Can there be standards for spontaneous speech?

Towards an ontology for speech resource exploitation

Dafydd Gibbon

Universität Bielefeld
Germany

LPSS - Taipei - 2006-11-17/19

The context - progress on resources?

- Previous work on standardising resources:
  - EAGLES hanbooks:

    Gibbon, Moore & Winski 1997

    Gibbon, Mertins & Moore 2000

- Things have come a long way since then:
  - SpontiSpeech
  - Multimodality
  - new language families with speech resources
  - new archiving and dissemination techniques and standards (XML, Unicode; OLAC; ELRA/ELDA; LDC; ...

MULTIMODALITY

OTHER LANGUAGES

Example: multimodal counting

Multimodal counting

What is so complicated about this?

What is so complicated about this?

- Modality:
  - vocal-acoustic (vs. clapping, stamping, snapping...)
  - manual-visual (vs. facial, gaze, posture, ...)

What is so complicated about this?

- Modality:
  - vocal-acoustic (vs. clapping, stamping, snapping...)
  - manual-visual (vs. facial, gaze, posture, ...)
- Structure:
  - coordination of parallel events
  - analysis of number morphology:
    - base 10? 5? 20? 12?

What is so complicated about this?

- Modality:
  - vocal-acoustic (vs. clapping, stamping, snapping...)
  - manual-visual (vs. facial, gaze, posture, ...)
- Structure:
  - coordination of parallel events
  - analysis of number morphology:
    - base 10? 5? 20? 12?
- Semantics:
  - set-denoting gesture (fingers of left hand)
  - deictic gesture (finger of right hand)

RELEVANT ISSUES

AT

LPSS CONFERENCE

LPSS topics - overview

- Issues of
  - spontaneous speech domains
  - spontaneous speech varieties
  - categories in spontaneous speech
    - linguistic
    - phonetic
  - methods of
    - corpus creation
    - analysis
  - technology
    - automatic speech recognition

LPSS topics: domains of speech

- Genres:
  - affected speech
  - task based communication
  - remembering
- Varieties
  - language variation
  - speech acquisition
  - foreign language learning
  - code-switching

- Phonetic and linguistic categories:
  - dialogue acts
  - floor-holding
  - silences, pauses
  - disfluencies, repairs
  - duration
  - intonation, prosody and understanding
  - prosodic hierarchy
  - voice quality

  - *Where is multimodality?*

LPSS topics: methods, applications

- Interaction in corpus creation:
    - map task
    - ethnographic observation; conversation
    - Questionnaire-based interview
  - Wizard of Oz; Monologues
- Analysis techniques
  - Tools for transcription and alignment
  - POS tagging
  - Machine Learning
- Speech technology
  - Automatic Speech Recognition
    - *Where is speech synthesis?*
    - *Can there be "spontaneous" speech synthesis?!*

SPONTANEITY - AND STANDARDS??

NEEDS AND STRATEGIES

Spontaneity

LinguistList survey (Fagyal 1995):

- 'Spontaneous speech' is a
  (1) type or 'mode' of speech production opposed to 'read-aloud' speech;
  (2) real-time generated, unplanned and non-rehearsed type of encoding linguistic information;
  (3) casual 'way of speaking' or 'style', characterizing informal speech situations;
  (4) naturally occurring, non-experimental type of speech event of any kind.

Cf. the survey of LPSS topics ...

Spontaneity

- Better, maybe:
  - Authentic speech
    - (as in Foreign Language Teaching)
  - Definition:

Authentic speech is speech which is not produced for the purpose of the study of speech.

The need and a strategy

- How can all the LPSS topics (and others) be systematically organised
  - in order to facilitate
    - coordination of resources
    - facilitation of interdisciplinary and international work on resources
  - and in particular
    - organisation and storage of resources
    - search for resources

The need and a strategy

- How can all the LPSS topics (and others) be systematically organised
  - in order to facilitate
    - coordination of resources
    - facilitation of interdisciplinary and international work on resources
  - and in particular
    - organisation and storage of resources
    - search for resources
- One currently popular strategy:
  - heuristic ontologies
  - from text technology, archive & library technology

SPECIFYING RESOURCES

What is an ontology?

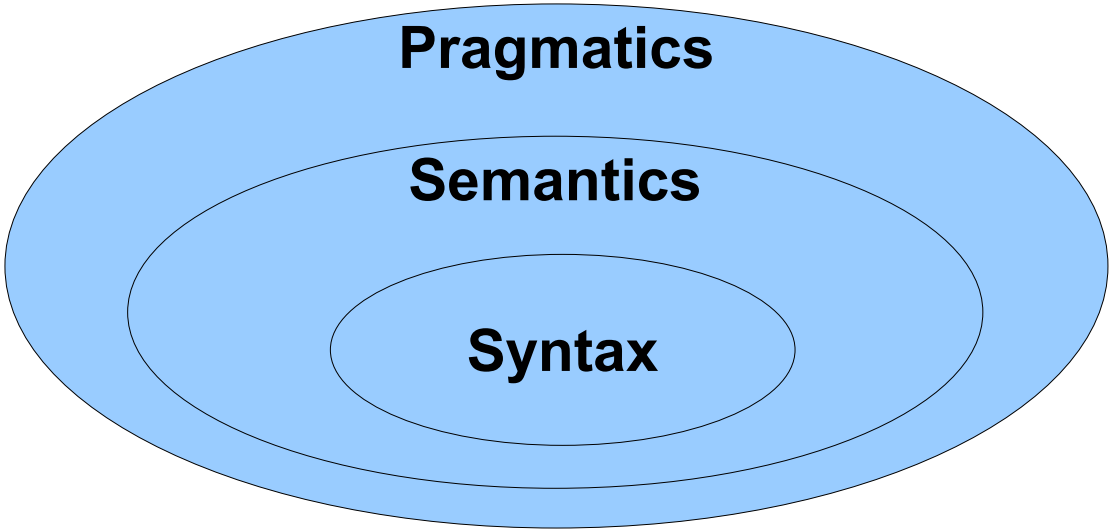
- The E-MELD project definition (Anon 2005):
  - An ontology here is essentially a machine-readable formal statement of a set of terms and a working model of the relationships holding among the concepts referred to by those terms in some particular domain of knowledge. Its purpose is not to define meaning, but to allow computers to navigate human knowledge in a way that mimics intelligence.

What is an ontology?

- The E-MELD project definition (Anon 2005):
  - An ontology here is essentially a machine-readable formal statement of a set of terms and a working model of the relationships holding among the concepts referred to by those terms in some particular domain of knowledge. Its purpose is not to define meaning, but to allow computers to navigate human knowledge in a way that mimics intelligence.
- Simplified:
  - An ontology is a highly structured terminological dictionary designed to facilitate search for information in some technical domain.

