Natural Language Information Assurance and Security (NL IAS):

Style Metrics from Semantic Analysis

Student: Lauren M. Stuart. Advisors: Prof. Julia M. Taylor, Prof. Victor Raskin

Stylometry assumes:
- Authors have a style
- that is expressed in their works,
- that can be quantified and measured, and
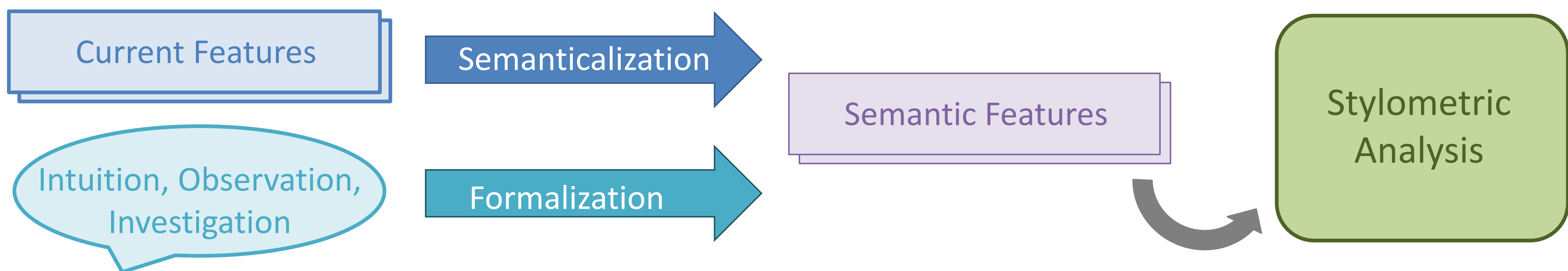- that can be distinguished from that of other authors

Applications of stylometry:
- Author recognition
- Authorship verification
- Stylistic deception
- Stylistic deception detection

Most style metrics focus on surface features of a text. Surface features may capture (or be influenced by) deeper phenomena or other events that may not be style (for example, text file encoding), and may not capture other intuitive notions of style.

Some qualities of author style may be better captured by, or in conjunction with, meaning-based features.
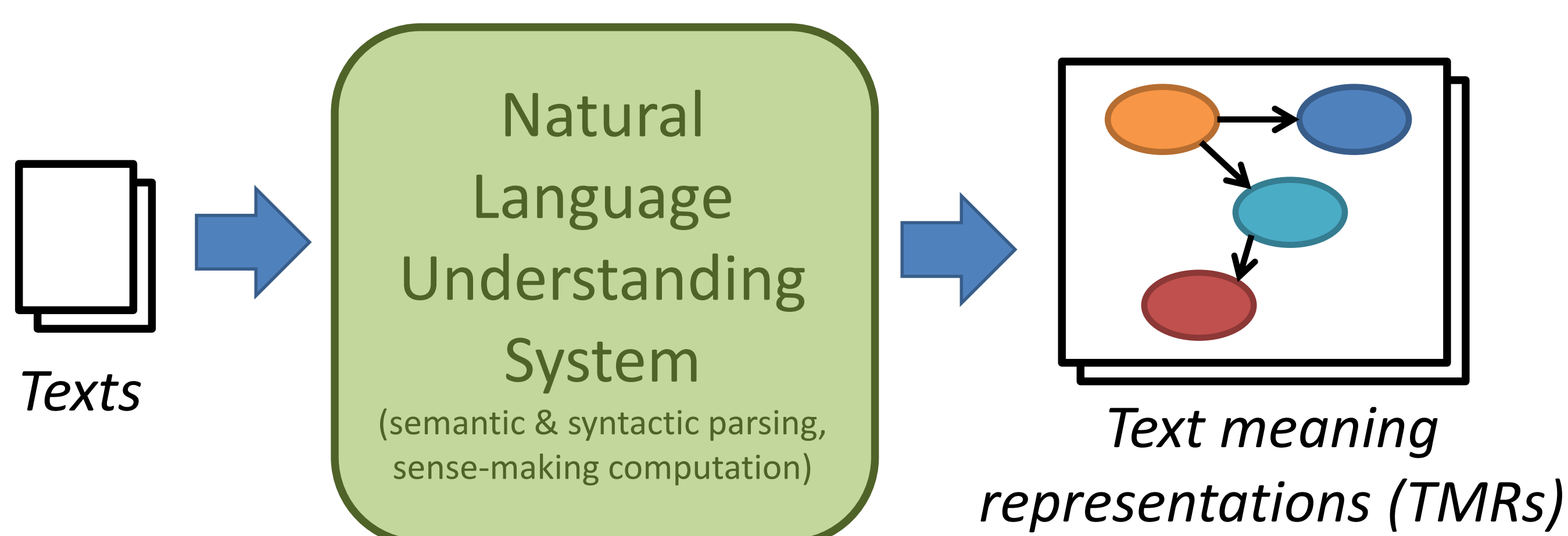
Moving Toward Semantic Stylometry

Current Features → Semanticalization → Semantic Features → Stylometric Analysis

Intuition, Observation, Investigation → Formalization → Semantic Features

Current Features	Examples	Semanticalizations
Function words	of	Relations: instrument, quantity, origin, product-of
	that	Addition/density of relations between objects
Letter trigrams	-ing, -est, pre-	Attributes: tense/aspect Concept names
Words, synonym set choices	book, letter, novel	Concept names: TEXT, BOOK

Current investigation: how do authors differ? What clues do they leave behind, and which of these are stylistic? Some ideas:

Semantic Map Features
Concept-property-filler triples
Tail length / Tree depth
Order of traversal
Concepts, relations per sentence, word

From Text to Semantic Representation

Texts → Natural Language Understanding System (semantic & syntactic parsing, sense-making computation) → Text meaning representations (TMRs)

TMRs are graphs of concepts (nodes) connected by properties (edges) representing the meaning of a text. Properties are directed edges; the nodes at which they terminate are fillers. Some properties are attributes: they have literal value fillers.

A basic unit of analysis: the concept-property-filler triple.

Concept —Property→ Concept (Filler)
Concept —Property→ Value (Filler)