

Corpus as Understanding

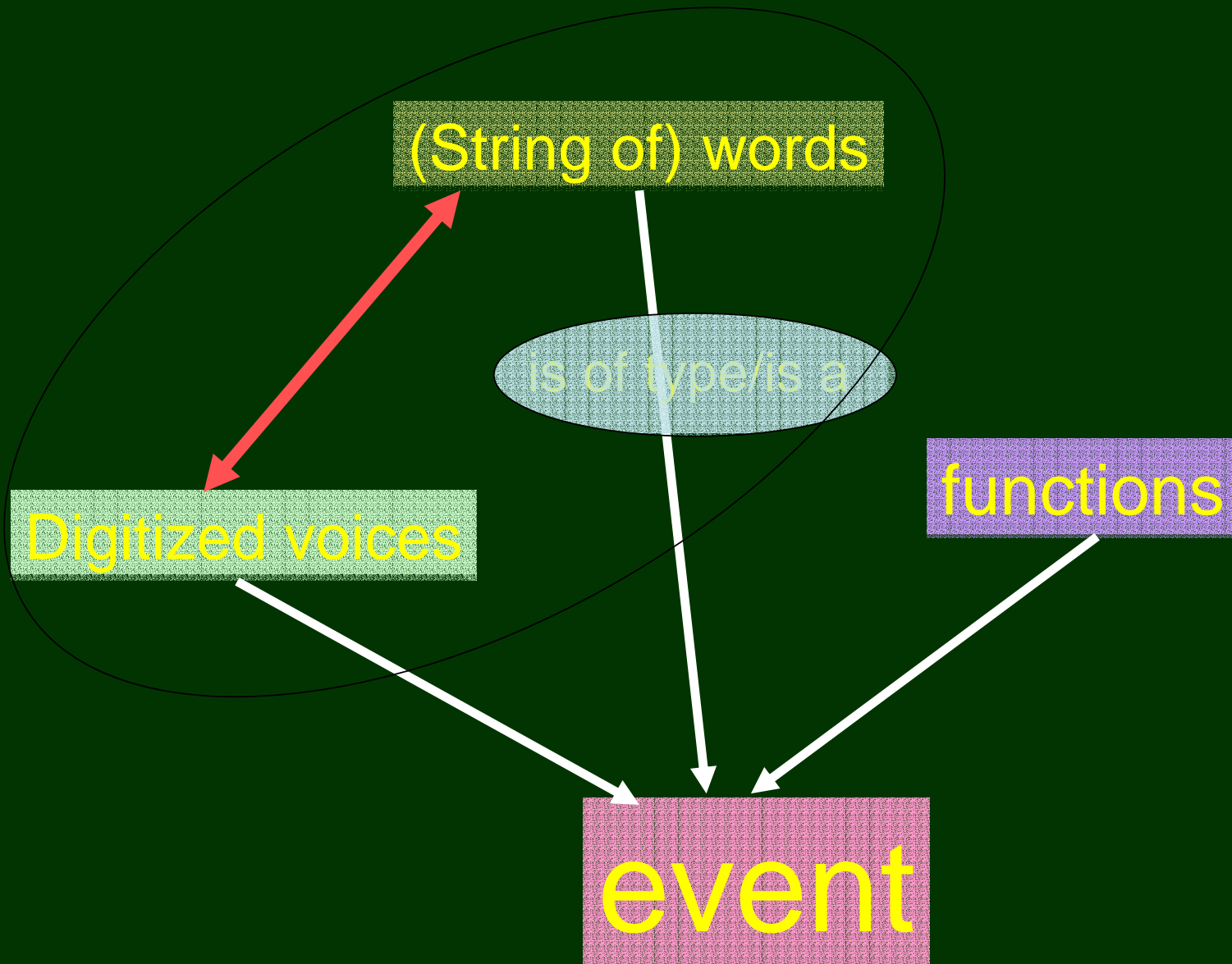
Syun Tutiya
(Chiba University)

- A linguistic corpus is a triple, $\langle G, P, R \rangle$.
 - G: Grounding information, time, space, participating agents
 - P: Perspectives from which to look at the grounded information, like phoneme, prosody, morpheme, word, phrasal structure, “meaning,” speech acts, discourse segments
 - R: Relations between various perspectives. Grammar can be described in terms of morphemes and their orders, etc

Speech corpus

- G: spatio-temporal region
- P:
 - Digitized sounds
 - (“accompanying non-verbal behaviors”)
 - (phonemes)
 - Words and non-words
 - “Grammar”
 - Turn(?)
 - Speech acts
 - Discourse structures
 - Tasks
- R: words/Grammar/Speech acts

Speech corpus as (contextual) dictionary



Corpus as understanding

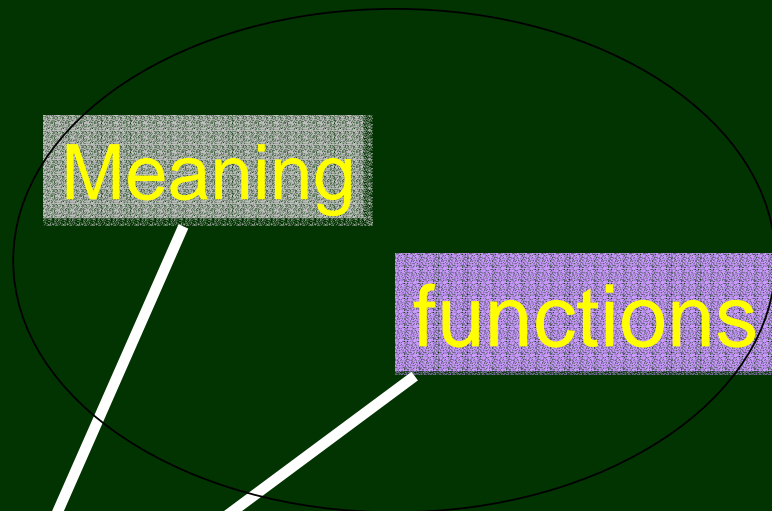
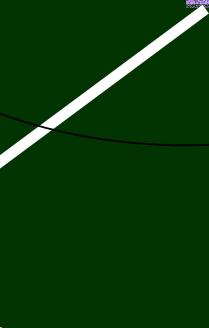
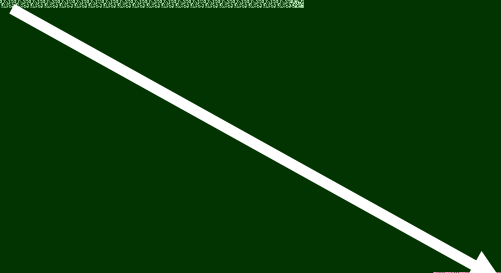
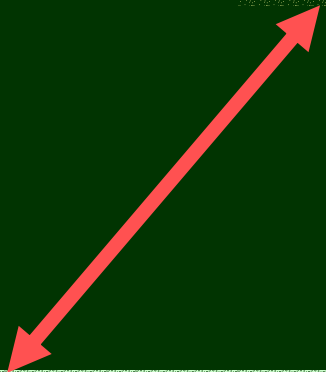
(String of) words

Meaning

functions

Digitized voices

event



Corpus as understanding

- Creation of corpus is understanding of events
- No oneway processing, but leveraged increase of “understanding”
- Maybe this is “fusion, “ but no idea idea.