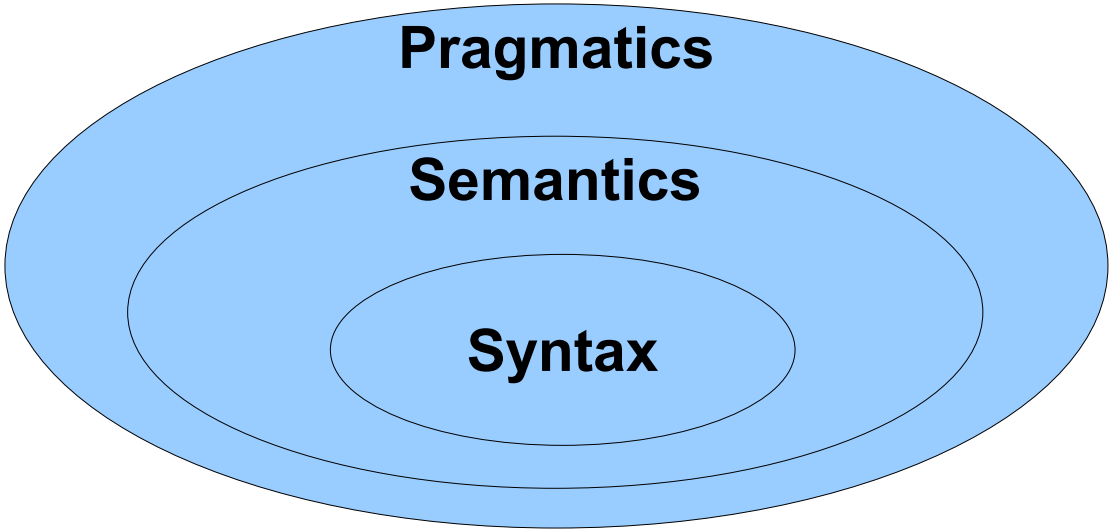
HOW TO DEAL WITH SPEECH?

Classic sign architecture is misguided

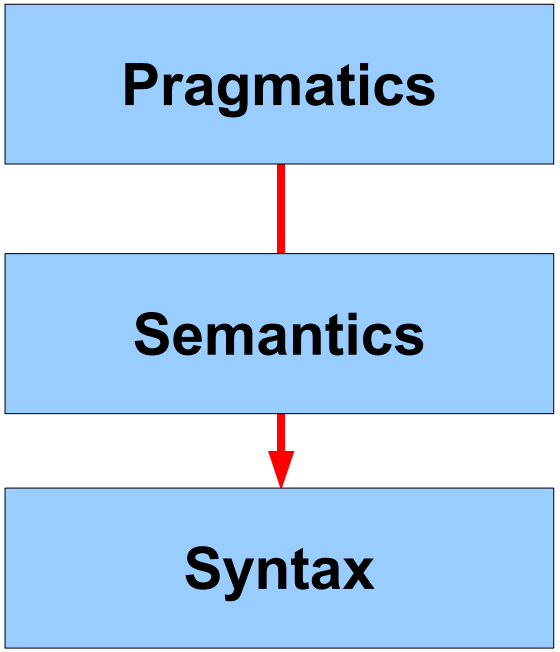


- Carnap:
  - set inclusion
  - onion model

Classic sign architecture is misguided

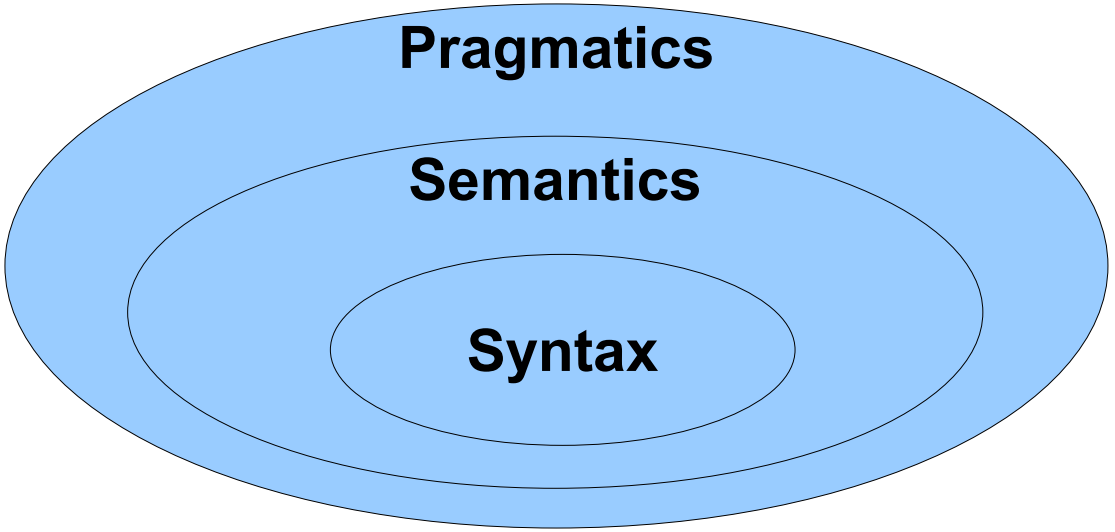


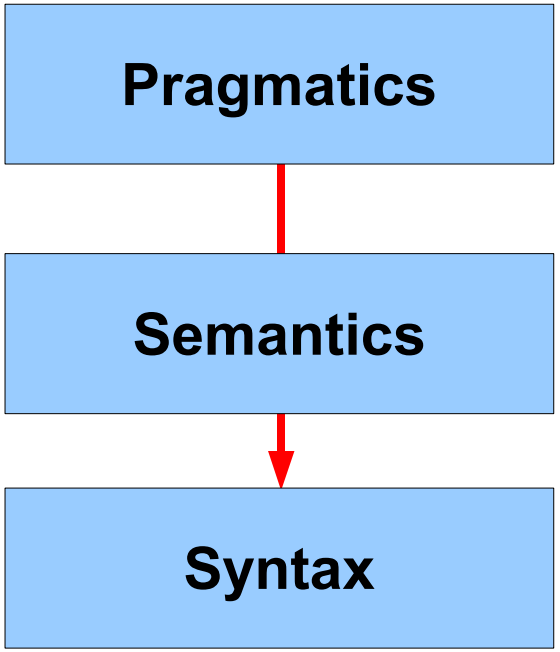
- Carnap:
  - set inclusion
  - onion model



- Linguistic folklore
  - generation model
  - production model

Classic sign architecture is misguided





- Carnap:
  - set inclusion
  - onion model
- Linguistic folklore
  - generation model
  - production model

BUT Pragmatics, Semantics and Syntax are ONTOLOGICALLY QUITE DIFFERENT TYPES !

Why misguided?

- The concepts comes from philosophical logic
  - Charles S. Peirce
  - Rudolf Carnap
- Philosophers and logicians are not interested in phonology and phonetics
- Consequently
  - there is no notion of *expression*
  - no *semiotic duality*
- But in Natural Languages, the SIGN relates to both
  - a CONTENT reality (as in logic)
  - a MODALITY reality

A FUNCTIONAL LINGUISTIC PERSPECTIVE

Non-uniqueness thesis:

- True of content:
  - Spontaneous speech is in many respects different from non-spontaneous speech
    - (whatever spontaneous speech may be)
    - based on additional memory structures, e.g. related to planning, or random access of written media
- BUT not of method:
  - Models of spontaneous speech are not formally different from models of non-spontaneous speech
    - except that reading alound (for example) allows additional memory support in planning prosody etc.
    - spontaneous speech DOES have sentences, though not all sentences may be completed
      - in fact, interlocutors often - rightly or wrongly! - complete them!

