

02 Basics of Spoken Language Documentation

Dafydd Gibbon

Universität Bielefeld

2015-11-11

Abidjan-Bielefeld Cooperation

Programme for today

- Some revision
- Documentation workflow
- Documentation principles
 - Requirements specification
 - Design
 - Implementation
 - Evaluation
- Document creation
- Introduction to annotation
 - Elan - multimedia & archiving
 - Praat - phonetics and phonology
- Practical work
 - discussion of report
 - creation and addition of graphics

The problem to be solved

How is professional documentation created?

What is professional documentation?

- What is its structure?
- What is its physical form?

Meetings:

1. Basics: text linguistics and document creation
- 2. Basics: Documenting spoken language**
3. Recording and annotation
4. Language documentation report

Language documents and their creation

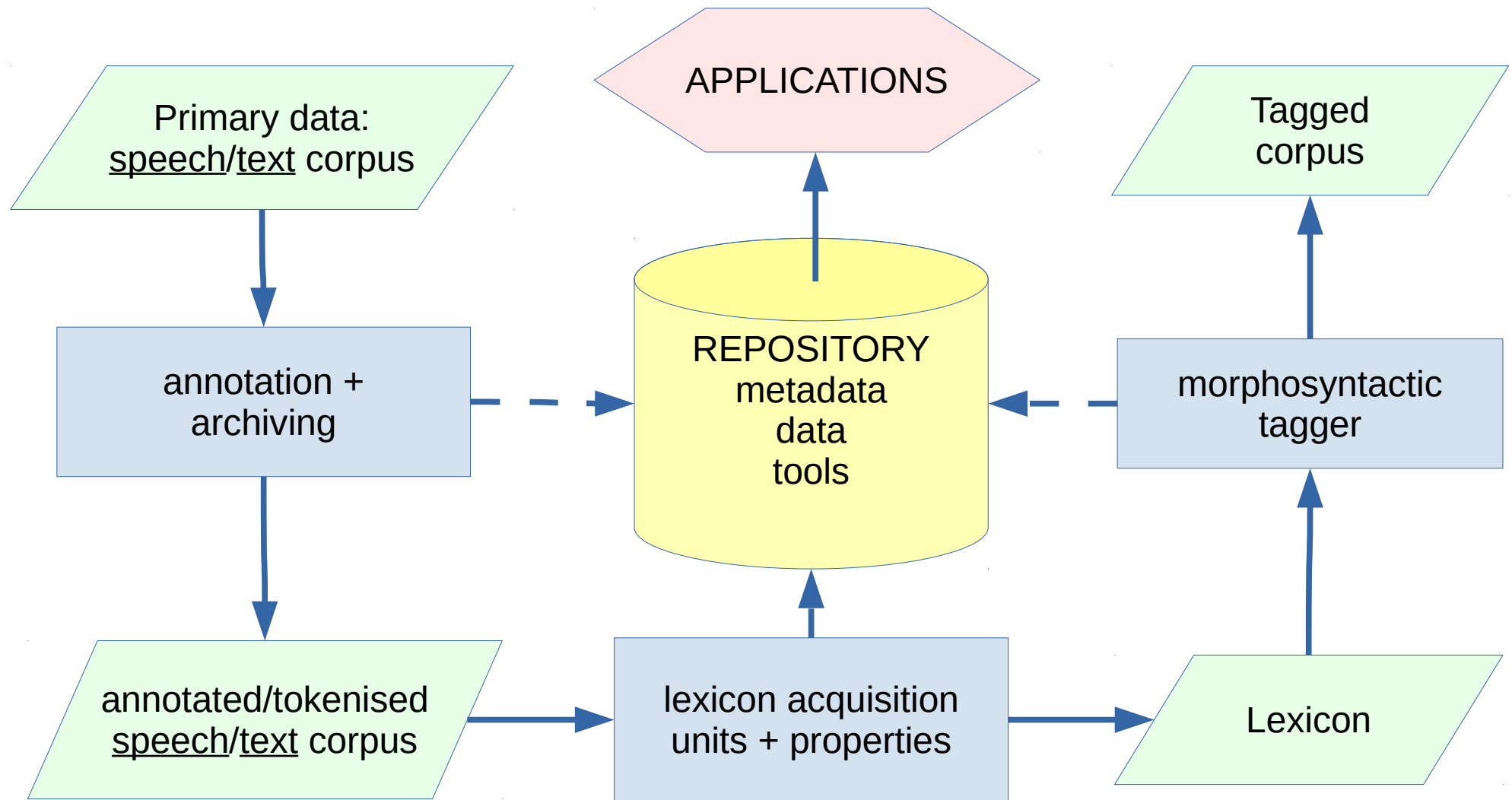
Language documentation is concerned with multimodal documents (artefacts such as texts, audio and video recordings, databases) and their creation with manual and computational methods.

