

## Introduction to the online Workshop on the Terraling Database

Hilda Koopman, UCLA, September 18 2020.

- Welcome!
  - What is Terraling?  
The Terraling database <http://terraling.com/>:
    - (flexible) collaborative research tool for theoretically guided cross-linguistic research
    - It is an open-access, open-ended, community based,
    - where linguists can both pose data questions to others and answer such questions for their native languages.
- For further information see [here](#)).
- Goal of workshop:
    - get you interested in the Terraling database [terraling.com](http://terraling.com), and collaborating,
    - familiarize you with the technical aspects
    - give you an idea what the database can do
    - so that researchers can carry out their own projects
    - Generally pave the road for future projects making use of the database, and generally stimulate comparative (and collaborative) research.
  - Program and structure of the days  
Day 1: more practical/technical.  
Day 2: more content oriented.

**Program** (all times are given in Central European Summer Time, which will mean, e.g., -6 hours for New York City/East Coast, -9 hours for Los Angeles/West Coast (A few) Times on the West Coast (PT) are indicated in red). Day 1. Friday September 18- 2020 (All times CET, in Pacific Time in red)

5.30 pm Zoom opens. Welcome! Meet and Greet: Virtual coffee/ or virtual happy hour.

8.45 AM PT

6 pm 6.20 pm: Introduction to the database (Hilda Koopman, UCLA)

PT 9 am-9.20

6.20 pm 6.40 pm: A guided tour through the database (Nina Haslinger, Göttingen)

6.40 pm 7.10 pm: Try-out session (practice/all participants, using break-out rooms).

7.10 pm 7.30 pm: Discussion (all participants) (Viola (moderator) and Nina)

7.30 pm 7.40 pm: Break

7.40 pm 8.25 pm: Terraling and Fieldwork (Ryan Bochnak, UBC)

10.40- 11.25 PT

8.25 pm 9 pm: Networking discussion (all participants) (Hilda (mod), and Nina)

## 0.1 (Very) Brief history of the project

- Chris Collins and Richie Kayne (2009) NYU. Workshop on feasibility of building on online database. Dennis Shasha (NYU), attended, came up with the plan how to do this, and developed the prototype database.<sup>1</sup>
- pilot SSWL
- HK took over from Chris Collins (2011).  
Based on lessons from working with the pilot– reprogrammed the backend, and broadened the vision.

Terraling: a webbased database application that allows setting up your own datasets online, using all the tools and infrastructure of the database.  
Current datasets (SSWL, Coordination and Conjunction, Cinque's Universal 20 database,...).

## 0.2 Why yet another database?

- Progress in theoretical linguistics: native speaker intuitions, introspection.  
...which has produced descriptive grammars volumes of entirely new knowledge, with the end nowhere in sight (see Kayne 2012).
- No crosslinguistic database that allows us to "imitate" where Theoretical progress has come from. Native speaker linguists intuitions on their native languages.

So we have to build one!

- Terraling allows accessing native speaker intuitions/introspection. "Imitate fieldwork, controlling context, etc".  
Can provide information about what is not possible.

## 0.3 What differentiates Terraling from existing databases, and how does it work?

Set up to support theoretical research in linguistics.

Can do what existing typological databases can do but can do more:  
+ Possibility to involve native-speaker linguists: allows access to intuitions/introspection  
+ Possibility to find out whether some things are possible or not, whether certain meanings for a given example are available in some scenarios.

Provides a flexible framework that researchers can use to set up their own projects.

(Built-in) Tools to explore the database.

<sup>1</sup>current development team: Dennis Shasha NYU (architect), Hannan Butt, and Shailesh Vasandani (backend)

## 0.4 Developing Content: Basic Database schema.

1. the database schema and the notion of property.
2. show an example, and compare with Wals
3. Explain one more powerful feature ( language level properties, and form level properties) that the database set up allows, and why that is particularly useful .
4. briefly introduce "roles" in the database

### 0.4.1 Illustrate the database schema and the notion of property.

"Content" is generated: Flexible database schema (relational database)

"Language", "Property": values Y/N (as well as NA))

Property: a characteristic that holds of a "language".

A question language experts are asked to answer for their native language.

A simple example: Word order property (from SSWL) SV= fall

	SV	VS	
1	Yes	No	English,...
2	No	Yes	Samoan,...
3	Yes	Yes	Italian...
4	No	No	0

#### Better empirical content:

Even where properties look similar (word order properties in SSWL and WALS), the classifications turn out to be quite different. In SSWL, they can become much more finegrained, without the need for reprogramming and reflect the empirical landscape more accurately

### 0.4.2 A concrete example (from SSWL) of a Terraling data schema

- Capturing word order properties for A(djective) N(oun) orders
- ..starting with the most general to gradually more refined properties.
- .. compare results from SSWL with WALS (see Dryer (2013)). (See.Koopman & Guardiano. (2021) for discussion.)
- Adjectives can either follow the noun, precede the noun, or do either. This can be captured by two (independent) properties, and shown in the table below: (This table shows the result of a a "cross" search in SSWL, one of the many searches that are possible)

SSWL		
Adj N	N Adj	Count?
Yes	No	59 ...Bangla
No	Yes	100 ..Swahili
Yes	Yes	87 ... <span style="border: 1px solid black; padding: 0 2px;">French</span> , Tagalog
No	No	3 .. No As

