

# A “Social Layer” in Typological Databases

Michael Cysouw

# Background: Sharing Information

- Database publication
- Database integration
- Accountability
- Differences of opinion about data
- Citability

# Two-layered Databases

- Social Layer (“notes”)
  - ▶ shared resource-based annotations
  - ▶ discussion about these annotations
- Personal Layer (“database”)
  - ▶ classifications in restricted vocabulary
  - ▶ no discussion about (personal) vocabulary

# Annotations in the Social Layer

- Include (internal structure of the note):
  - ▶ Reference to relevant pages/sections/lines
  - ▶ Interpretations of a source
  - ▶ New glossings of examples given in the source
  - ▶ Language specific form
- Linked to a resource
- Unchangeable after publication of the note
- Citable for use in database or discussion

# Drehu

## Language Feature Information:

feature:

[Alignment of Case Marking of Full Noun Phrases](#)

value:

[Active-inactive](#)

Data Sources:

- Moyses-Faurie 1983: 147
- [Show Example 1](#)

## References:

- Moyses-Faurie, Claire. 1983. Le drehu, langue de Lifou (Iles Loyauté). Paris: SELAF.
- Nichols, Johanna. 1992. Linguistic Diversity in Space and Time. London and Chicago: University of Chicago Press.
- Tryon, Darrell T. 1967a. Dehu Grammar. Pacific Linguistics, Series B-7. Canberra: ANU.
- Tryon, Darrell T. 1967c. Dehu-English Dictionary. Canberra: The Australian National University.
- Tryon, Darrell T. 1967d. English-Dehu Dictionary. Canberra: The Australian National University.

## 89 available features for Drehu:

### Phonology

[1. Consonant Inventories](#) - Moderately large

## Drehu

troa traqa la nekönatr  
 OBLIG arrive the child  
 'the child must come' p. 147

kola huliwa hnei wamo  
 DUR work AG Wamo  
 'Wamo is working' p. 147

öhnyi angajoxu hē hnei sinewenyi  
 find big.chief ACC AG Sinewenyi  
 'Sinewenyi has found the big chief' p. 147

source: Moyses-Faurie 1983

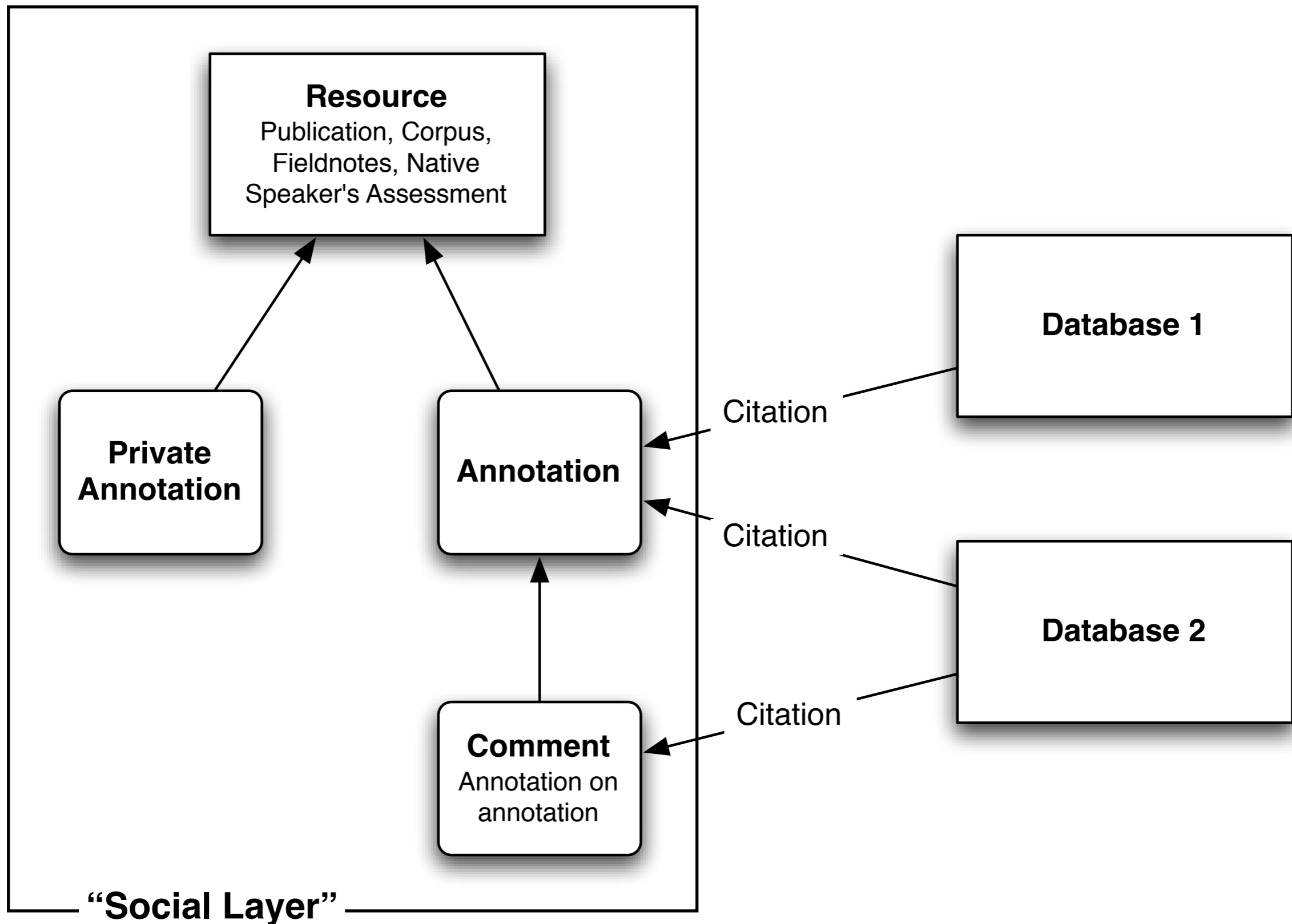
note: In the non-Past, illustrated here, the agent marker is used obligatory with transitive subjects and optionally with agentive intransitive subjects; in the Past, it is used with all subjects (though with some exceptions for inanimates). Pronouns apparently follow the same pattern, though few possibilities are illustrated.