The multimodal documents are designed to have standardised properties pertaining to

- reusability (for purposes other than the original purpose)
- sustainability (for survival over long periods of time)
- interoperability (for use in practical environments)
- ubiquity (for general accessibility)
- coherence (for organising documents and databases)

These properties are supported by using a text linguistic frame of reference.

Language Documentation: simplified workflow

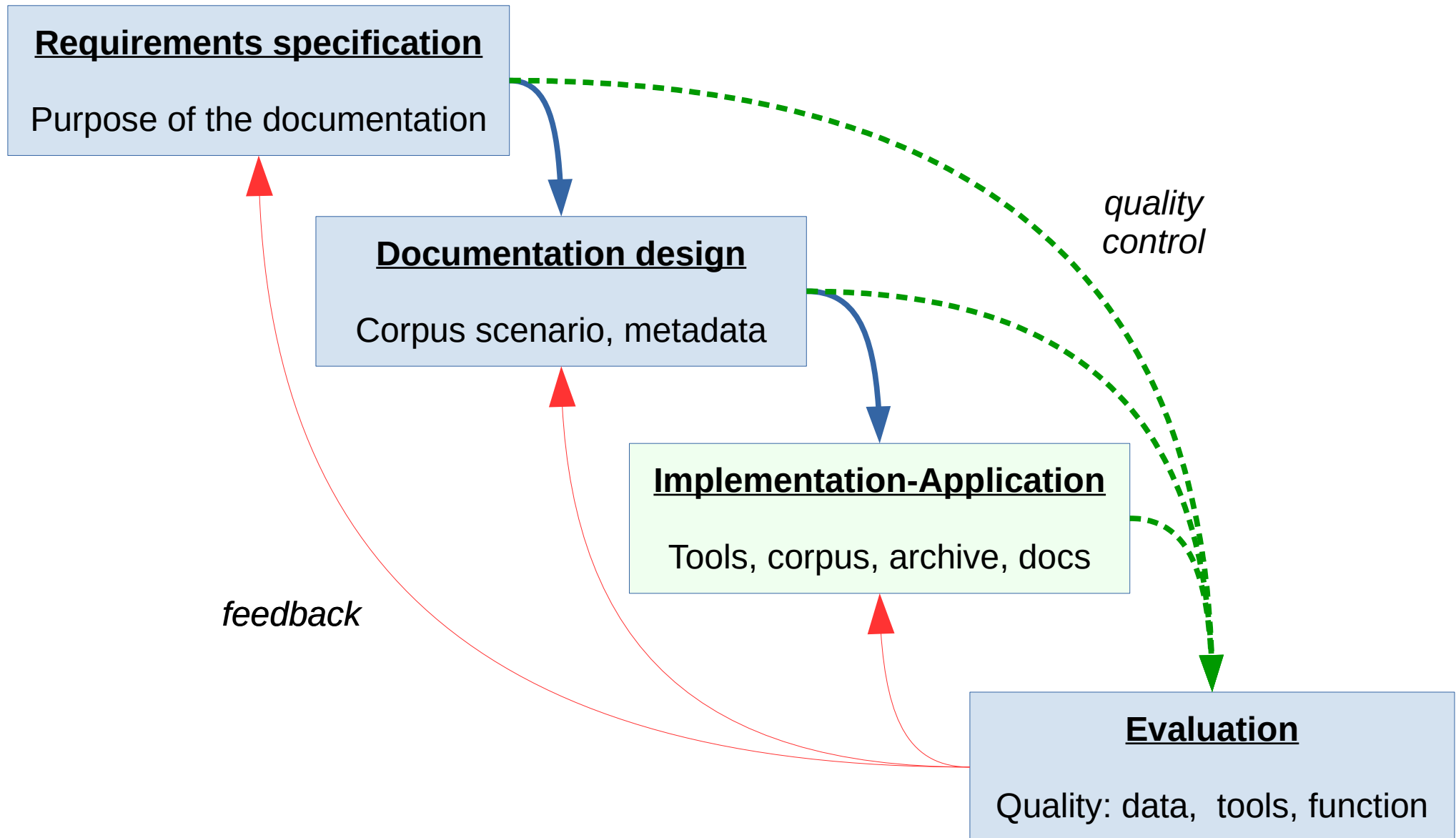


What is language documentation?

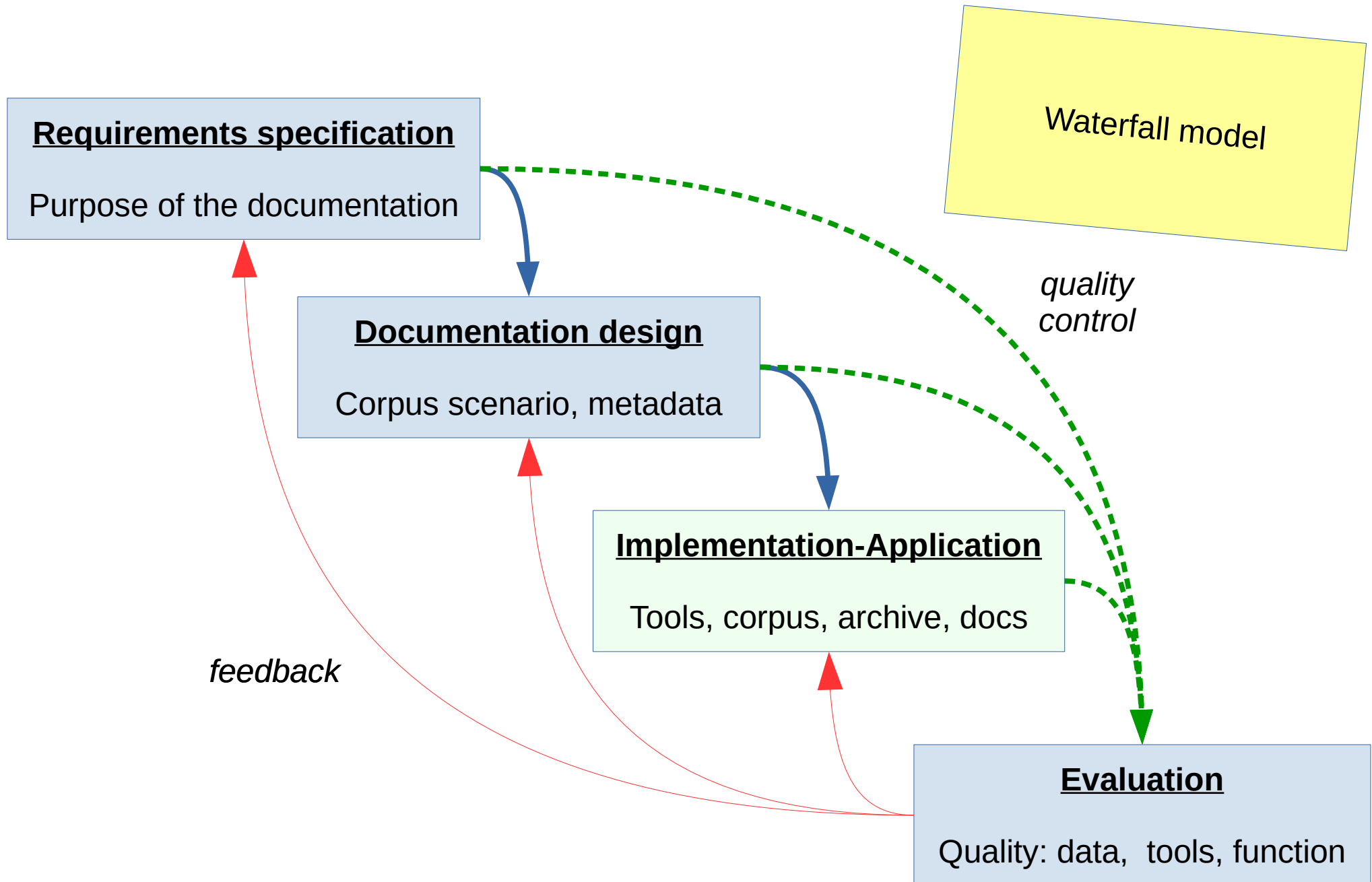
Language documentation is a manual and computational methodology and a product, with a broader linguistic context:

- Discovery methods:
 - documentation: recording of observed data
 - description: generalisation over documentation
 - explanation: predictive theories of descriptions
- Evaluation methods:
 - testing of explanation, description, documentation
 - extension of results to new data
- Application methods (manual and computational):
 - Storage media: databases for
 - standardisation, sustainability, reusability, interoperability
 - Production of objects with tools:
 - books and digital media/texts, dictionaries, grammars

General principles: documentation project development



General principles: documentation project development



Requirements specification (pre-collection phase 1)

- Purpose
 - heritage materials
 - lexicon
 - grammar
- Users
 - community
 - public
 - specialists
- Use cases
 - museum
 - teaching
 - research
- Creators
 - professionals
 - researchers
 - students
 - community

Documentation design (pre-collection phase 2)

- Components:
 - scenario
 - participants
 - media
 - tools
 - task network:
 - Pert chart
 - schedule (timeline):
 - Gantt chart
- Documentation:
 - metadata
 - IMDI
 - OLAC
 - archive:
 - database
 - materials
 - description
 - transcription
 - annotation

Implementation (collection phase)

- Training
- Tools
 - recording device
 - video
 - audio
 - sensors
 - microphones, headphones
 - media
- Instructions
 - prompts
 - wordlists ...
- Finance
- Travel
- Permissions
- Data collection:
 - Instruction
 - general
 - experimental
 - Corpus
 - recording
 - archiving
 - documentation

