Natural Language Processing

- Language processing problem
  - Processing written text using
    - Syntax, semantic knowledge of the language
    - Real world information
  - Processing spoken language using
    - All information above
    - Knowledge about phonology
    - Additional information to resolve ambiguity
- English
  - Most rules apply to other evolved languages also

English – Useful and difficult

- Incomplete descriptions
  - I called her to ask her to the movies and she agreed (Can be vague or precise!)
- Same expression means different things
  - Where is the fire? (Manage infinite world with finite words!)
- New words, phrases, meanings
  - I will google it! (Language can evolve)
- Different ways to say the same thing
  - He was born on the 20th/His birthday is the 20th (When we know a lot, sentences imply each other)

Understanding and NLP

- Understanding
  - Process of mapping from an input form to a more immediately useful form
- Language can be
  - A set of strings without reference to any world or task to be performed! – not useful in AI!
- NLP in AI
  - Task at hand
  - Target representation appropriate for task at hand
  - Source language and the mapping into target representation
  - When we talk about one, the other is always present

NLP Steps

- Morphological Analysis
  - Individual terms/words are analysed into components
  - Non-word tokens are separated from words
- Syntactic analysis
  - Linear sequences transformed into structures that show how words are related
  - Sequence will be rejected if language rules are violated
  - The boy the go to store the
- Semantic analysis
  - Structures created in syntactic analysis are assigned meanings
  - **Mapping** is made between syntactic structures and objects in task domain
  - Structure without such mapping are rejected
  - Eg: colourless green ideas sleep furiously!

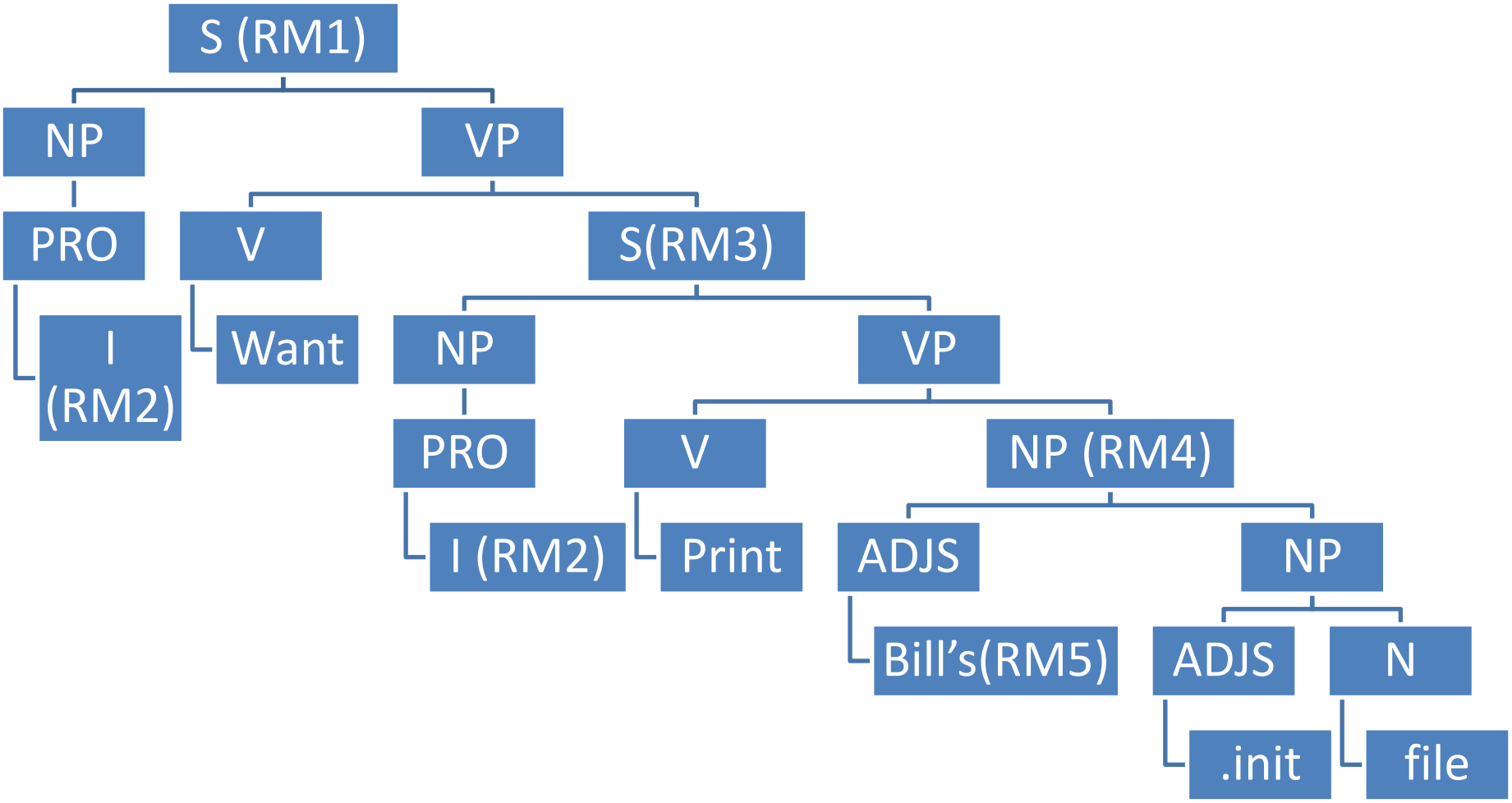
NLP Steps – Contd.

