Logical Metalanguage for Linguistic Description

Hossep Dolatian

Stony Brook University

August 10, 2018

Motivation

- **Background**: Linguistics is the scientific description of language
- **Problem**: What metalanguage do you use to describe language?
- **Solution**: Develop a computational+logical metalanguage for linguistic description
- **Illustration:** Use it to describe Armenian
- **Implementation:** Develop software to run logical formula for linguistics (phonology & morphology)

What's linguistics for

- What do linguists do?

What's linguistics for

- What do linguists do?
  - Describe

What's linguistics for

- What do linguists do?
  - Describe
  - Analyze

What's linguistics for

- What do linguists do?
  - Describe
  - Analyze
  - Share
- Like with who?

What's linguistics for

- What do linguists do?
  - Describe
  - Analyze
  - Share
- Like with who?
  - Other language-specialists

What's linguistics for

- What do linguists do?
  - Describe
  - Analyze
  - Share
- Like with who?
  - Other language-specialists
  - Theoretical linguists

What's linguistics for

- What do linguists do?
  - Describe
  - Analyze
  - Share
- Like with who?
  - Other language-specialists
  - Theoretical linguists
  - Computational linguists
    1. Develop computational resources
    2. Industrial NLP applications

What's the problem?

- What do you need to share knowledge?
  - The right **metalanguage** which is:

1. **accurate**
2. **precise**
3. **human-readable**
4. **stable shelf-life**
5. **flexible**
6. **implementable**

What's the problem?

- What do you need to share knowledge?
  - The right **metalanguage** which is:

1. **accurate**
2. **precise**
3. **human-readable**
4. **stable shelf-life**
5. **flexible**
6. **implementable**

- But linguists often use prose+formalisms that don't fit all 6 needs

A linguistic problem in Armenian

- What's Armenian?
  - Indo-European language spoken in the Caucuses & Middle East
  - Under-studied & low-resource
  - Focus on morphology and phonology
- What's going in Armenian?
  - Orthography shows long sequences of consonants

Cons.	2	2	3	4	4	6
Meaning	'money'	'low'	'grand'	'to jabber'	'to find'	'to grumble'
Spelling	tram	stor	hsga	pndrel	tntrel	krtmnjal

A linguistic problem in Armenian

- What's Armenian?
  - Indo-European language spoken in the Caucuses & Middle East
  - Under-studied & low-resource
  - Focus on morphology and phonology
- What's going in Armenian?
  - Orthography shows long sequences of consonants

Cons.	2	2	3	4	4	6
Meaning	'money'	'low'	'grand'	'to jabber'	'to find'	'to grumble'
Spelling	tram	stor	hsga	pndrel	tntrel	krtmnjal
Pron.	təram	əstor	həsga	pəndrel	təntərel	kərtmənjal

  - But all are broken up by the vowel /ə/ 'puppet'

A linguistic problem in Armenian

- Research question:
  - Given a spelled word, how do you pronounce it? (=Text-to-Speech problem)

A linguistic problem in Armenian

- Research question:
  - Given a spelled word, how do you pronounce it? (=Text-to-Speech problem)
- Research Answer?

A linguistic problem in Armenian

- Research question:
  - Given a spelled word, how do you pronounce it? (=Text-to-Speech problem)
- Research Answer?
  - Position of the vowel is predictable
  - Depends on #, type, and position of consonants

A linguistic problem in Armenian

- Two patterns:

1. Repeated Cs:

Meaning	'to be astonished'	'to jabber'	'to find'
Spelling	prpral	tntrel	pndrel

A linguistic problem in Armenian

- Two patterns:

1. Repeated Cs:

Meaning	'to be astonished'	'to jabber'	'to find'
Spelling	prpral	tntrel	pndrel
Pron.	pərpəral	təntərel	pəndrel

2. sC:

Meaning	'money'	'low'
Spelling	tram	stor

A linguistic problem in Armenian

- Two patterns:

1. Repeated Cs:

Meaning	'to be astonished'	'to jabber'	'to find'
Spelling	prpral	tntrel	pndrel
Pron.	pərpəral	təntərel	pəndrel

2. sC:

Meaning	'money'	'low'
Spelling	tram	stor
Pron.	təram	əstor

A linguistic problem in Armenian

- What have linguists done with this problem?

A linguistic problem in Armenian

- What have linguists done with this problem?
  - Theoretical linguistics:
    - Vaux 1998: Most complete description with a sketch of an algorithm
  - Computational linguistics:
    - Me: Implemented algorithm in Python
  - His knowledge+my implementation cover a lot of ground but...