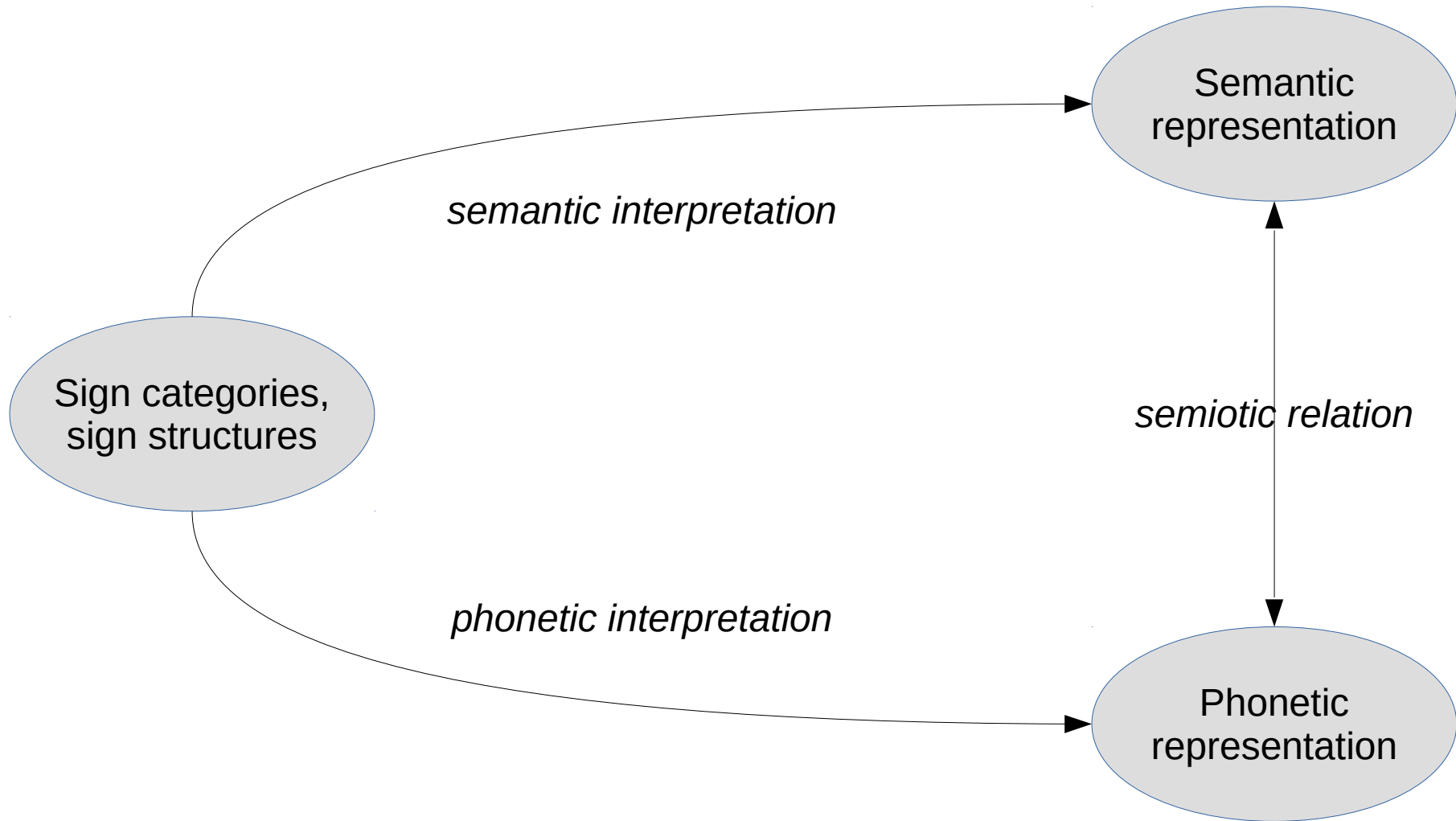
Evaluation

- Internal evaluation:
 - requirements ok?
 - design ok?
 - implementation ok?
 - evaluation procedure ok?
- External evaluation:
 - project proposal ok?
 - description?
 - tasks?
 - timeline?
 - personnel?
 - finance?
 - results ok?
 - corpus?
 - tools?
 - documentation?
 - archive?

