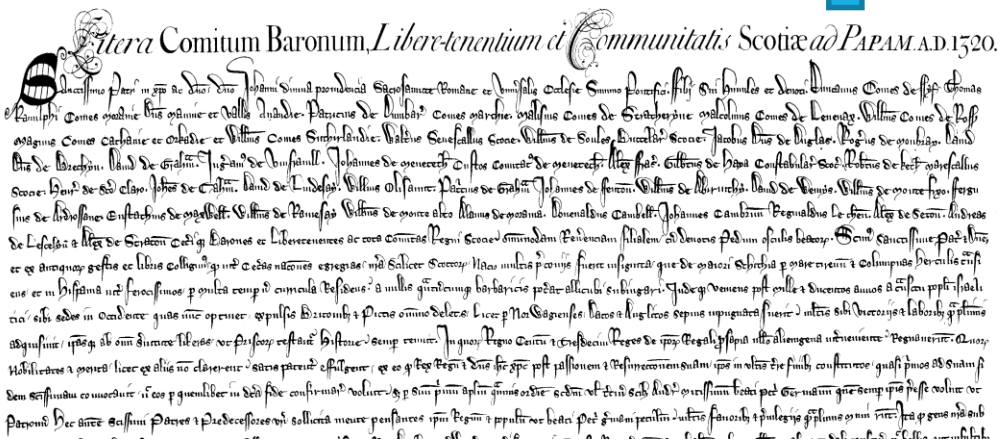
Bird’s Eye View

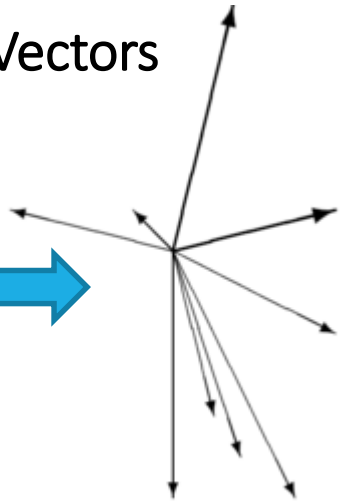
Semantic Analysis in NLP

Logical Forms

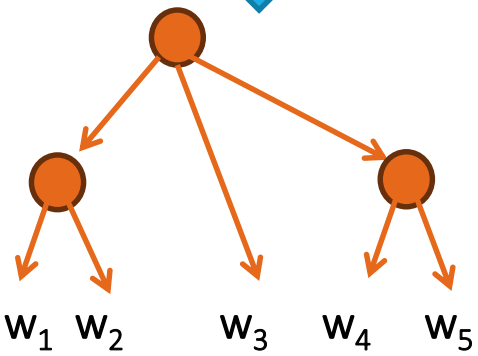
$\lambda x.p_1(a,b) \wedge p_2(c,x)$



Vectors



Trees/DAGs

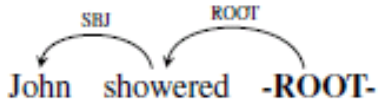


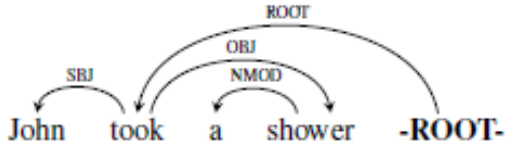
KBs

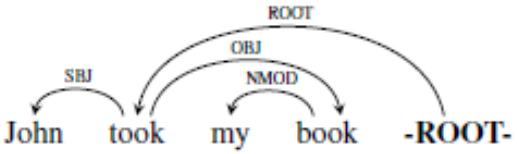
Symbolic Semantic Representation

- The focus of this tutorial: symbolic, sentence-level (or few sentences max)
- By “semantic” NLPers refer to many different things
- Some of which:
  - Representation that supports inference
  - Representation relates to the text to some extra-linguistic semantics (grounding)
  - The compositional structure of a sentence/text
  - An invariant of “meaning-preserving” variation (translation or paraphrase)

