CS 6890: Lecture 7

Vladimir Kulyukin
Department of Computer Science
Utah State University

Outline

- Natural Language Generation under Meaning⇔Text Theory (MTT)

Two Common Methods

- Canned text
- Template filling

NLU and NLG

- Natural language understanding is a process of constructing a meaning representation from a linguistic input.
- Natural language generation is a process of constructing a linguistic output from a meaning representation (non-linguistic input).

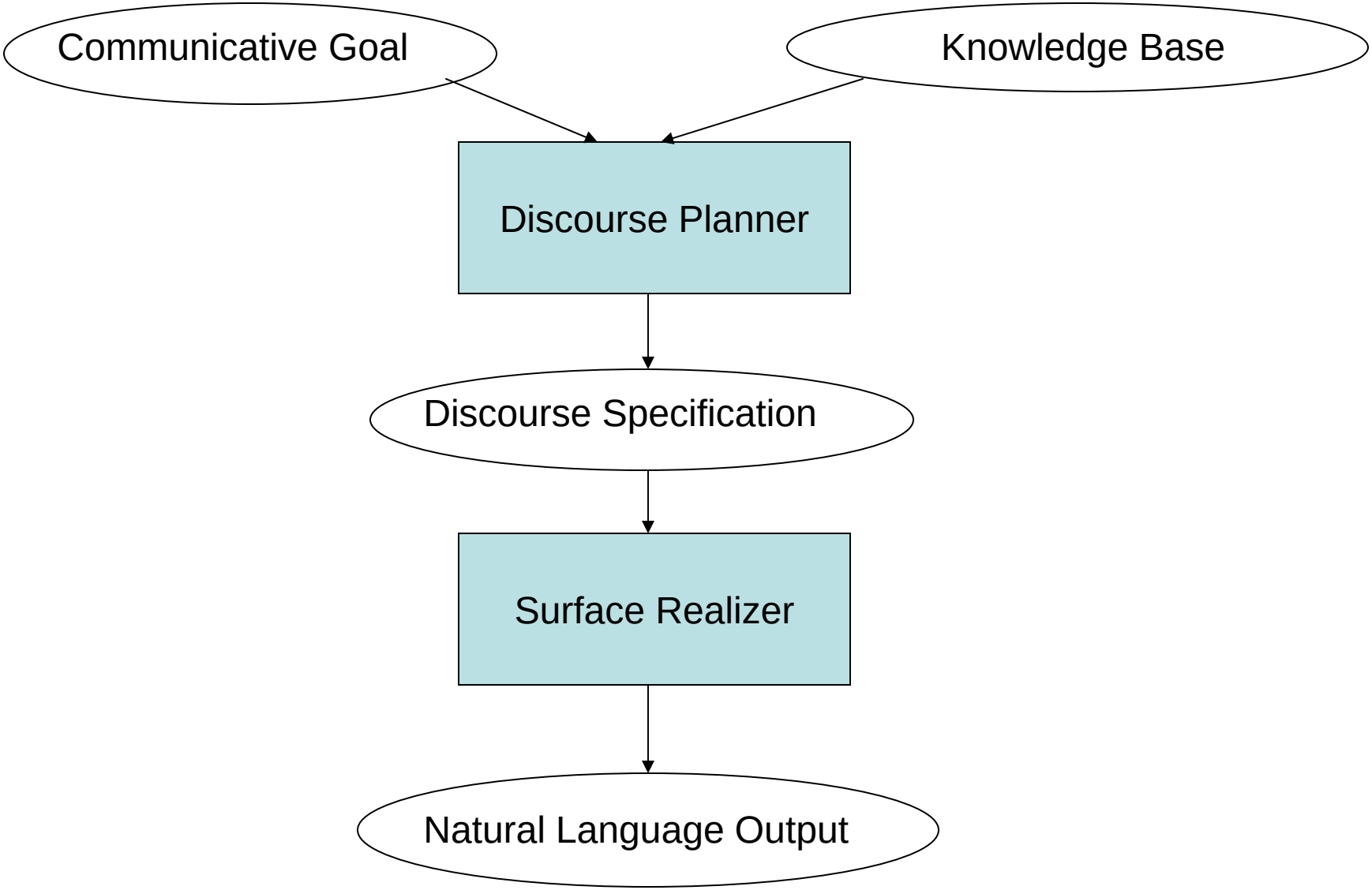
NLU and NLG

- The inputs to NLG systems vary.
- Most input formalisms are domain specific.
- NLU can be seen as a hypothesis management process.
- NLG is a selection process.

NLG as Selection

- Content Selection
  - what to say and when
- Lexical Selection
  - what words to use to express the chosen content
- Sentence Structure
  - Aggregation: the chosen content must be apportioned into phrases, sentences, paragraphs, etc.
- Discourse Structure
  - If the NLG system deals with multi-sentence units, they must be structurally coherent.

An NLG System Architecture



Meaning⇔Text Theory

- MTT is a theoretical framework for the construction of Meaning⇔Text models of natural languages.
- A Meaning⇔Text model (MTM) is a model that can generate a set of texts each of which corresponds to a specific meaning and, vice versa, given a text, map that text to a set of meanings associated with it.

Denumerable Sets

- A set is denumerable if and only if it can be counted.
- More precisely, a set is denumerable if and only if it there is a one-to-one mapping between the elements of the set and the set of natural numbers.
- Any finite set is denumerable.
- For infinite sets, we have to come up with a one-to-one mapping to show that they are denumerable.

Example

- The set of perfect squares is denumerable. Here is a mapping:
- 0 → 1
- 1 → 4
- 2 → 9
- 3 → 16
- …

Example 2

- The set of primes is denumerable.
- Here is a mapping:
- 0 → 2
- 1 → 3
- 2 → 5
- 3 → 7
- 4 → 11
- …

MTT Tenets

Tenet 1

