gloss
G+2		
ngaupügü	FF, MF	grandfather
ngitsü	MM, FM	grandmother
G+1		
uü	F, FB, FFB	father
ama, ata, isi	M, MZ, MMZ	mother
sogu	MB	maternal uncle
etsi, ipügü	FZ	paternal aunt
G0		
hisuügü	B, FBS, MZS	brother
hasü	eZ, FBeD, MZeD	older sister
ikene	yZ, FByD, MZyD	younger sister
hisü	B, FBS, MZS	brother
hãü	MBCh, FZCh	cousin
G−1		
mukugu	S, ZS	son
indisü	D, ZD	daughter
hatuü	BS	nephew
hati	BD	niece
G−2		
higü	SS, SD, DS, DD	grandson/-daughter

Table 1. Kuikuro consanguineal kin terms (male ego)

Term	Denotata	English gloss
G+2		
ngaupügü	FF, MF	grandfather
ngitsü	MM, FM	grandmother
G+1		
uü	F, FB, FFB	father
ama, ata, isi	M, MZ	mother
ijogu	MB	maternal uncle
etsi, ipügü	FZ	paternal aunt
G0		
hisuügü	B, FBS, MZS	brother
hinhano	eB, FBeS, MZeS	older brother
hisü	yB, FByS, MZyS	younger brother
ingãdzu	Z, FBD, MZD	sister
hãü	MBCh, FZCh	cousin
G−1		
mugu	S, BS	son
indisü	D, BD	daughter
hatuü	ZS	nephew
hati	ZD	niece
G−2		
higü	SS, SD, DS, DD	grandson/-daughter

* The tables make use of the commonly used abbreviations for kin relations: F = father, FF = father's father, M = mother, Z = sister, B = brother, S = son, D = daughter, Ch = child, e = elder, y = younger, etc.

BODY PARTS

the existence of alternative terms for the same body part may prove to be an interesting source of information.

Franchetto's example:

In Kuikuro, lines on the palms of hands are also called *katuga etoho* 'used for (the) *mangaba* (resin ball) to come', and the upper central region of the forehead and the thigh can be referred to with *katuga agitoho* 'used to throw (the) *mangaba* (resin ball);' both designations refer to an ancient and abandoned ritual game.