Semantic Structures: Stability to Paraphrasing







"John showered" ≈ "John took a shower" ≠ "John took my book"

Hebrew:
ג'ון התקלח
John showered-himself

Hebrew:
ג'ון לקח את
הספר שלי

Semantic Structures: Stability to Paraphrasing

Syntactic Schemes	Semantic Schemes
founding of the school	founding of the school
president of the United States	president of the United States
United States president	United States president

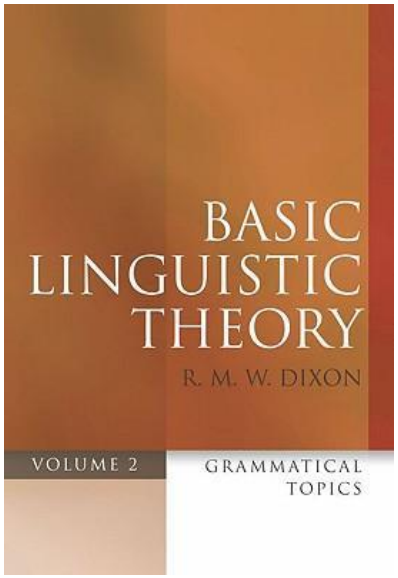
Why not just have Syntax?

- Syntactic structure is very useful, but
  - Syntactic schemes often under-specify, or are orthogonal to semantic distinctions
  - Syntax varies considerably across languages (translation divergences; e.g., Dorr, 1994)
- Accessibility to non-expert annotators
  - Syntactic annotation requires highly proficient annotators
  - Can semantic structure be more accessible?

UCCA: Design Principles

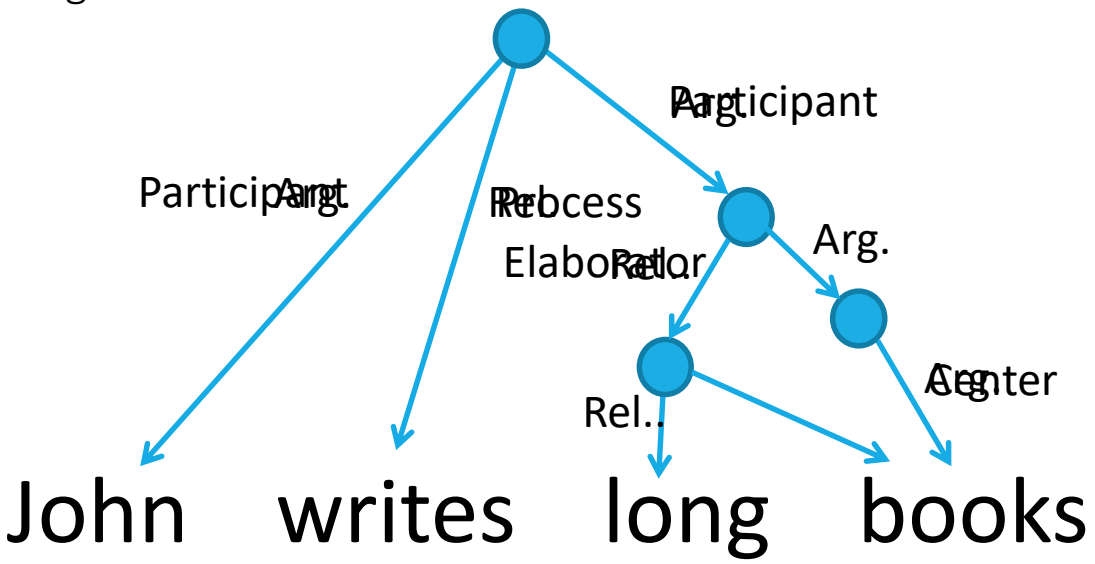
1. Abstract away from formal variation
2. Cross-linguistic applicability
3. Accessibility to non-expert annotators
4. Modularity