So what is needed?

- The most general category is the SIGN
- Then comes the STRUCTURE of the sign:
  - SYNTAGMATIC STRUCTURE
    - combinatorial / compositional / part-whole structure
    - meronomy
  - PARADIGMATIC STRUCTURE
    - classificaton, implication, inheritance, ISA-relation, same/different criteria
    - taxonomy
  - INTERPRETATIVE STRUCTURE
    - realisation, manifestation, expression
    - relation with reality:
      - content domain semantics
      - modality domain “semantics”

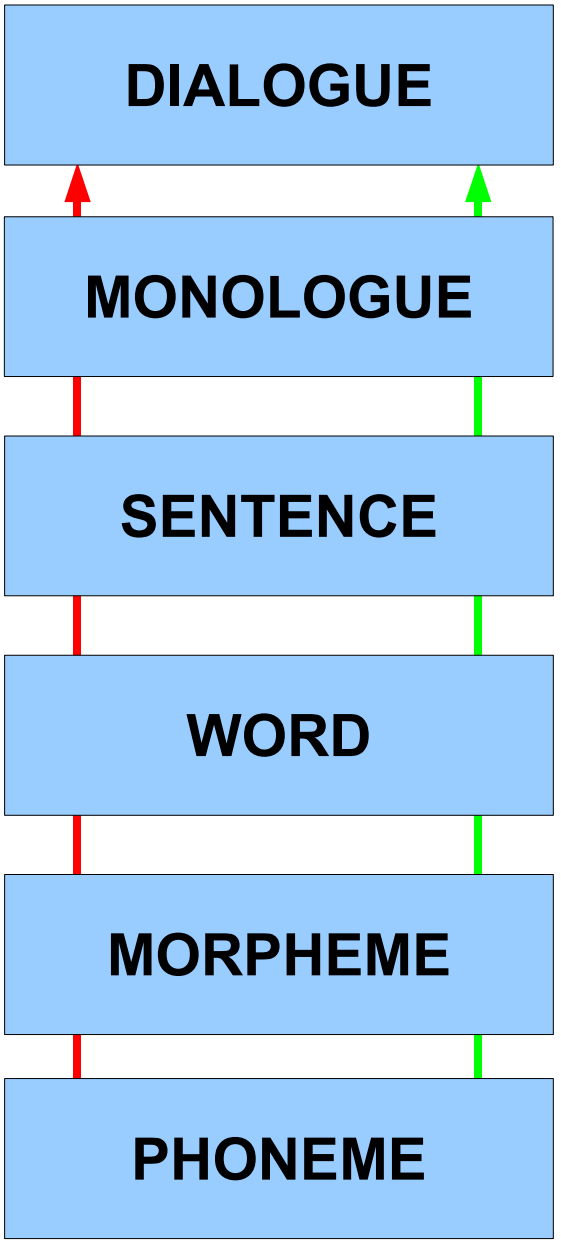
Examples of categories & relations

- SYNTAGMATIC
  - Grammar: *SUBJECT*, *VERB*, *OBJECT*
  - Phonology: syllable *ONSET*, *NUCLEUS*, *CODA*
- PARADIGMATIC
  - Grammar: *noun*, *verb*, *preposition*, ...
  - Phonology: *consonant*, *vowel*, *fricative*, *tone*, ...
- INTERPRETATIVE
  - Grammar: *Sentence Prosody* (intonation etc.)
  - Phonology: *Word Prosody* (phonemes, accent, tone etc.)

Compositional categories

- Major compositional categories:
  - Ranks
    - dialogue
    - monologue
    - sentence
    - word
    - morpheme
    - phoneme
- Minor compositional categories:
  - Within each rank
    - concatenative, linear precedence (LP) structures
    - hierarchical (immediate dominance, ID) structures
    - and sometimes parallel, synchronous structures

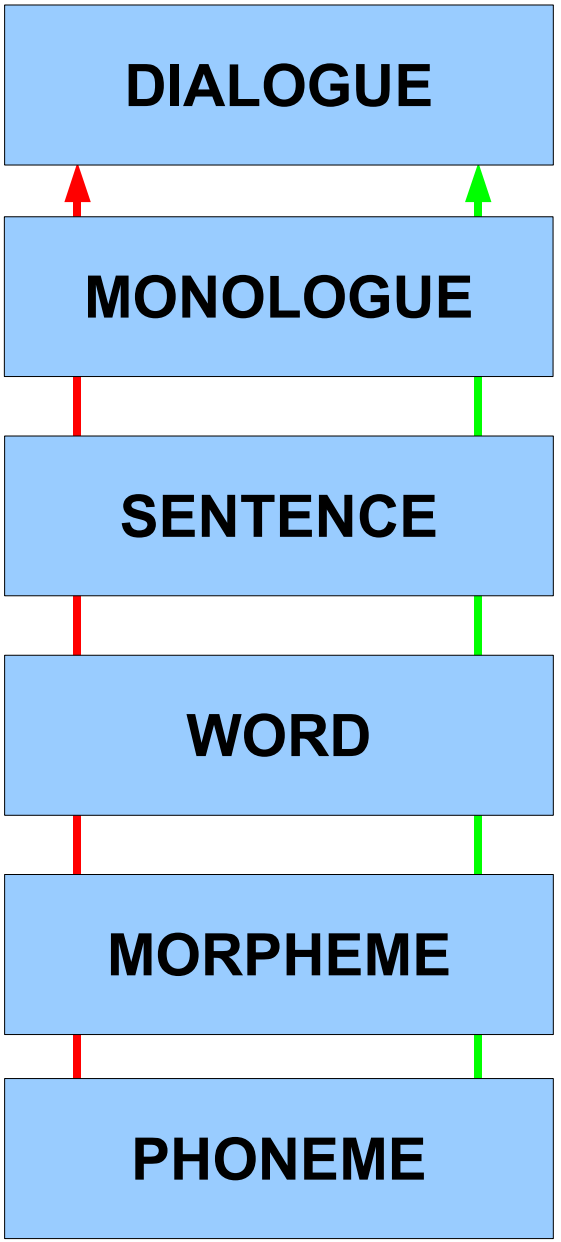
Syntagmatic relations: ranks (simplified)



Two types of "natural" compositionality:

- classic (combinatorial) compositionality:
  - the *P* of a unit is a function of the *P* of its parts