A natural language is a many-to-many mapping between an infinite denumerable set of meanings and an infinite denumerable set of texts.

Tenet 1 Implications

- MTT deals with formal representations of meanings and texts.
- Relation is the basic mechanism of language, i.e. dependency is preferred to constituency.
- Cognitive validity of those representations is of interest but, strictly speaking, is beyond the scope of MTT.

Tenet 1 Formalization

- Let SemR be a formal language that represents meanings of some natural language L.
- Let PhonR be a formal language that represents utterances (phonological sequences) of the same natural language L.
- Let S = {$SemR_i$} be the infinitely denumerable set of meanings.
- Let P = {$PhonR_i$} be the infinitely denumerable set of utterance representations.

Tenet 1 Formalization

- Let f be a one-to-many mapping from S to the power set of P.
- Let g be a one-to-many mapping from P to the power set of S.
- L = <f, g>.
- $f(SemR_i) = \{PhonR_j \mid PhonR_j$ is a phonological representation of $SemR_i\}$.
- $g(PhonR_i) = \{SemR_j \mid SemR_j$ is a semantic representation of $PhonR_i\}$.

Tenet 1 Implications

- Synonymy is a mapping of one meaning representation to more than one text.
- Polysemy is a mapping of one text to more that one meaning representation.

Tenet 1 Implications

- To model a language is to describe how it deals with synonymy, especially paraphrases.
- Meaning is the invariant of paraphrases.
- Lexical choices are made prior to grammatical choices.
- Grammar is but a set of generalizations over a lexicon.

Two Questions about Tenet 1

- There are two questions that one may want to ponder with respect to Tenet 1.
  - Is there a meaning representation that has no corresponding utterance representation, i.e. that cannot be mapped to any text?
  - Is there an utterance representation that has no corresponding meaning representation?

Tenet 2

Viable linguistic descriptions are obtained only if the native speaker's perspective is adopted.