Theoretical foundations:
Mostly, *Basic Linguistic Theory*, a typological descriptive framework by *R.M.W. Dixon*



UCCA: Formalism

- Terminals
- Units
- Relations and arguments
- Categories
- Layers



UCCA's Foundational Layer

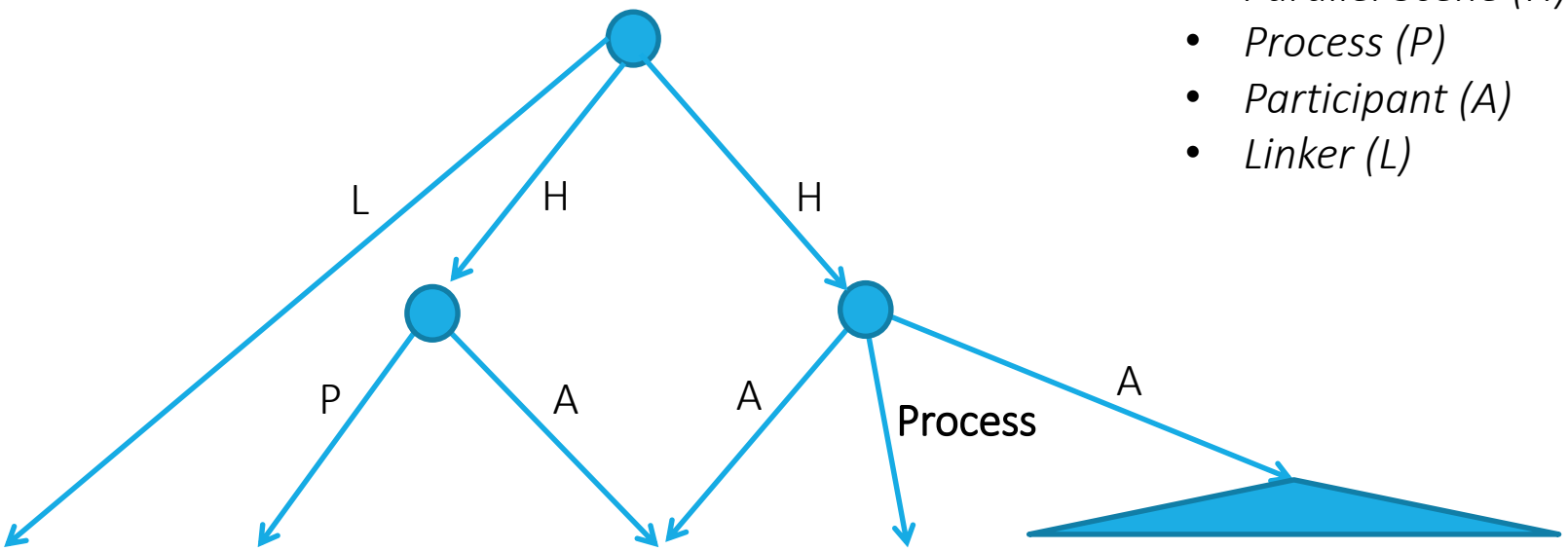
- **Focus:** semantic heads, predicate-argument relations and linkage between them
- Maximally coarse-grained (14 categories)
- Based on the semantic aspect of Basic Linguistic Theory's definition of a clause

UCCA's Foundational Layer: Scenes

"After graduation, Sorkin moved to New York City where he worked odd jobs including delivering telegrams, and driving a limousine."