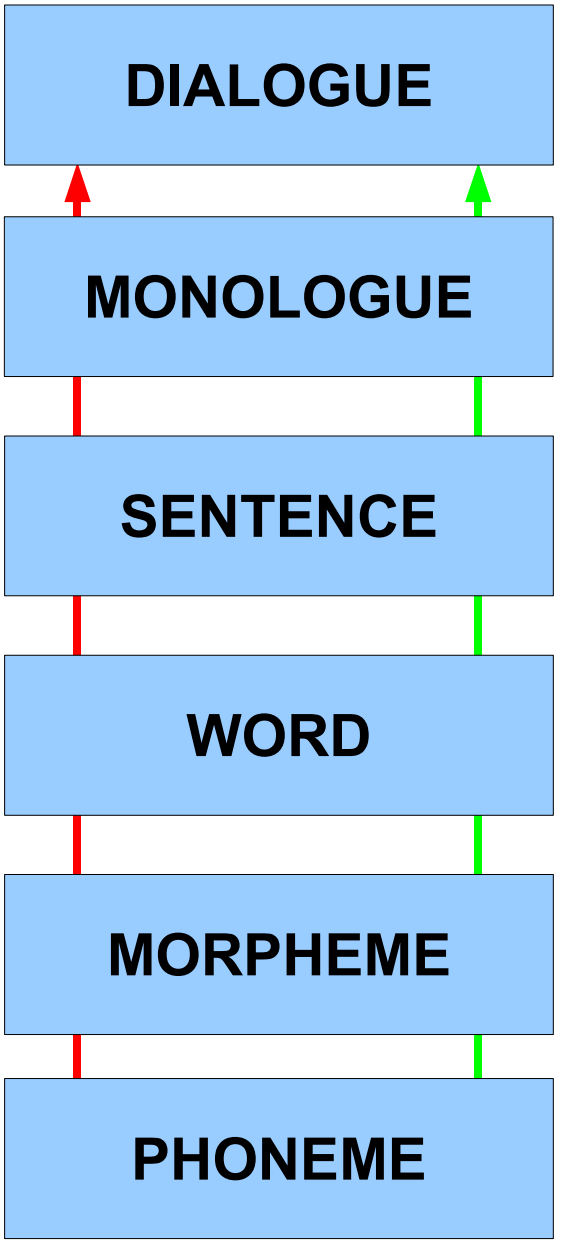
Syntagmatic relations: ranks



Two types of "natural" compositionality:

- opaque (lexical) compositionality:
  - phonematic
  - morphematic
  - elliptical
  - idiomatic

Syntagmatic relations: ranks



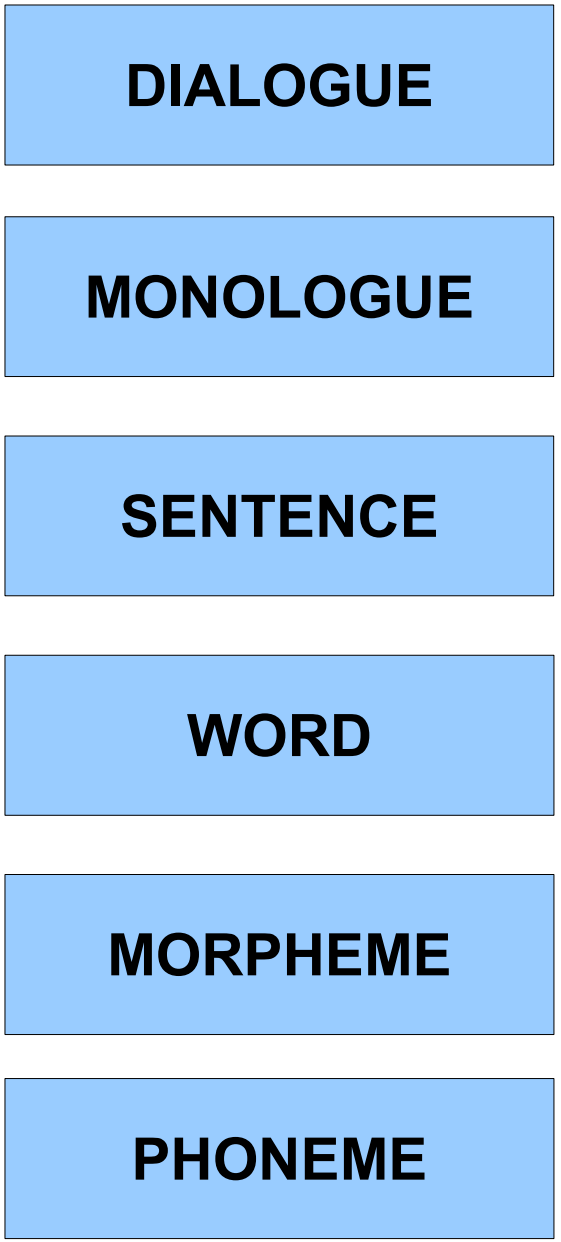
Two types of "natural" compositionality:

- classic (combinatorial) compositionality:
  - the *P* of a unit is a function of the *P* of its parts
- opaque (lexical) compositionality:
  - phonematic
  - morphematic
  - elliptical
  - idiomatic

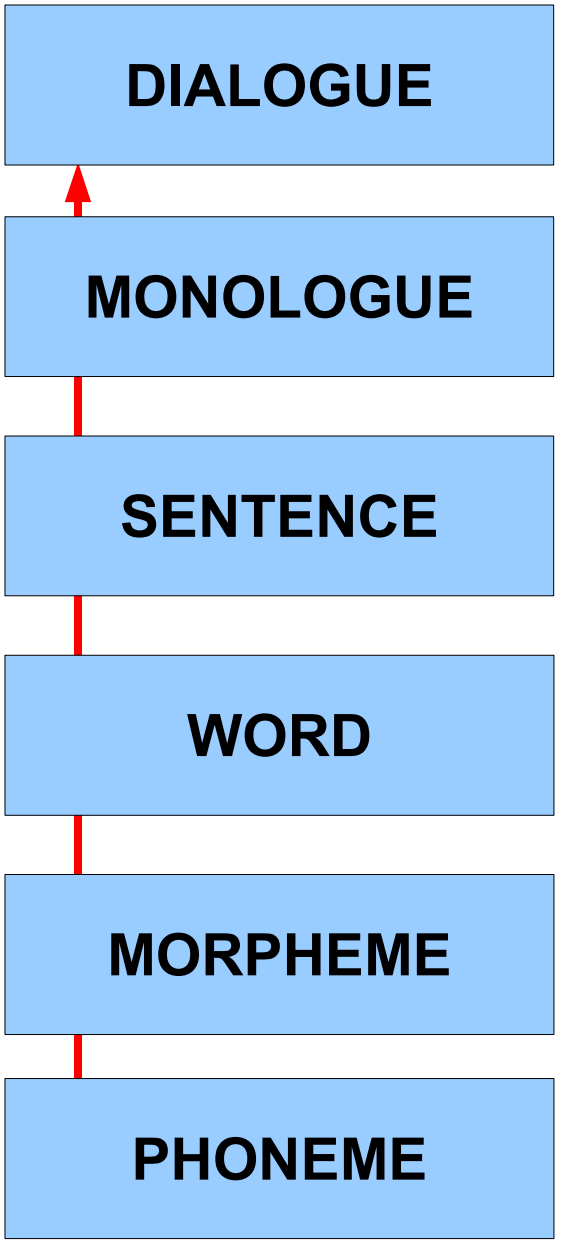
Rank-specific semantics

- Dialogue:
  - turn-taking
- Monologue:
  - narration, argumentation, ...
- Sentence:
  - propositionality
- Word:
  - sets, relations, property ascription, ...
- Morpheme:
  - grounding of words
- Phoneme:
  - contrast