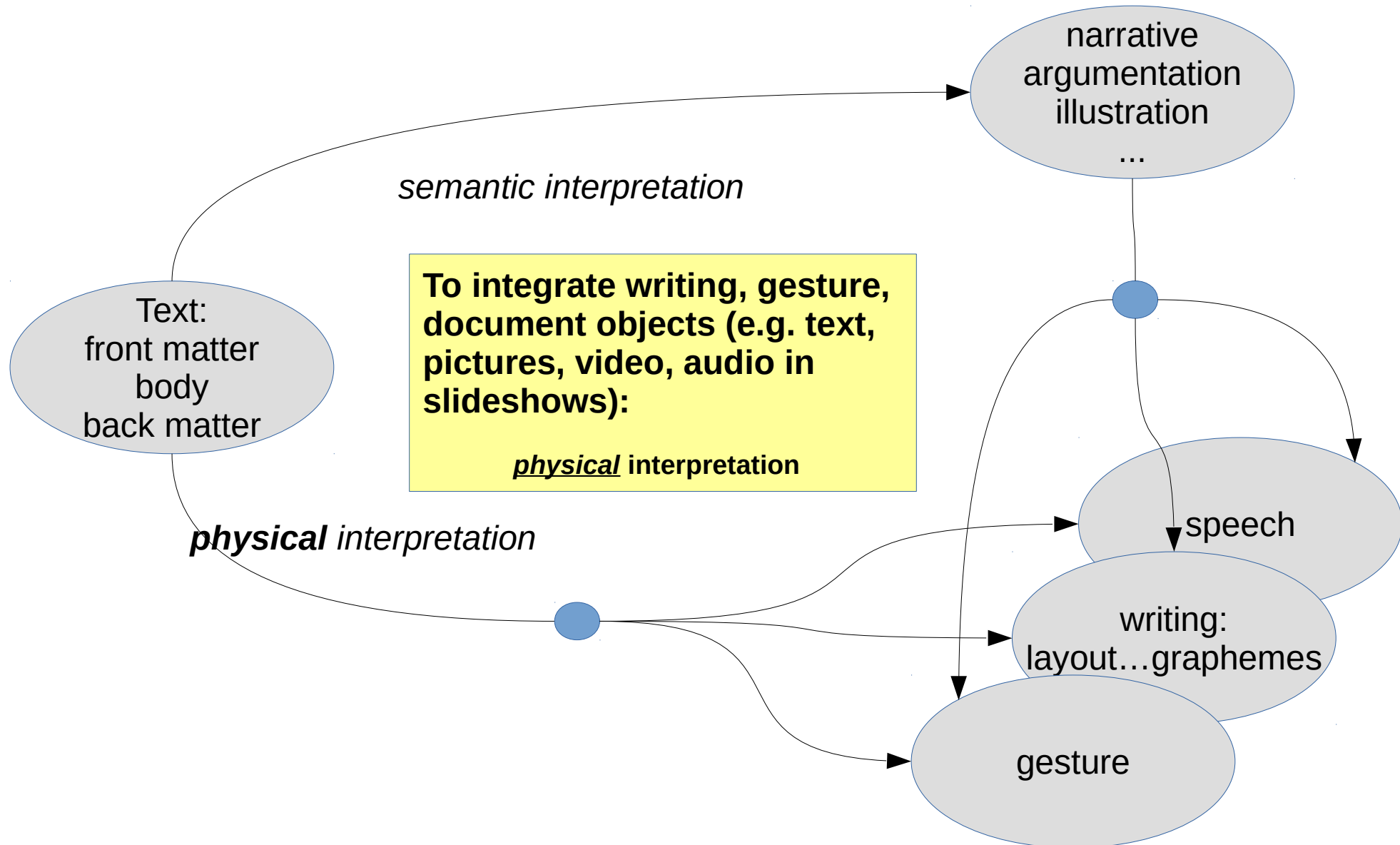
Evaluation

- Internal evaluation:
 - requirements ok?
 - design ok?
 - implementation ok?
 - evaluation ok?
 - External evaluation:
 - project proposal ok?
 - results ok?
 - tools?
 - documentation?
 - archive?
- The multimodal documents are designed to have standardised properties pertaining to
- reusability (for purposes other than the original purpose)
 - sustainability (for survival over long periods of time)
 - interoperability (for use in practical environments)
 - ubiquity (for general accessibility)
 - coherence (for organising documents and databases)