- Discourse integration
  - Meaning of a sentence may depend on the sentences that precede it
  - Meaning of a sentence may influence the meanings of sentences that follow it
  - Eg: The sword shined. He wanted it. He always had.
- Pragmatic analysis
  - Structure representing what was said is reinterpreted to determine what was actually meant.
  - Eg: Do you know what time it is?
- No clear boundaries between these steps
- Can happen sequentially or in parallel

NLP Steps – Example

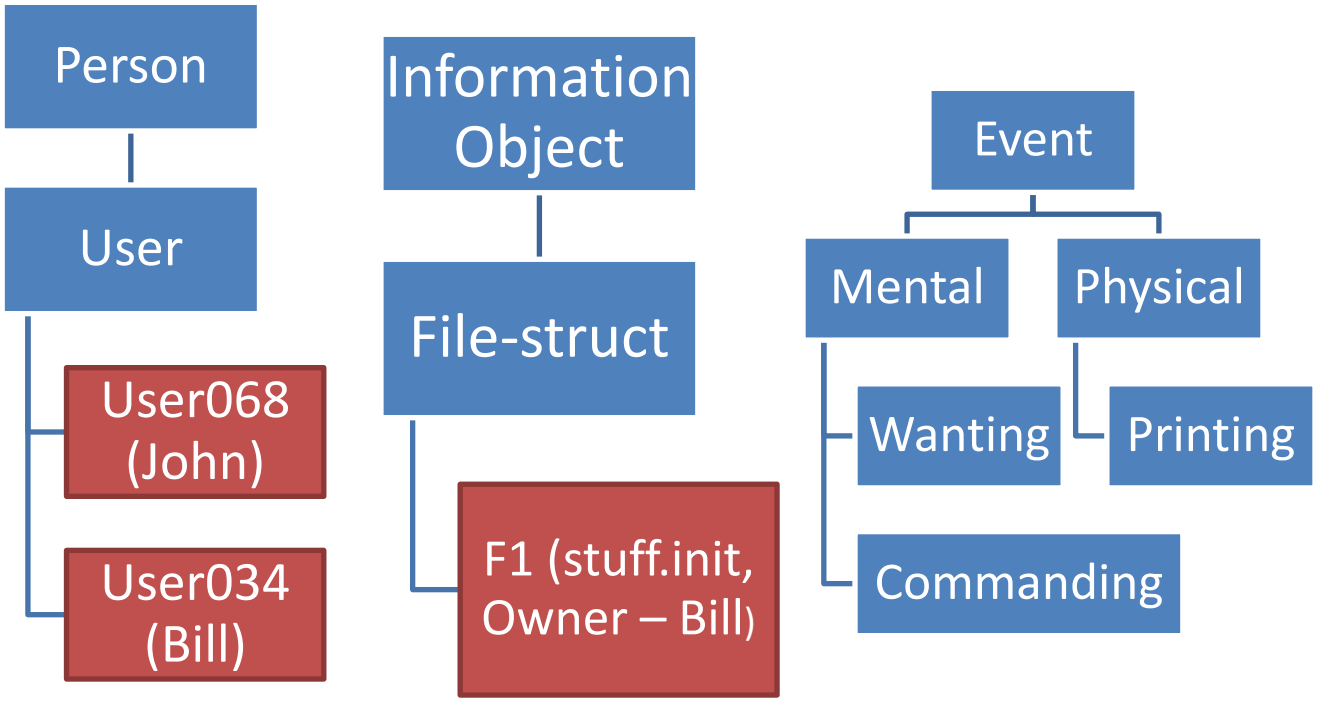
- Process ***I want to print Bill's .init file***
- Morphological analysis
  - Separate sentence into tokens
  - Bill's: ***Bill*** and possessive suffix ***s***
  - .init: recognize the file extension, and that it's an adjective
- Syntactic analysis
  - Build structural description of sentence: parsing
    - Convert flat list of words into a structure that defines the units forming the flat list