A linguistic problem in Armenian

- Problem: The original source used prose

A linguistic problem in Armenian

- Problem: The original source used prose
- Effect: This caused problems with:

A linguistic problem in Armenian

- Problem: The original source used prose
- Effect: This caused problems with:

1. & 2. Accuracy & precision: vagueness in how to order certain processes

Meaning	'to grumble'	'to jabber'	'to find'
Spelling	krtmnjal	tntrel	pndrel

A linguistic problem in Armenian

- Problem: The original source used prose
- Effect: This caused problems with:

1. & 2. Accuracy & precision: vagueness in how to order certain processes

Meaning	'to grumble'	'to jabber'	'to find'
Spelling	krtmnjal	tntrel	pndrel
Process A→B	kərtmənjal ✓	təntərel ✓	pəndərel ✗
Process B→A	kərtmənjal ✓	təntrel ✗	pəndrel ✓

  - Source didn't "explicitly" say what's the right order for A & B

A linguistic problem in Armenian

- Problem: The original source used prose
- Effect: This caused problems with:

1. & 2. Accuracy & precision: vagueness in how to order certain processes

Meaning	'to grumble'	'to jabber'	'to find'
Spelling	krtmnjal	tntrel	pndrel
Process A→B	kərtmənjal ✓	təntərel ✓	pəndərel ✗
Process B→A	kərtmənjal ✓	təntrel ✗	pəndrel ✓

  - Source didn't "explicitly" say what's the right order for A & B

- Problem trickled down to other properties

A linguistic problem in Armenian

- Problem: The original source used prose
- Effect: This caused problems with:

1. & 2. Accuracy & precision: vagueness in how to order certain processes

Meaning	'to grumble'	'to jabber'	'to find'
Spelling	krtmnjal	tntrel	pndrel
Process A→B	kərtmənjal ✓	təntərel ✓	pəndərel ✗
Process B→A	kərtmənjal ✓	təntrel ✗	pəndrel ✓

  - Source didn't "explicitly" say what's the right order for A & B

- Problem trickled down to other properties

3. Human-readable?
4. Long shelf-life?
5. Flexible?
6. Implementable?

A linguistic problem in Armenian

- Problem: The original source used prose
- Effect: This caused problems with:

1. & 2. Accuracy & precision: vagueness in how to order certain processes

Meaning	'to grumble'	'to jabber'	'to find'
Spelling	krtmnjal	tntrel	pndrel
Process A→B	kərtmənjal ✓	təntərel ✓	pəndərel ✗
Process B→A	kərtmənjal ✓	təntrel ✗	pəndrel ✓

  - Source didn't "explicitly" say what's the right order for A & B

- Problem trickled down to other properties

3. Human-readable?
4. Long shelf-life?
5. Flexible?
6. Implementable?

→ Original resource had good research quality but metalanguage problems

Solution: logical metalanguage

- Problem in a nutshell

Solution: logical metalanguage

- Problem in a nutshell
  - Research is about sharing knowledge but that needs an efficient metalanguage

Solution: logical metalanguage

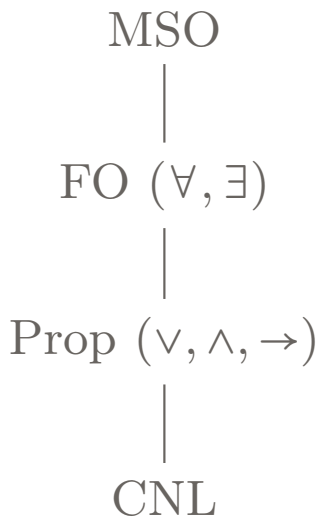
- Problem in a nutshell
  - Research is about sharing knowledge but that needs an efficient metalanguage
- Solution: Formal logic

Solution: logical metalanguage

- Problem in a nutshell
  - Research is about sharing knowledge but that needs an efficient metalanguage
- Solution: Formal logic
  - **Synthesize** work from theoretical linguistics, formal logic, and computational linguistics.
  - **Adapt** formal logic as a metalanguage for linguistics
  - **Develop** working software to convert formal logic into running code for linguistics

Formal logic

- What's logic?
  - Logic is the metalanguage used for describing complex patterns in mathematics (Courcelle, 1997; Engelfriet and Hoogeboom, 2001)
  - There are different types of logical languages with different degrees of power



Why logic

- Benefits of logic?
  1. Accurate
  2. Precise
  3. Human-readable
  4. Long shelf-life
  5. Flexible
  6. Implementable
- Why not something else like FSM?