Tenet 2 Implications

- The mappings are described through a formal device called Meaning⇔Text Model (MTM).
- An MTM simulates the performance of a native speaker of L.
- An MTM is a functional construct.

Tenet 2 Implications

- MTT accepts introspection as a valid research method.
- Dictionaries and grammars must describe active not passive knowledge of language.

MTM Representation Levels

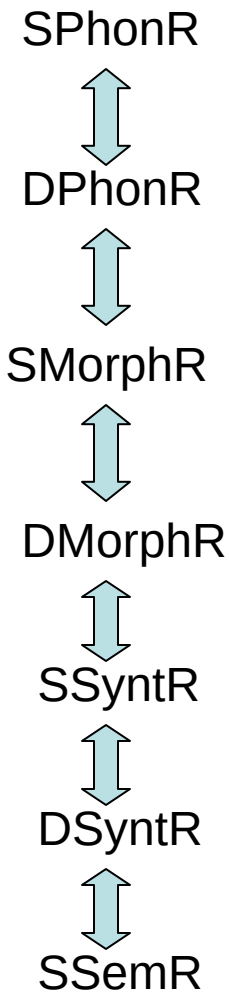
MTM Representation Levels

- Semantics (SymR)
- Deep Syntax (DSyntR)
- Surface Syntax (SSyntR)
- Deep Morphology (DMorphR)
- Surface Morphology (SMorphR)
- Deep Phonology (DPhonR)
- Surface Phonology (SPhonR)

Representation Structures

- Each level has a core representation structure around which the process of synthesis/analysis revolves.
- The semantic level has a *semantic network (a graph)*.
- The syntactic level has a *dependency tree*.
- The morphological level has a *string of linearly ordered word-forms* which make up the utterance.
- The phonological level has a *string of phonemes*.

MTM Representation Levels



Two Types of Rules in MTM

- Correspondence rules state correspondences b/w fragments of representations of two adjacent levels:
  - X level n ⇔Y level n+1 | Conditions
- Equivalence rules state equivalences b/w representations of the same level:
  - X level n = Y level n | Conditions

SemR ➔ DSyntR

SemR ➔ DSyntR

- MTT does not address the World ⇔ SemR question.
- There are four constituents of SemR:
- Semantic Structure (SemS)
- Semantic-Communicative Structure (Sem-CommS)
- Rhetorical Structure (RhetS)
- Referential Structure (RefS)

What is Meaning?

- Meaning is the invariant of paraphrases.
- *To know one's stuff* means the same as *to know one's job well*.
- *To call the shots* is the same as *to be in charge*.
- Basic insight: for a native speaker it is a lot easier to say that an sentence S_1 means the same as a sentence S_2 than to describe the meaning of S_1.

What is Meaning?

- Exact synonymy is rare.
- Most of the time, E1 and E2 are approximately synonymous.
- In ordinary communication, many subtle differences are ignored.

Semantic Structure: SemS

- Semantemes are lexical meanings of a particular language L.
- SemS is a graph whose nodes are semantemes and whose arcs are labeled with numbers that indicate arguments (semantic actants) of predicate-argument relations.
- Example: The meaning of criticize is a three-place predicate ‘X criticizes Y for Z’.

SemS

- A seme is an elementary unit of meaning.
- A semanteme is a set of semes.
- The process of decomposing a semanteme into semes is called semantic decomposition.
- Semantic decomposition is a very complex knowledge engineering effort.
- Semantemes are denoted by singular quotes, e.g. ‘criticize’.

Communicative Structure: Sem-CommS

- Sem-CommS specifies which part of the SemS is the Theme (what is being talked about) and which is the Rheme (what is communicated about the Theme)
- The book is on the table.
- Theme = the book.
- Rheme = is on the table.