Language Profile

arrange the languages by

languages ▾

COPY LIST



# Citation & Audit Trail

- every entry in a database can easily be made citable (“only” controlled vocabulary)
- database entries **cite** annotations
- annotations **cite** sources
- **sources determine language !**

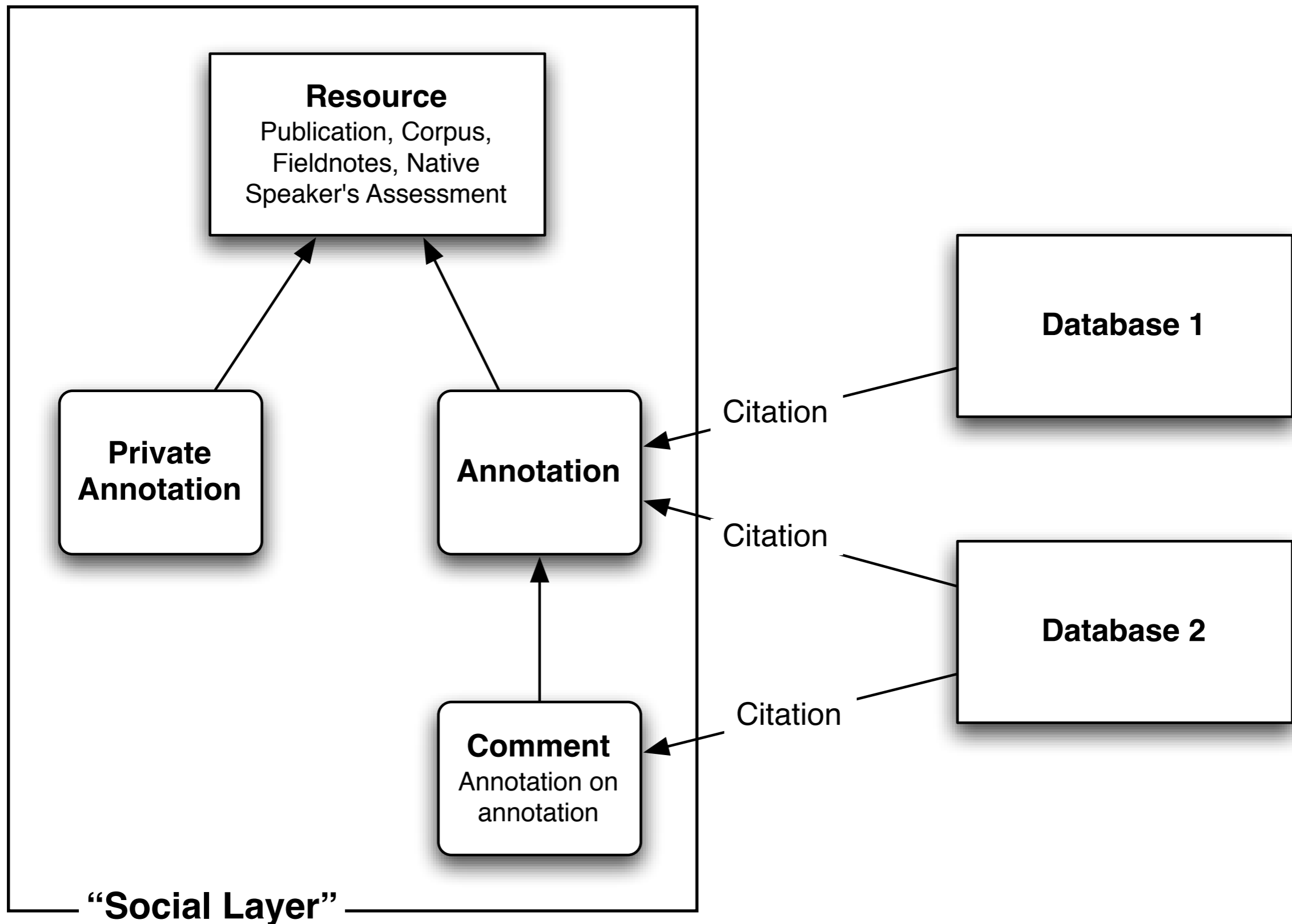
# Notes are easily searchable

- Notes contain **humanly readable text**, which can be searched (e.g. text mining)
- Notes are **classified** by the attributes (“features”) that cite them
- Attributes can be **organized** by the notes that are cited



# Sharing

- Notes are **shared** for use by other databases
- Different interpretations of the **sources** can be discussed in the social layer
- Different interpretations of the **attribute** should result in a new database



# Sources determine Language ...

- Old problem: what is a Language ?
  - ▶ Answer: depends on the granularity of view
- Proposal: **Languoid** for all “Language-like” entities
  - ▶ term combining **lects** (down) and **families/areas/typological groups/ad-hoc groups** (up)
- Lowest level Languoid is the **Doculect**
  - ▶ Many kinds of resources possible (published, corpora, native speaker’s assessment, etc.)
- Databases refer to doculects (via notes)
  - ▶ Typology depends on the grouping of doculects.