- WALS, has one feature, "Order of Adj N", which has four values: "N Adj", "Adj N", "no dominant order" and "only internally headed relative clause". (no possibility to build on this schema to make the classification more precise)

Table 1: A Comparison of Adj N orders in WALS and SSWL

WALS		SSWL		
Order of Adj		Adj N	N Adj	Count?
Adj N	.. Bengali..	Yes	No	59 ...Bengali
N Adj	.. <span style="border: 1px solid black; padding: 1px;">French</span> , Swahili,..	No	Yes	100 ..Swahili
No dom. order	.. Tagalog ..	Yes	Yes	87 ... <span style="border: 1px solid black; padding: 1px;">French</span> , Tagalog
Intern headed	...	No	No	3 .. No As

- Problems:

- in WALS : French and Swahili belong to the same type (wrong). (because of the (undefinable) notion of dominant order)
- in SSWL; they belong to different types (correctly: this is what we observe. YSWYG)

**No loss of information! No dominant order**

- but in SSWL, French and Tagalog belong to the same type, raising further questions.
- This is the best we can do with these properties!
- Can we do better using the basic database schema and explore further questions?
- Does the semantic category matter?

Well-known: Adjectival hierarchy R & Shih (1991), Dixon & (eds), Cinque (1994, 2010)?

- Formulate new properties *not in the SSWL database yet*

- (1)
  - a. "Adj<sub>color</sub> N: Yes/No",
  - b. "N Adj<sub>color</sub>: Yes/No"
- (2)
  - a. "Adj<sub>size</sub> N: Yes/No"
  - b. "N Adj<sub>size</sub>: Yes/No".

- Yields a finer typology:

	Adj <sub>size</sub> N	N Adj <sub>size</sub>	Adj <sub>color</sub> N	N Adj <sub>color</sub>	Count?
1	No	Yes	No	Yes	..Swahili
2	Yes	No	No	Yes	..French
3	Yes	Yes	No	Yes	..Italian
4	Yes	Yes	Yes	Yes	Tagalog

**Better empirical content:**

Allows building much more finegrained typologies, thus reflecting the empirical landscape more accurately.

Even where things look similar (word order properties in SSWL and WALS), the classifications turn out to be quite different, and the possibility to expand the dataset over time is vastly different.

This can even be further refined it using the double depth feature: Language and Forms.

### 0.4.3 Basic set up of a group allows one level-properties or two level properties

- One level: ("depth 0")
  - (3) "Language" - property: values are linked to "Language".
  - (4) For Example: 160 properties and 400 languages  
These will have to be answered for each Language.  
An expert will be asked to set the values for these for each Language
  
- Two levels: "depth 0" and "depth 1".
  - "Parents", and "children"  
"Languages" and "Forms"
  - (5) Just as "Children" are linked to "Parents", Forms are linked to Languages.
    - a. **Properties of Language:** Language-level properties
    - b. **Properties of Forms:** Form-level properties.
  
- An example of Form-level properties:
  - Suppose you want to find out properties of (special) existential construction in a language.
  - And a language has 2 (or more) such constructions.
  - You have decided you have 7 properties you want to find out about existential constructions. We can make these properties form-level properties:

Languages	Forms
...	Form1=be existentials Form2= have existentials Form3= give existential
	properties "coda obligatory?: yes/no" properties "pivot definite?: yes/no" etc ...

- Making the properties form-level properties, will allow coding the properties of each Form (existential construction)

This will show exactly how each existential construction behaves w.r.t these properties, and compare where they are similar and where they differ.

Before turning over the floor to Nina: for the guided tour, let me introduce one last feature:

#### 0.4.4 Roles in the database

- Roles in the database: who does what, and what can each role do:
  - Administrators (administration, set up a group, add members, vet members (make member into "expert" of a specific language), create languages, create properties, change language data, delete etc...)
  - Property authors/Teams: develop content (properties, values. and queries/property definitions
  - Experts (Contributors) answer the queries for their language, or a language they have deep knowledge of, set values and provide examples (with comments where relevant) that illustrate the values ("populate the database").  
Experts have sole control over their language, and can change data, add data, change values, add comments at any time

## References

- Cinque, Guglielmo. 1994. On the evidence for partial N-movement in the Romance DP. In L. Rizzi R. Zanuttini J. Koster, J.-Y. Pollock (ed.), *Paths towards Universal Grammar*, 85–110. Georgetown University Press Washington, DC.
- Cinque, Guglielmo. 2010. *The Syntax of Adjectives: A Comparative Study*. MIT Press. <https://doi.org/10.7551/mitpress/9780262014168.001.0001>.
- Dixon, Robert MW & Alexandra Y Aikhenvald (eds) (eds.). 2004. *Adjective classes: A cross-linguistic typology*, vol. 1. Oxford University Press.
- Dryer, Matthew S. 2013. Order of Adjective and Noun. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/87>, [Accessed on 2019-05-01].
- Koopman, Hilda & Cristina Guardiano. 2021. Managing data in TerraLing, a large-scale cross-linguistic database of morphological, syntactic, and semantic patterns. In Bradley McDonnell Eve Koller Berez-Kroeker, Andrea & Lauren Collister (eds.), *The Open Handbook of Linguistic Management*, MIT Press.
- R, Sproat. & C. Shih. 1991. The Cross-Linguistics Distribution of Adjectival Ordering Restrictions. In C. Georgopoulos & R. Ishihara (eds.), *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda*, Dordrecht: Kluwer.