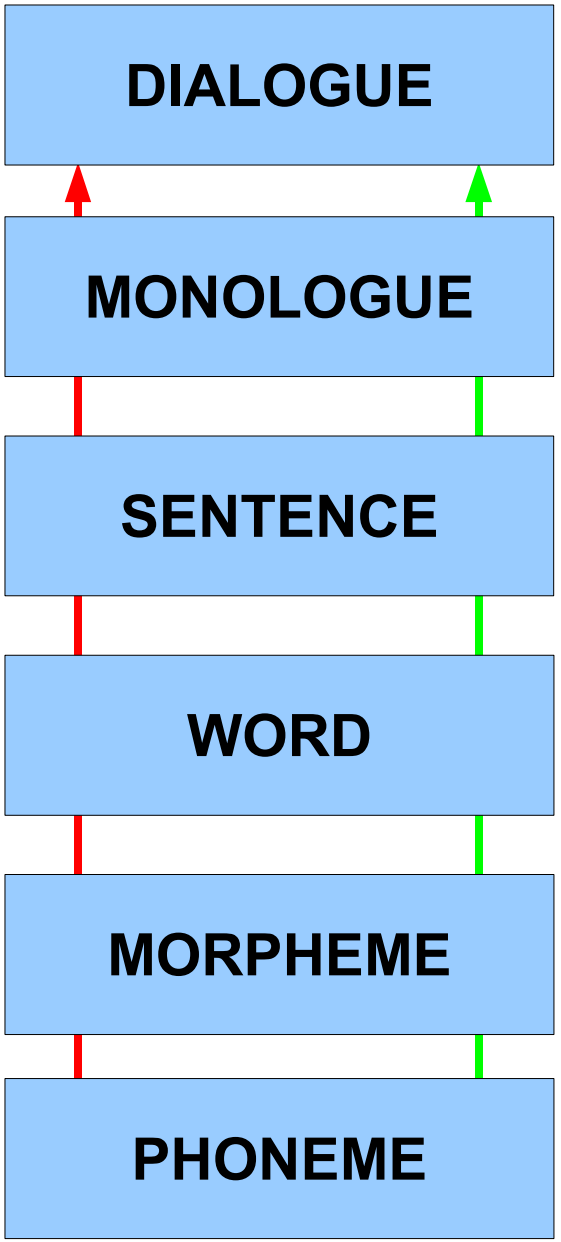
Ranks



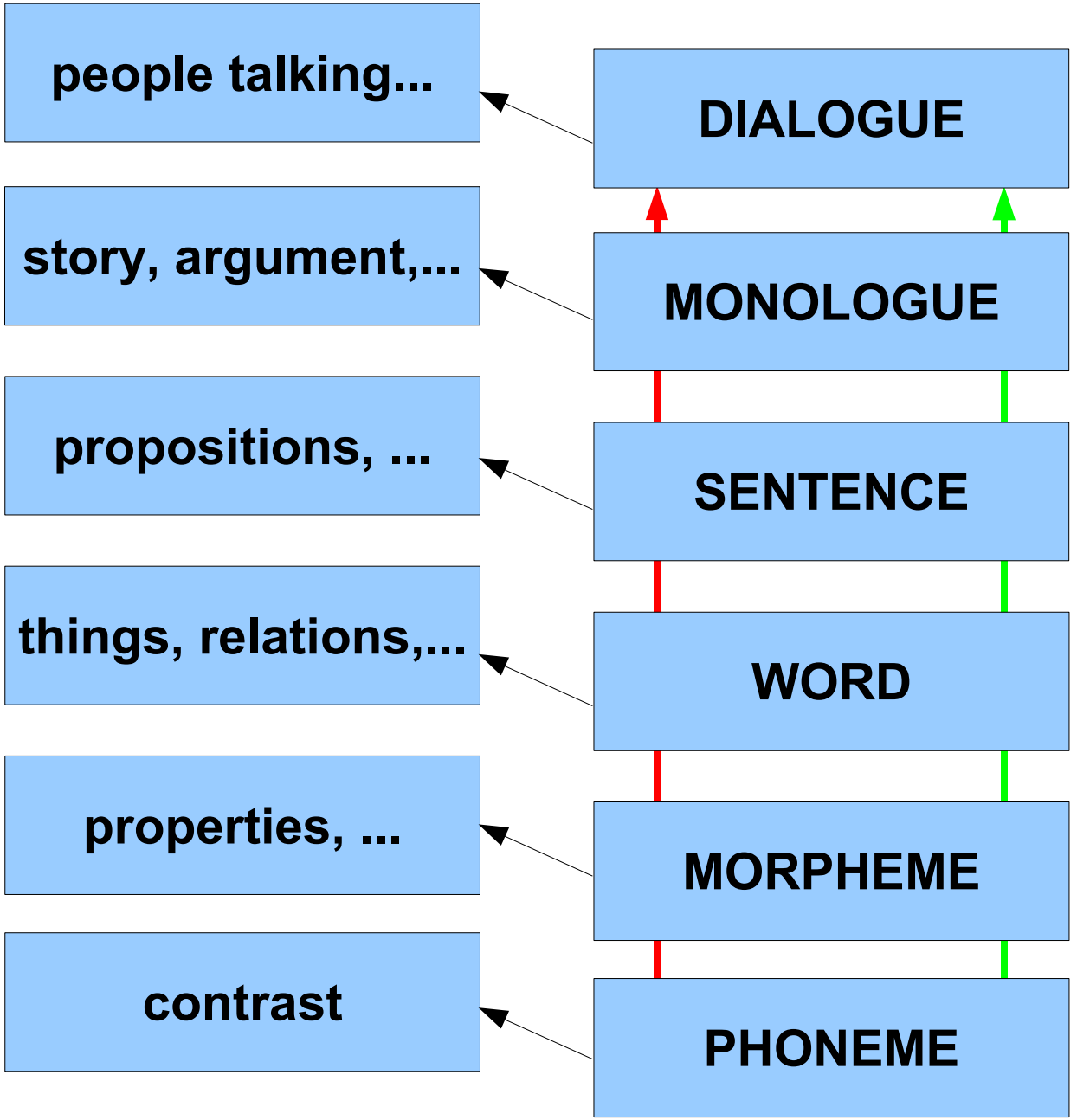
Ranks and combinatorial transparency



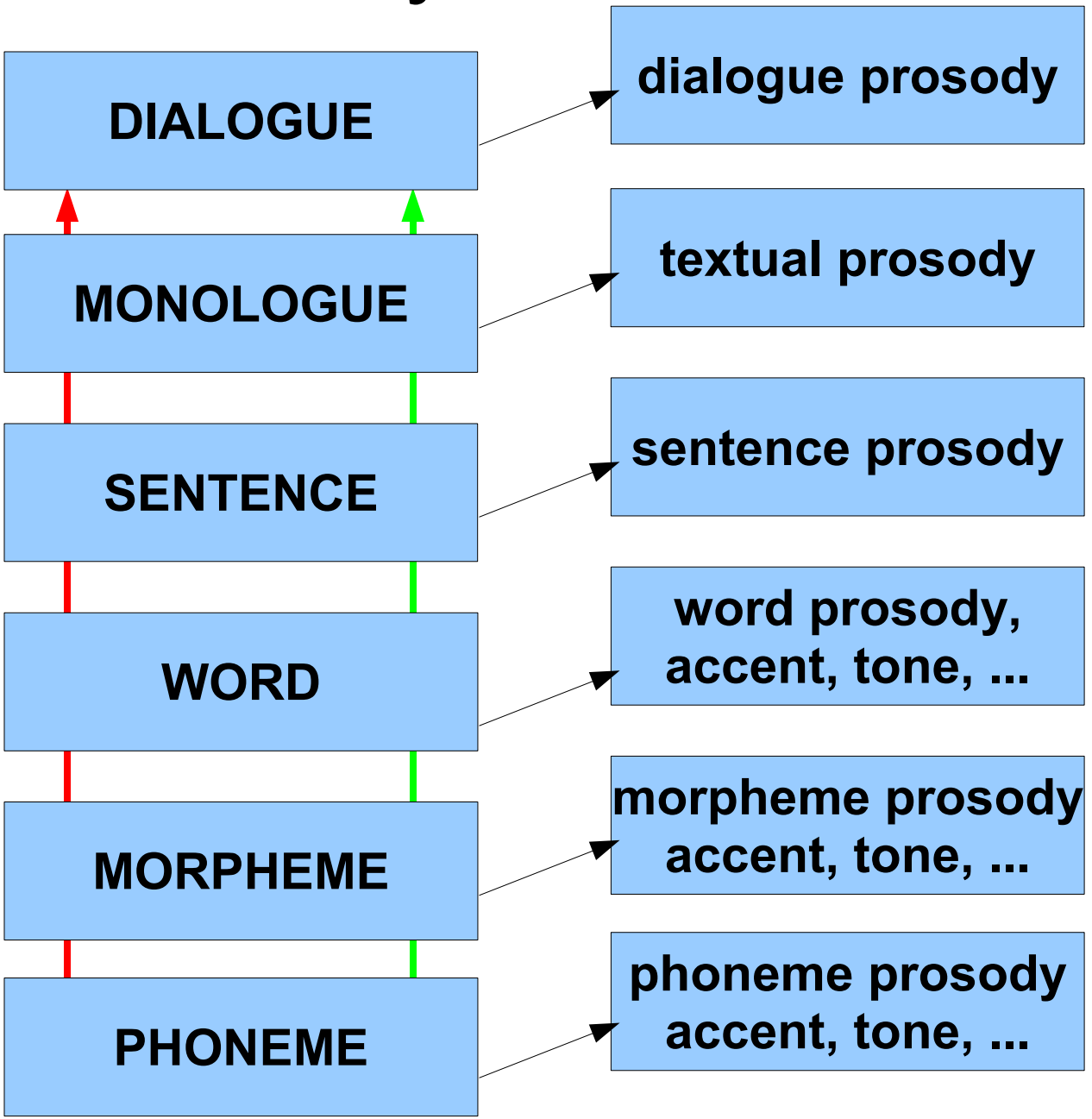
Ranks and combinatorial opacity



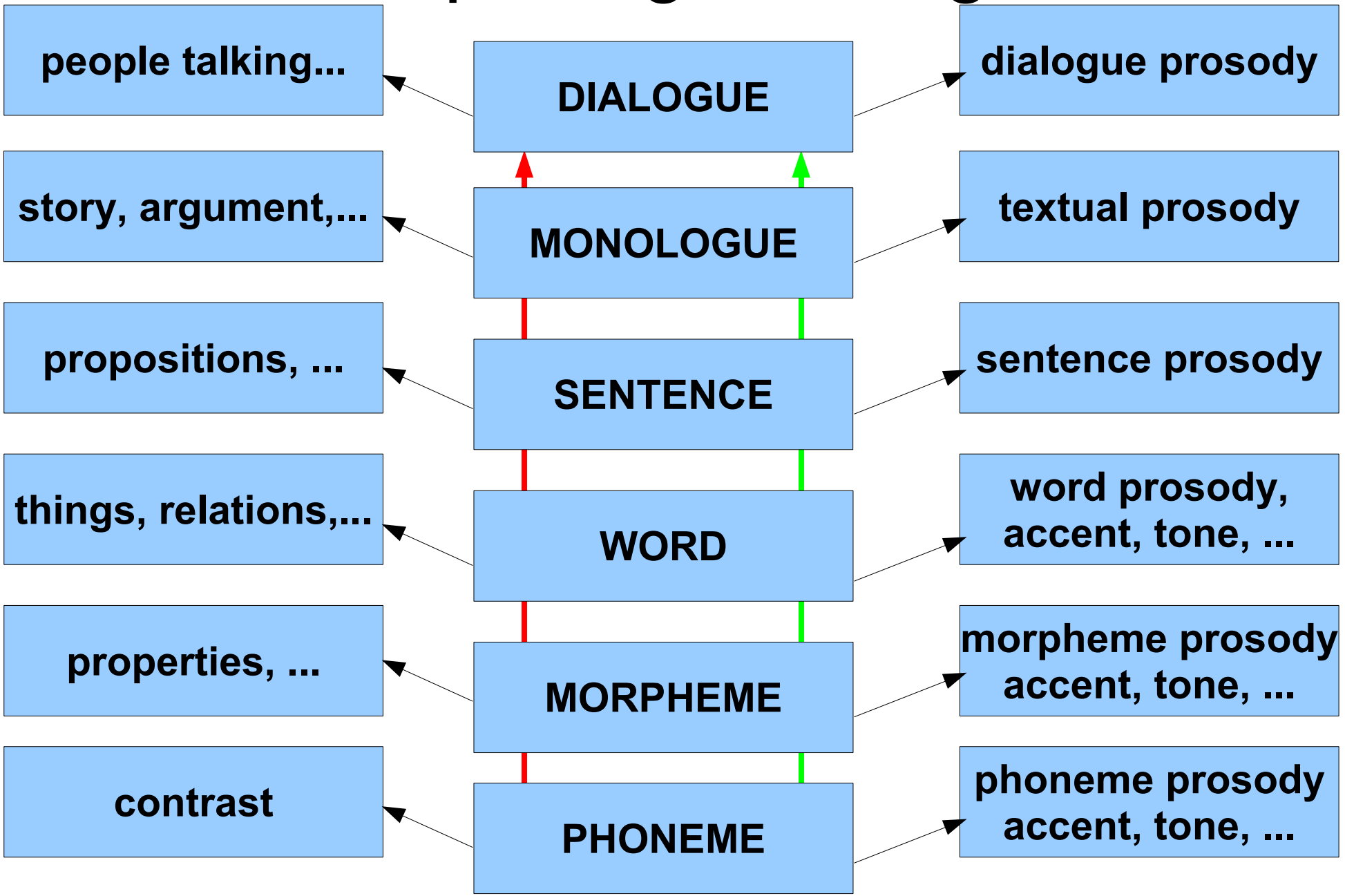
Ranks and domain semantics



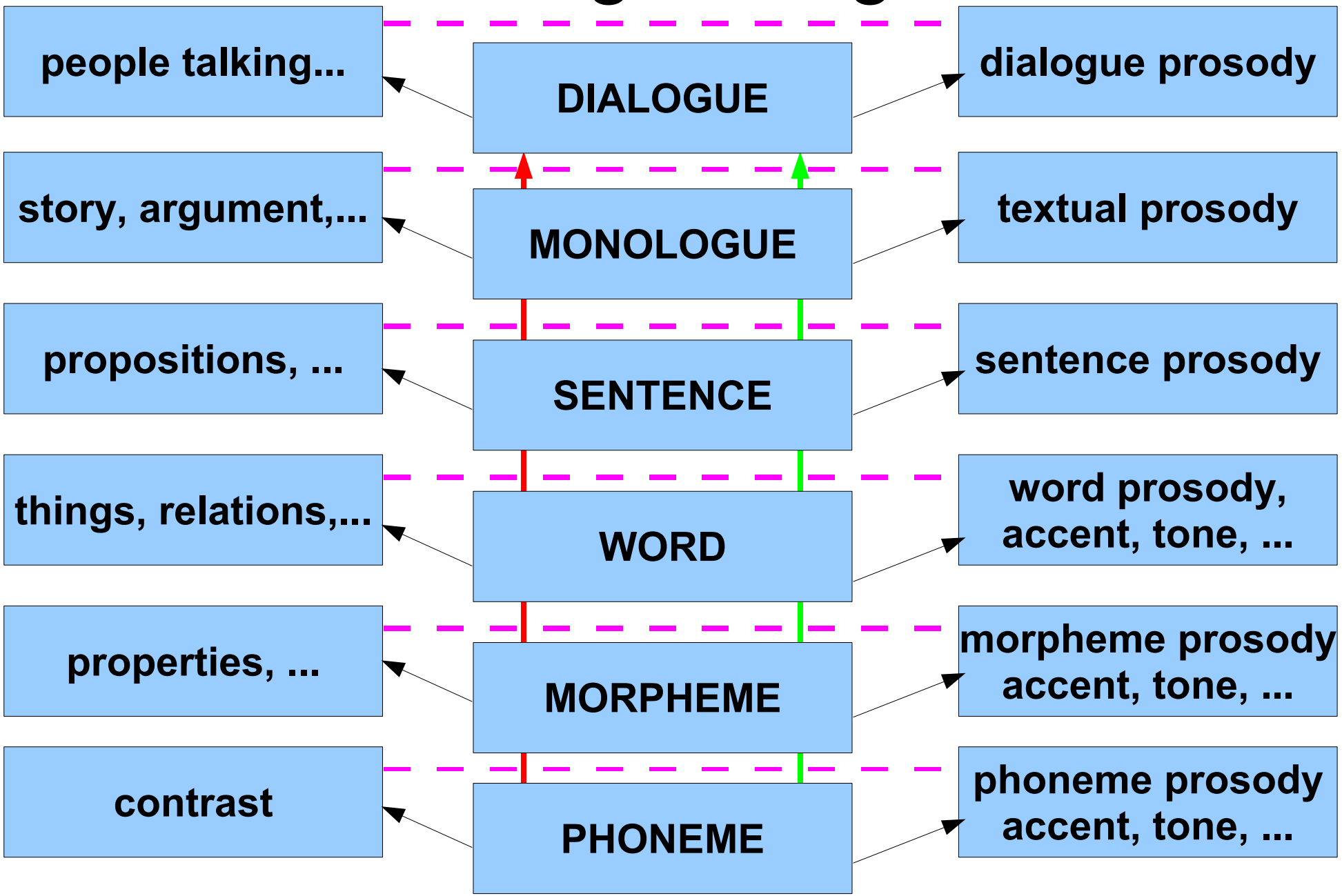
Ranks and modality semantics



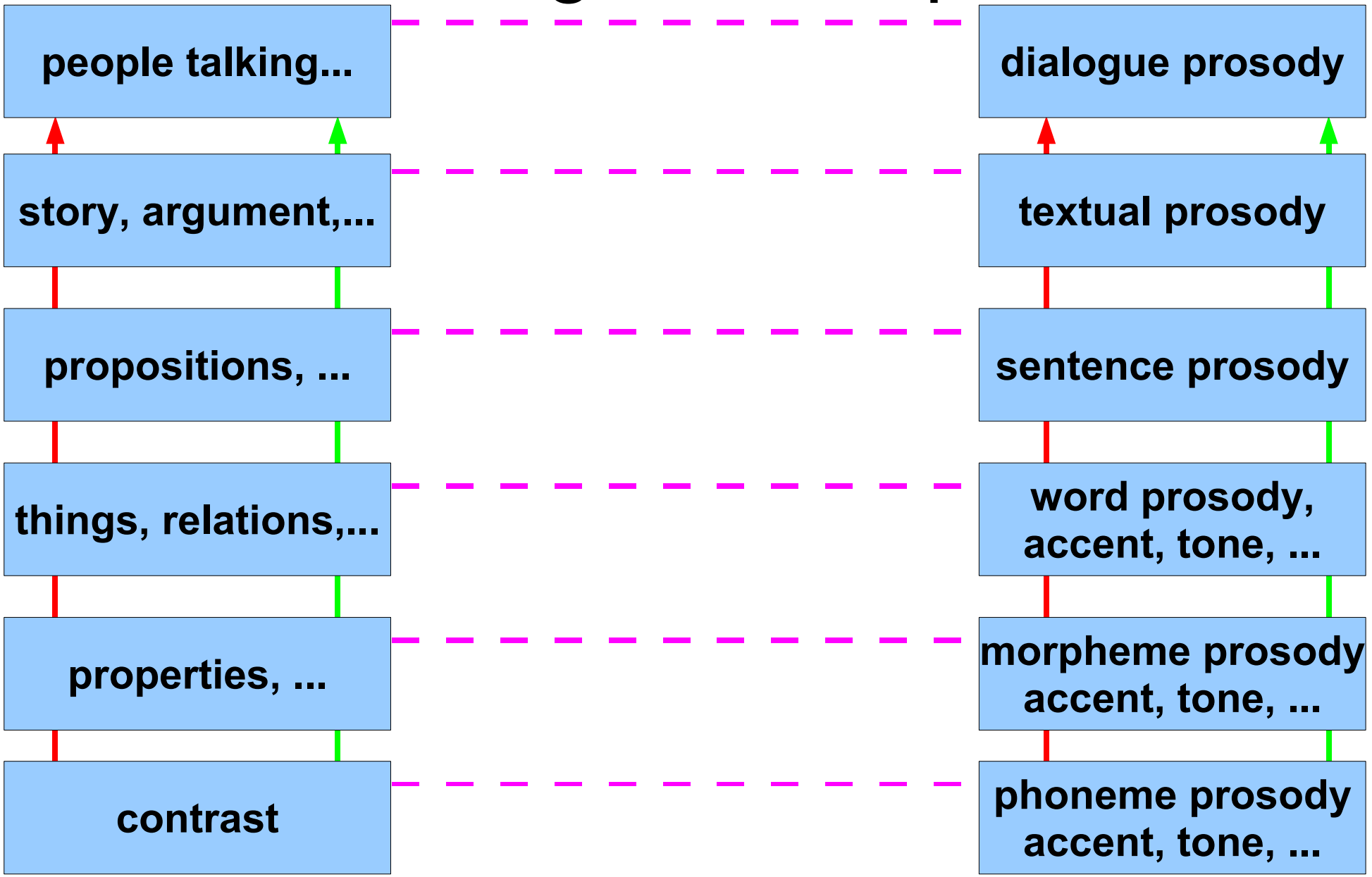
Ranks: putting it all together



Ranks: adding the sign relation



Ranks as signs: a simpler view



ONTOLOGICAL ASPECTS

OF

MARKUP

Adding detail: Time Types

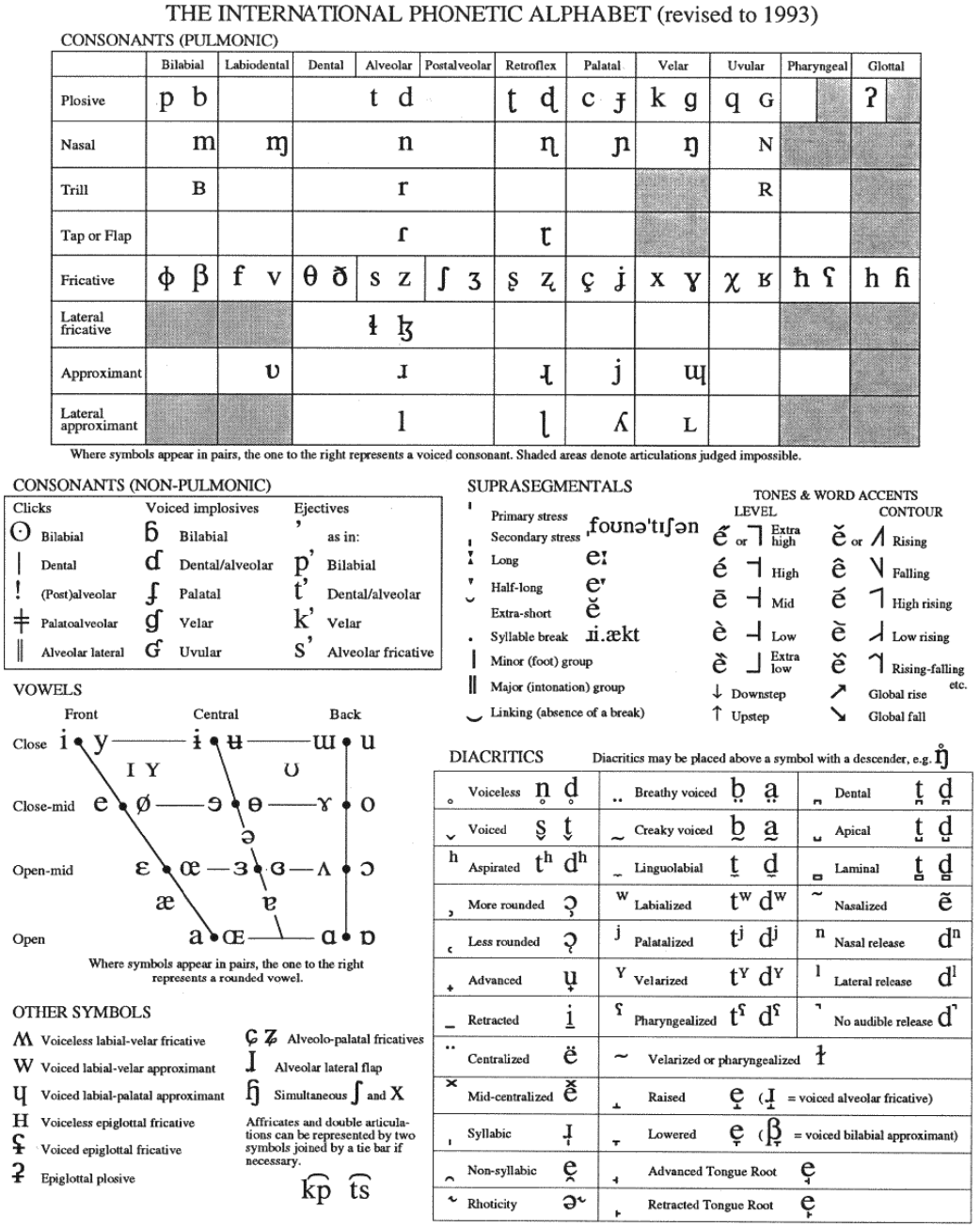
- *Absolute Time* in signal phonetics: time points and intervals determined by calibrated physical measurement.