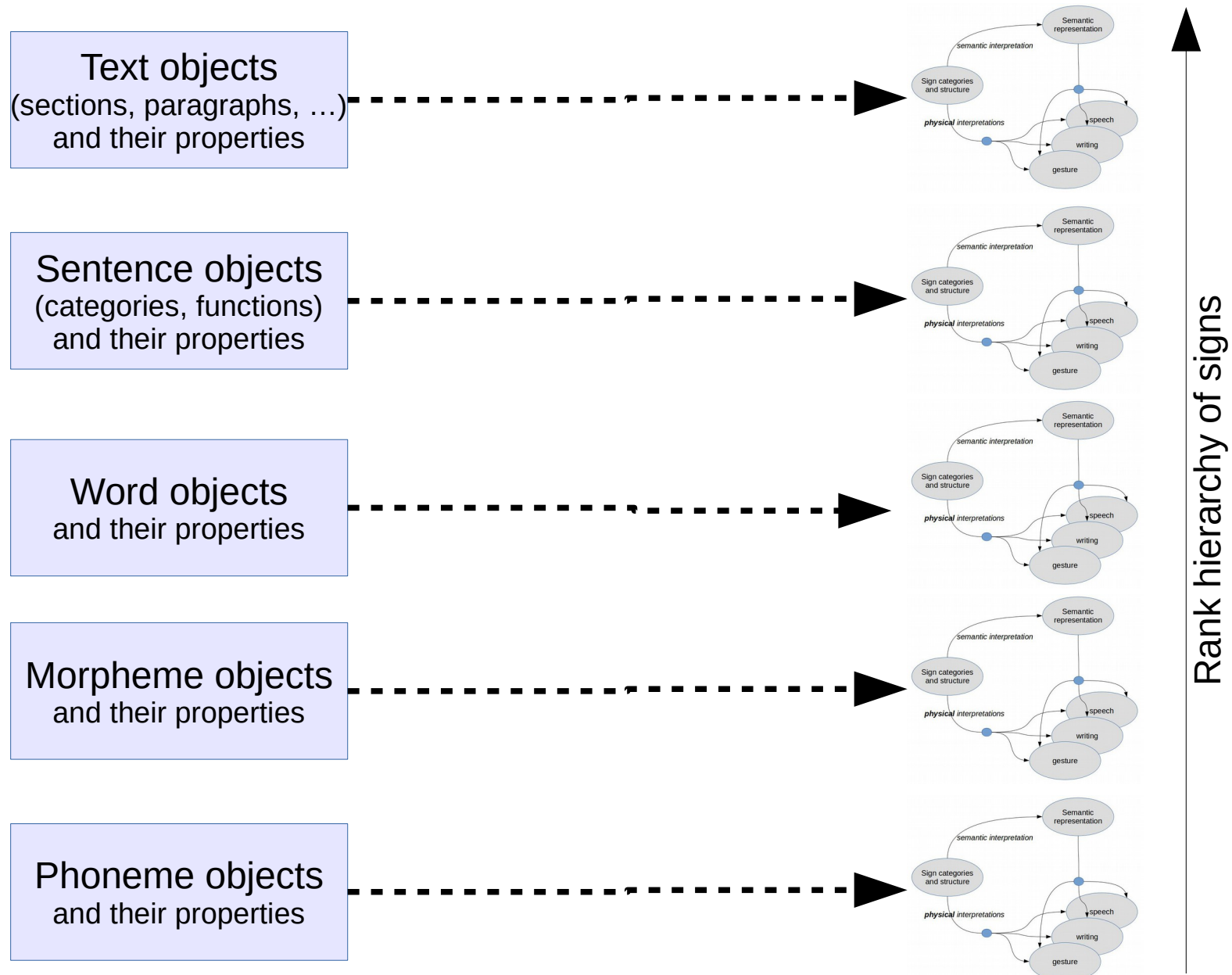
Documentation: a general sign model



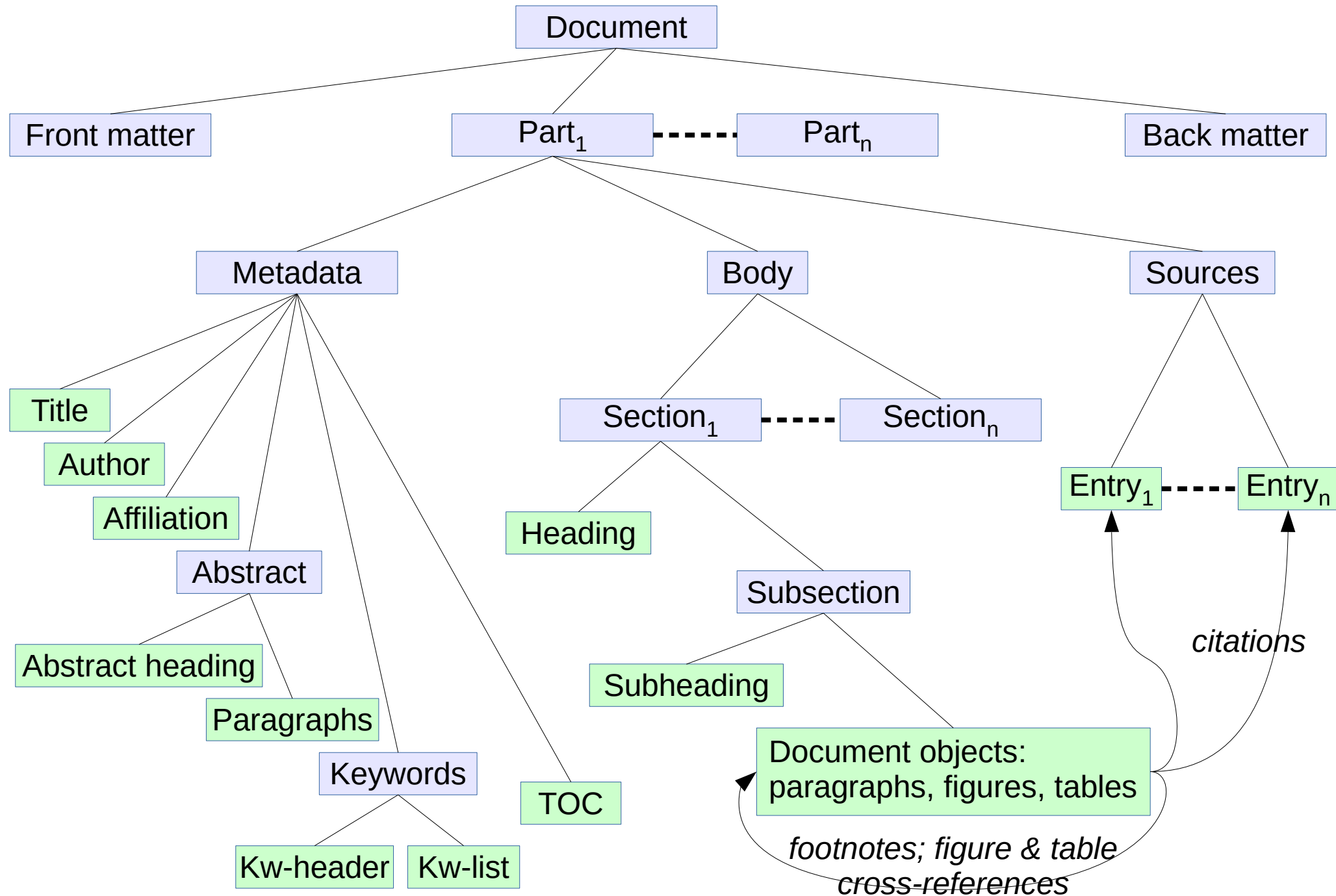
Documentation: a general sign model



Text linguistics: a general sign model



Structure of a typical report document or article



Time for practical work ...