wliciting e.g. body-parts

may result in getting numerals

FROM PREHISTORY TO THE INVENTION
OF THE COMPUTER

THE UNIVERSAL HISTORY OF NUMBERS

GEORGES IFRAH
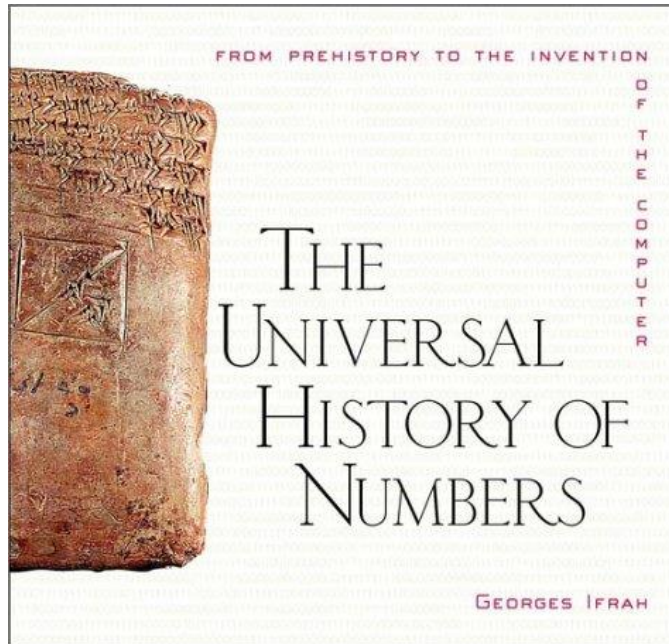
|			
1	Right hand little finger	22	left hand little finger
2	Right hand ring finger	23	Right breast
3	Right hand middle finger	24	Left breast
4	Right hand index finger	25	Right hip
5	Right thumb	26	Left hip
6	Right wrist	27	Genitals
7	Right elbow	28	Right knee
8	Right shoulder	29	Left knee
9	Right ear	30	Right ankle
10	Right eye	31	Left ankle
11	Nose	32	Right foot little toe
12	Mouth	33	Next toe
13	Left eye	34	Next toe
14	Left ear	35	Next toe
15	Left shoulder	36	Right foot big toe
16	Left elbow	37	Left foot big toe
17	Left wrist	38	Next toe
18	Left thumb	39	Next toe
19	Left hand index finger	40	Next toe
20	Left hand middle finger	41	Left foot little toe
21	Left hand ring finger		
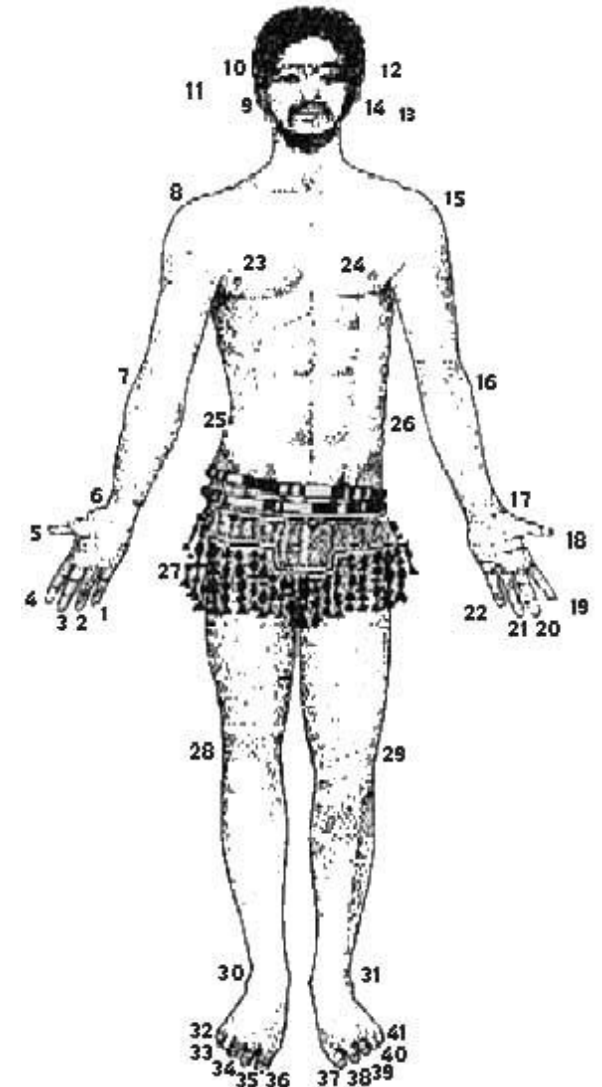
Numeral systems

TAHUTAN

1	tahi
10	onohuu
20	takau
200	au
2,000	mano
20,000	tini
20,000	tufa
2,000,000.	pohi

In counting fish, and all kinds of fruit, except breadfruit, the scale begins with **tauna** *pair* and then, omitting onohuu, they employ the same words again, but in a modified sense.

Takau becomes 10, au 100, etc.; but as the word "pair" is understood in each case, the value is the same as before.

2 (units) = 1 tauna = 2.
10 tauna = 1 takau = 20.
10 takau = 1 au = 200.
10 au = 1 mano = 2000.
10 mano = 1 tini = 20,000.
10 tini = 1 tufa = 200,000.
10 tufa = 1 pohi = 2,000,000.

endangered Tahutan numerals

MATERIAL CULTURE or ARTEFACTS
TERMS RELATING TO SUBSISTENCE ACTIVITIES

The terminology relating to the building and structure of the traditional house may prove relevant, for example, if some of its parts are named after human body parts, as well as being useful from a comparative perspective.

Terms: plants, verbs denoting actions and events in the agricultural economy, often extractable from origin myths on cultivated plants, the lexicon relating to types of swidden agriculture and phases of cultivation enables the extraction of important data on the organization of agricultural work as well as comparative observations.