NLP Steps – Example

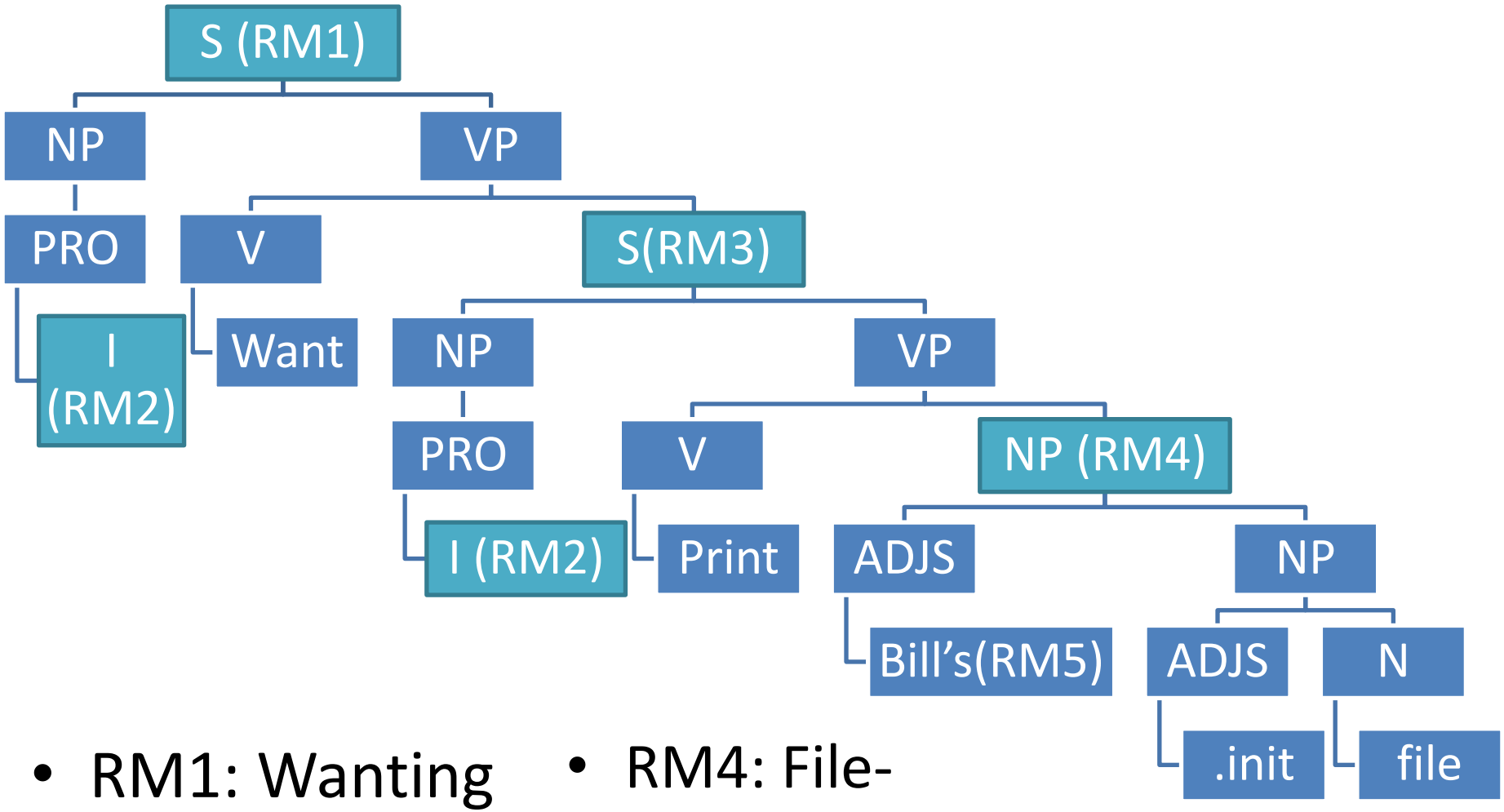


Semantic analysis

- Goals:
  - Map individual words into appropriate objects in the knowledge base / database
  - Create correct structures represent the meanings of the individual words combined with each other
- Assume we have a frame based knowledge base



NLP Steps – Example



- RM1: Wanting
- RM2
- RM3: Printing
- RM4: File-struct (.init)
- RM5: Person (Bill)

NLP Example – Contd.

- Discourse integration
  - Who am “I”?
  - Who is Bill? Which Bill?
  - Need to maintain a model of current discourse (context)
- Pragmatic analysis
  - Understand what needs to be done
  - Declaration: record and we are done
    - Eg: Sun rises in east
  - Others (intention, possibility, etc.): apply a set of rules that characterize cooperative dialogues
    - User claims to want to do something
    - System is capable of doing it
    - Translate from knowledge-based representation to a command

Syntactic processing

- Parsing
  - Flat input sentence into a hierarchical structure that corresponds to the *units of meaning* in the sentence
- Grammar: a declarative representation of the syntactic facts about the language
- Parser: A procedure that compares the grammar against the input sentence to produce parsed structures
- Parsing types
  - Top Down Parsing:
    - Begin with start symbol
    - Apply grammar rules forward until symbols at terminals correspond to components of sentence
  - Bottom-up parsing
    - Begin with sentence
    - Apply grammar rules backward until a single tree with terminals as words in sentence is formed with top node as start symbol
  - Choose based on branching factor