- *Relative Time* in 'interpretative phonetics', phonology and prosody, and defines points, intervals and precedence/overlap relations in time with no explicit assignment to Absolute Time.
- *Categorial Time* at underlying lexical and grammatical levels, esp. categories linked by algebraic operations such as concatenation, including abstract [DURATION] features.

Annotation “tiers”: some clarification:

- parallel signal *streams*
  - time functions describing continuous or discrete sampled speech signals
- partially aligned with parallel annotation *tracks*
  - time functions describing discrete, categorial sequences of events with signal streams
- in turn often derived from phonological *tiers*
  - linguistic constructs defining partially aligned trajectories through a feature space in Relative Time, as in autosegmental and other prosodic phonologies

Segmental Event Representations

- IPA chart
  - X-SAMPA - the IPA without lipstick
- An excellent basis for an ontology of segments because
  - widely accepted
  - massively tested for over 200 years
  - checked by expert commission
  - widely used in dictionaries



Previous proposals:

- Grammar:
  - Farrar & Langendoen, the General Ontology for Linguistic Description (GOLD)
- Contributions to a segmental ontology:
  - Aristar
  - Kamholz
  - IPA
- But what about *prosody*? There is no equivalent. So at least:
  - ToBI
  - IntSint
  - SAMprosa
  - IPA

BACK TO

MULTIMODALITY

An even harder example

WhatIsTheLadyDoing-1

An even harder example

WhatIsTheLadyDoing-2

Conclusion

- So we are a long way from a full ontology
- And maybe there will never be one
- But after all, all we need is a heuristic for searching our resources
- So there is still hope...
- ... in collaboration with others:

  EMELD: <http://www.emeld.org/>

  GOLD: <http://www.linguistics-ontology.org/>

- **OLAC: <http://www.language-archives.org/>**
- ... and, for example:

... and a proposal

- Proposal:
  - In the context of COCOSDA and O-COCOSDA
  - A new *Handbook of Speech Resources*, but this time more oriented towards work on
    - different language families
    - SpontiSpeech
    - multimodality
    - new archiving and dissemination techniques and standards