Rhetorical Structure: RhetS

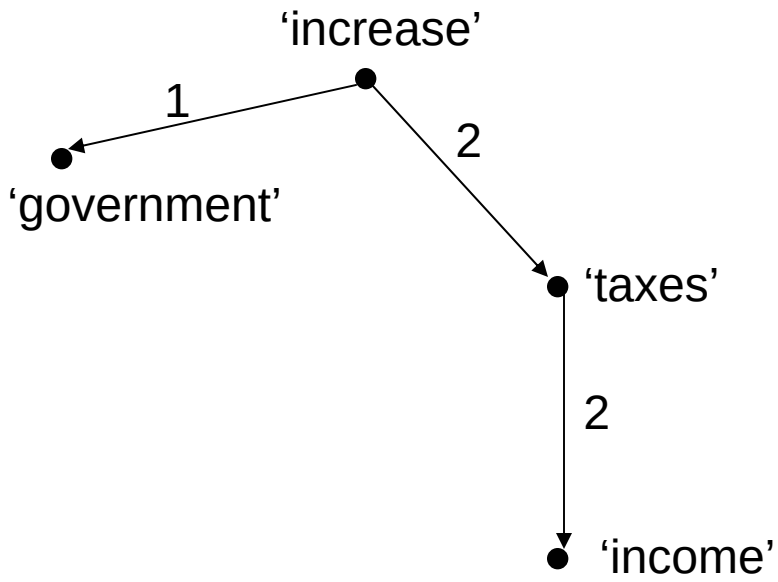
- RhetS specifies the style of SemS: neutral, official, colloquial, ironic, poetic, etc.
- Sem-CommS and RhetS determine the paraphrasing power of the SemS.
- Speaking is a series of choices b/w equivalent ways of expressing a certain meaning.

Extended Example

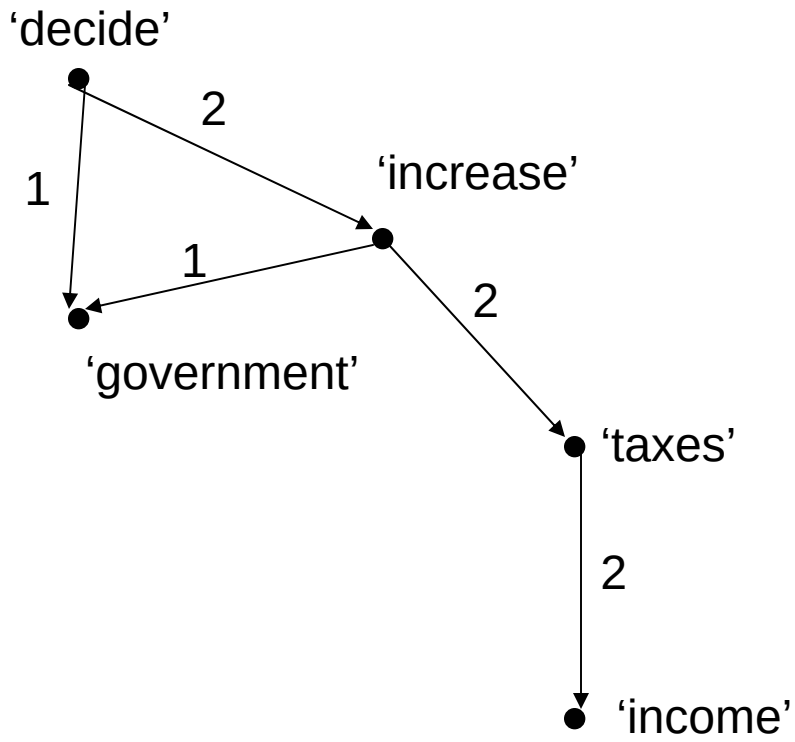
Sample Situation

- Decision: the government of some country X decided to increase income taxes.
- Criticism: the media of X intensely criticized the government for its decision to increase income taxes.