Multiple possible interpretations

- Process of understanding is a search process
  - Large number of possible interpretations
  - Find one that meets all constraints posed by the sentence
  - Explore all paths? Or a single most likely one?
- Eg: Have the students who missed the exam …
  - Two paths for processor
    - Have as a main verb: Have the students who missed the exam take it today
    - Have as an auxiliary verb in an interrogative sentence: Have the students who missed the exam taken it today?

Multiple possible interpretations – Contd.

- Can be handled in four different ways
- All paths
  - Follow all possible paths
  - Build all possible intermediate components
  - Many components will be ignored later
  - Disadvantage: Lot of unnecessary processing
- Best path with backtracking
  - Follow one path
  - Record at every branch, the information necessary to make another choice if chosen path fails to completely interpret the sentence
  - Disadvantages: Lot of information stored (and time) and same component may be analysed many times!

Multiple possible interpretations – Contd.

- Best path with patch up
  - Follow one path at a time
  - When error is detected, explicitly shuffle components that are already formed
  - Usually more efficient than previous two
  - But requires interactions among the rules of grammar to be made explicit in rules for moving components
- Wait and see
  - Follow one path, but don't make decisions about the function of components
  - Wait till enough information to make correct decision is available
  - Very efficient, but buffer should be able to hold the look-ahead information!