UCCA's Foundational Layer: Scenes

- *Parallel Scene (H)*
- *Process (P)*
- *Participant (A)*
- *Linker (L)*

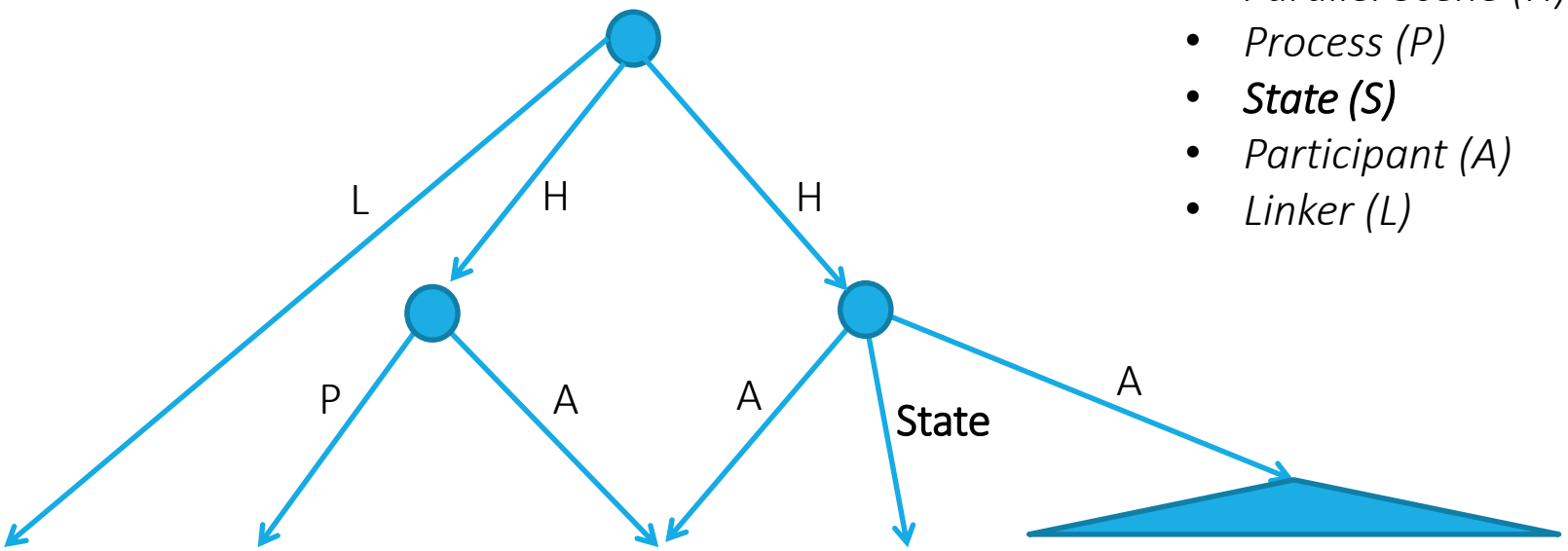


"After graduation, Sorkin moved to New York City"

- Nouns/adjectives/prepositions (etc.) can evoke scenes
- Participants need not be syntactic arguments

UCCA's Foundational Layer: States

- *Parallel Scene (H)*
- *Process (P)*
- ***State (S)***
- *Participant (A)*
- *Linker (L)*

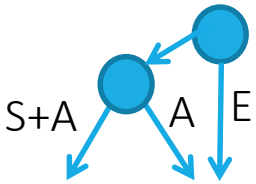


"After graduation, Sorkin stayed in New York City"

Sensitivity to Content, not Syntactic Categories

- Verbs can be adverbials:
  - "John **began**$_D$ swimming"
- Prepositions can be many things:
  1. Case markers: "Yossi lives **in**$_R$ Jerusalem"
  2. Linkers: "**After**$_L$ graduation, Sorkin moved to NYC"
  3. Scene-evokers: "The tree is **in**$_S$ the garden"

- Possessives can mark
  1. A State: "**your**$_{S+A}$ **car**$_A$"
  2. Part-whole relation: "**your**$_C$ **hands**$_C$"
  3. Participation: "**your**$_A$ **kick**$_P$"



Legend:
- *Process (P)*
- *State (S)*
- *Participant (A)*
- *Adverbial (D)*
- *Elaborator (E)*
- *Center (C)*

Inter-Scene Linkage

"After graduation, Sorkin moved to New York City where he worked odd jobs including delivering telegrams, and driving a limousine."