SHAMANISM / RITUALS

Kuikuro lexicon

a key thematic field for many societies

Ideally, the lexicon would include all the terms designating supernatural beings or entities, explicating them individually and as a whole, and associating them with etiology, the classification and denomination of illnesses, cures, rituals, masks

```
\lx     itseke
\entyp  root
\lc     itseke
\ph     [iˈt͡seke]
\ps     N
\ge     hyper-being
\xkk    tinegetinhüha ugei itsekeinha
\te     I am afraid of the hyper-beings
\xkk    itseke ingilüha kupehe kukapüngu igakaho
\te     we see the hyper-beings before we die
\xkk    kagamuke kaginenügü itseke heke
\te     the hyper-being frightened the child
\defkk  itseke ekisei kukengeni, kugehüngüha ekisei, inhalüha
        ingilüi; itseke kukilüha ngiko heke kukengeni heke;
        itseke ekisei kukotombani kukügünuhata.
\defe   Itseke is that which eats us, it is not a person, it
        cannot be seen; we say that itseke is something which
        eats us; itseke is that which hurts (otomba-) us with
        invisible arrows when we are sick. Itseke is a super-
        natural being, a spirit, a 'beast;' it dwells in the
        forests, rivers and lakes; it causes illness and
        death; only shamans and the sick can see them.
\cf     kuegü, otomba
```

HOW / WHERE TO DOCUMENT (THIS KIND OF) LEXICAL KNOWLEDGE ?

from a previous lecture

Metalinguistic knowledge,

manifest in the native speakers' ability to provide interpretation and systematizations for linguistc units and events

(Himmelmann 1998)

Table 2. Extended format for a language documentation

Primary data	Apparatus	
	Per session	For documentation as a whole
recordings/records of observable linguistic behavior and metalinguistic knowledge	*Metadata* *Annotations* – transcription – translation – further linguistic and ethnographic glossing and commentary	*Metadata* *General access resources* – introduction – orthographical conventions – glossing conventions – indices – links to other resources … *Descriptive analysis* – ethnography – descriptive grammar – dictionary

Woodbury, Anthony C. 2003.
"Defining documentary linguistics",
in: P.K. Austin (ed.) *Language Documentation and Description* 1: 35-51.

Bridging «linguistic behavior» and «metalinguistic knowledge»

«[…] for example, if you are making a thesaurus, you don't want to just find the names of different grasses in your corpus, you also want the resultant list to be discussed and gone over by speakers who are authorities on grasses to make sure you have the field properly covered and to generate good definitions. Rather, what a documentation-oriented view says is that **the discussions of grass names** should themselves be videotaped or tape recorded and **should themselves become a part of the whole corpus**; as should any and all grammatical elicitation of the traditional kind. Moreover, years from now, it will be the grass name attestations and grasses discussion tapes, and not the dictionary, that you will consider as the final document on grass names.»

PRINCIPLES OF LANGUAGE DOCUMENTATION

☐ done in **interdisciplinary teams**: linguists + other specialists,

☐ outsiders and insiders: speech community is involved not only as

„informants"

☐ done for **various purposes**

☐ published electronically (and on paper)

☐ stored in digital archives

☐ contains written, audio and video material

☐ documents the use of the language in various situations

SlideShare

BRIDGING PRIMARY DATA, ANNOTATION AND ANALYSIS

Gaby Cablitz. 2011. The making of a multimedia encyclopaedic dictionary for and in endangered speech communities. In: *Documenting endangered languages*, eds. G. Haig et al., 223-261.

Software: LEXUS and ViCoS (MPI Nijmegen, not longer supported)

IMPORTANCE AND USE OF DICTIONARIES

Since dictionaries are more visible or salient to the public than are grammars or text collections, community members are more likely to have strong opinions about dictionaries.

Speakers may object to dictionaries for religious, puristic, or political reasons.

Other speakers will consider the dictionary (particularly a handsome paper copy to be displayed on a bookshelf) to be an emblematic part of their linguistic heritage.

Extra care and diplomacy in securing community support and collaboration is crucial.

(Chelliah, Shobhana L. & Willem J. Reuse 2011. *Handbook of descriptive linguistic fieldwork*. Chapter 9: "Lexicography in Fieldwork")