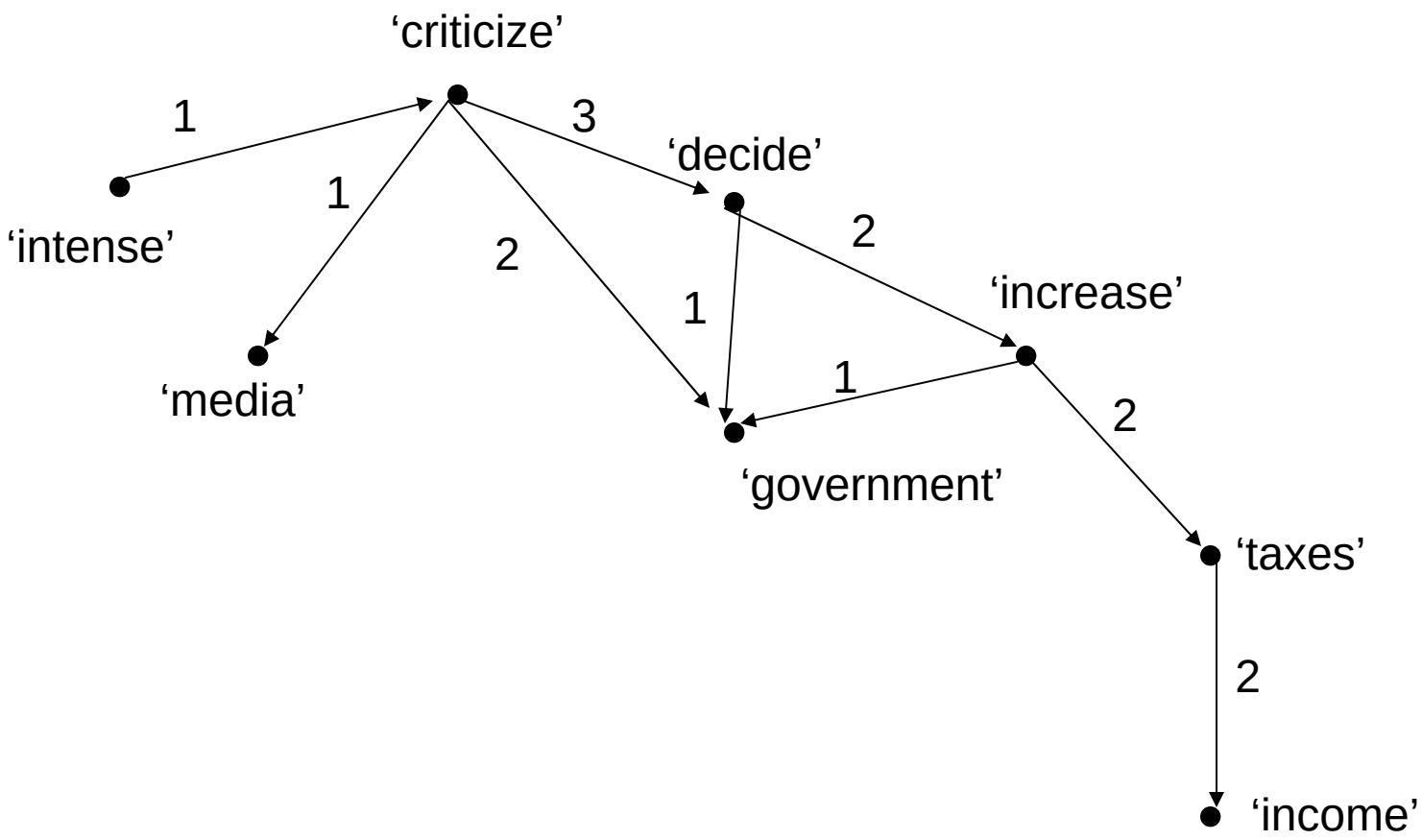
SemS



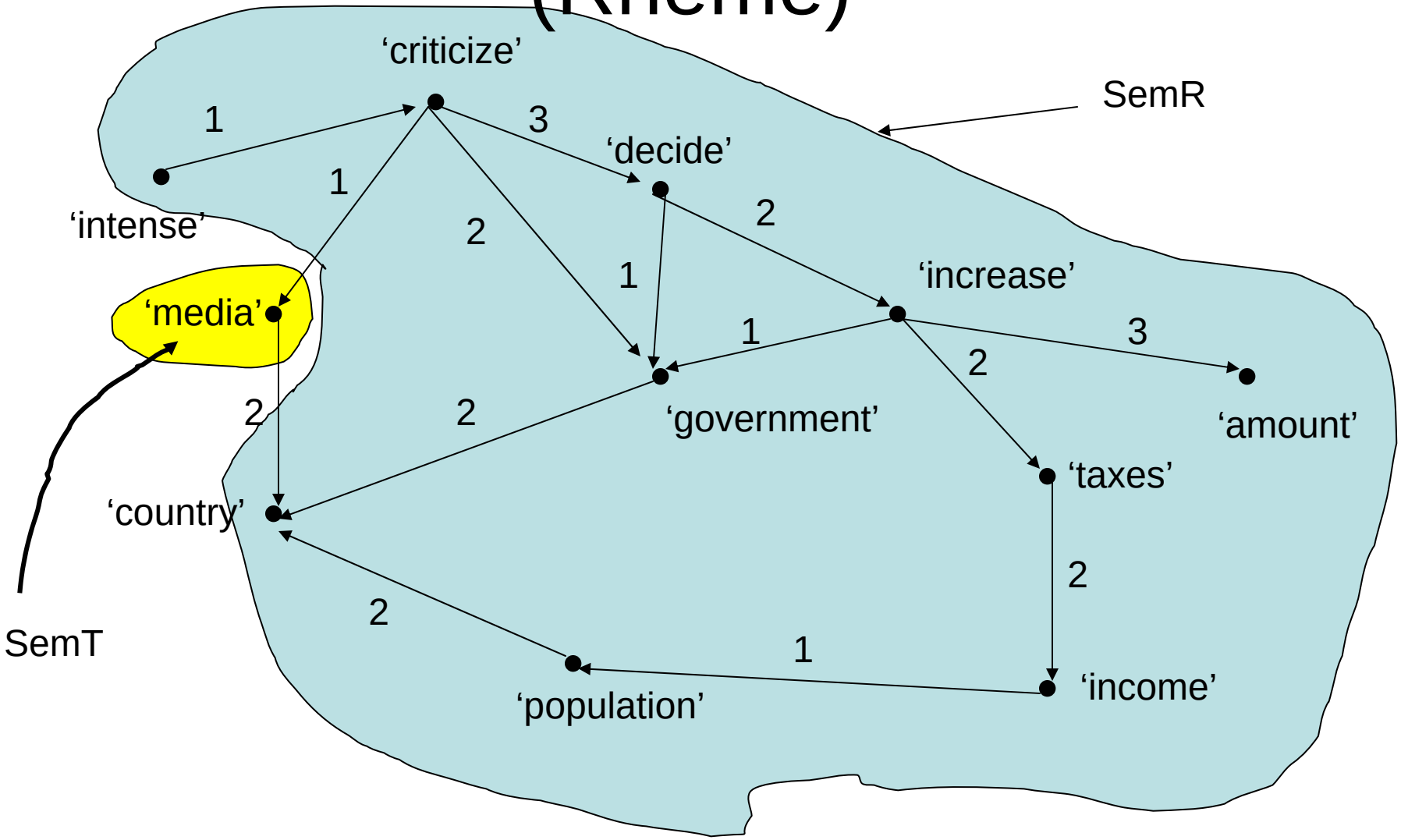
SemS



SemS


‘criticize’
1
3
‘decide’
‘intense’
1
2
2
‘increase’
1
‘media’
1
2
‘government’
‘taxes’
2
‘income’

SemS: SemT (Theme) and SemR (Rheme)



References

Jasmina Milicevic. A Short Guide to the Meaning-Text Linguistic Theory. Journal of Koralex, vol. 8, pp. 187-233, 2006.