Scenes:

- "graduation$_P$ … Sorkin$_A$"
- "Sorkin$_A$ moved$_P$ [to New York City]$_A$"
- "he$_A$ worked$_P$ [odd jobs]$_A$"
- "he$_A$ … delivering$_P$ telegrams$_A$"
- "he$_A$ … driving$_P$ [a limousine]$_A$"

"and"

"after"

"where"

"including"

Coarse-grained, Refinable

- Two additional layers for UCCA:
  1. Semantic roles / Preposition supersenses (Schneider et al., 2018; Prange et al., 2019):

*Possession is **not** scene-evoking:*

- Kinship: "John's sister"

- Part-Whole: "The car's windshield"

Possession is scene-evoking:

- Agent: "John's kick saved the game"

- Patient: "The boy's murder was never reported"

- Ownership: "John's computer"

Coarse-grained, Refinable

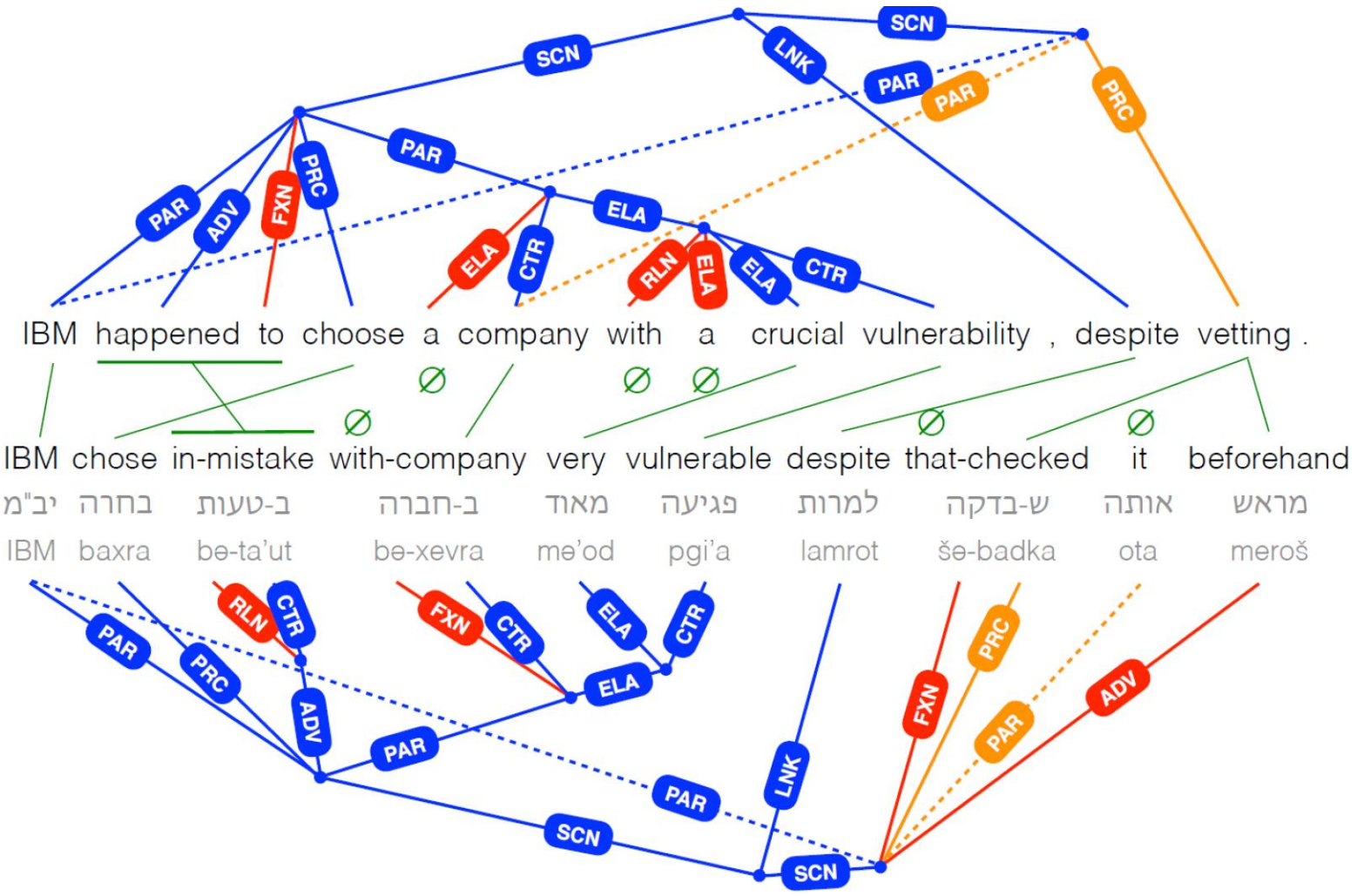
- Two additional layers that refine the foundational layer:
  1. Semantic roles / Preposition supersenses (Schneider et al., 2018; Prange et al., 2019a):

     *Possession is **not** scene-evoking:*
     - Kinship: "John's sister"
     - Part-Whole: "The car's windshield"

     Possession is scene-evoking:
     - Agent: "John's kick saved the game"
     - Ownership: "John's computer"

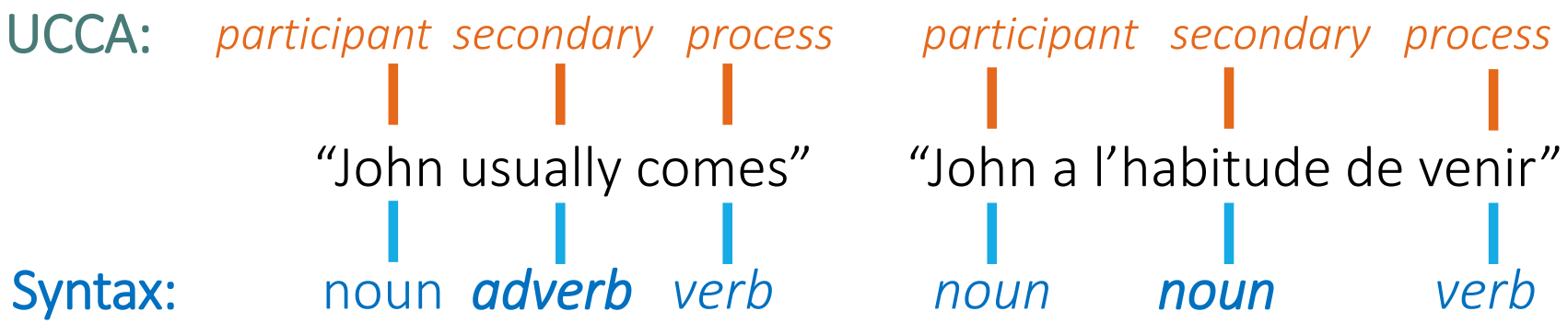
  2. Coreference resolution (mentions are constrained by UCCA's foundational layer; Prange et al., 2019b)

Translation Divergences / Stability



A Corpus Study of Translation Stability

- How well does UCCA *preserve* structure across English-French translations?



Cross-linguistic Stability: Results

- Scene divergences:

  → 92% of the English Scenes and 95% of the French Scenes have a correspondent on the other side

- Comparison to syntax: stability of the number of Scenes/paragraph vs. the number of clauses/paragraph

  → UCCA is more stable than PTB-style syntactic trees

UCCA: Cross-linguistic Stability

Nevertheless, occasional UCCA divergences are found:

Officers were probing the increasing gloom with their night glasses

Les officiers **armés$_S$ de leur lorgnette de nuit** fouillaient l'obscurité croissante