Why logic

- Benefits of logic?
  1. Accurate
  2. Precise
  3. Human-readable
  4. Long shelf-life
  5. Flexible
  6. Implementable
- Why not something else like FSM?
  - Easy to implement but not very human readable for complex generalizations

Why logic

- Benefits of logic?
  1. Accurate
  2. Precise
  3. Human-readable
  4. Long shelf-life
  5. Flexible
  6. Implementable
- Why not something else like FSM?
  - Easy to implement but not very human readable for complex generalizations
  - Logic is higher-level language
  - Logic can be converted to FSMs to get its benefits

Logic in linguistics

- How has logic been used in linguistics?

Logic in linguistics

- How has logic been used in linguistics?
  - Has been applied to sentence structure (Rogers, 1998)
  - Heinz & colleagues have partially adapted it for phonology
  - Looks like FO can handle bulk of phonology & morphology.
- What will I do?

Logic in linguistics

- How has logic been used in linguistics?
  - Has been applied to sentence structure (Rogers, 1998)
  - Heinz & colleagues have partially adapted it for phonology
  - Looks like FO can handle bulk of phonology & morphology.
- What will I do?
  - Describe a language's phonology with logic
  - Develop software to convert logic into code for linguistics (Python, Prolog, Haskell)

Armenian illustration

- Subpattern in pronouncing 4 consonants: Repeated Cs

Meaning	'to be astonished'	'to jabber'	'to find'
Spelling	prpral	tntrel	pndrel
Pron.	pərpəral	təntərel	pəndrel

- Add /ə/ after 1st and 3rd if they're the same consonant; otherwise after just the 1st

Armenian illustration

- How should you pronounce *ksksel*?

Armenian illustration

- How should you pronounce *ksksel*?
- Look at *ksksel* as an array-like graph

k		s		k		s		e		l
cons		cons		cons		cons		vow		cons
0	◁	1	◁	2	◁	3	◁	4	◁	5

Armenian illustration

- How should you pronounce *ksksel*?
- Look at *ksksel* as an array-like graph

k	s	k	s	e	l
cons	cons	cons	cons	vow	cons
0 ◁	1 ◁	2 ◁	3 ◁	4 ◁	5

- Conditions can be formalized with

Armenian illustration

- How should you pronounce *ksksel*?
- Look at *ksksel* as an array-like graph

k		s		k		s		e		l
cons		cons		cons		cons		vow		cons
0	$\lhd$	1	$\lhd$	2	$\lhd$	3	$\lhd$	4	$\lhd$	5

- Conditions can be formalized with

$$\texttt{first}(y) \stackrel{\text{def}}{=} \neg\exists x[x \lhd y] \tag{1}$$

Armenian illustration

- How should you pronounce *ksksel*?
- Look at *ksksel* as an array-like graph

k		s		k		s		e		l
cons		cons		cons		cons		vow		cons
0	$\lhd$	1	$\lhd$	2	$\lhd$	3	$\lhd$	4	$\lhd$	5

- Conditions can be formalized with

$$\texttt{first}(y) \stackrel{\text{def}}{=} \neg\exists x[x \lhd y] \tag{1}$$

$$\texttt{third}(z) \stackrel{\text{def}}{=} \exists x, y[\texttt{x} \wedge x \lhd y \wedge y \lhd z] \tag{2}$$

Armenian illustration

- How should you pronounce *ksksel*?
- Look at *ksksel* as an array-like graph

k		s		k		s		e		l
cons		cons		cons		cons		vow		cons
0	$\lhd$	1	$\lhd$	2	$\lhd$	3	$\lhd$	4	$\lhd$	5

- Conditions can be formalized with

$$\texttt{first}(y) \stackrel{\text{def}}{=} \neg\exists x[x \lhd y] \quad (1)$$

$$\texttt{third}(z) \stackrel{\text{def}}{=} \exists x, y[\texttt{x} \wedge x \lhd y \wedge y \lhd z] \quad (2)$$

$$\texttt{same}(x, y) \stackrel{\text{def}}{=} (\texttt{p}(x) \wedge \texttt{p}(y)) \vee (\texttt{t}(x) \wedge \texttt{t}(y)) \vee \ldots \quad (3)$$

Armenian illustration

- How should you pronounce *ksksel*?
- Look at *ksksel* as an array-like graph

k	s	k	s	e	l
cons	cons	cons	cons	vow	cons
0	1	2	3	4	5

- Conditions can be formalized with

$$\texttt{first}(y) \stackrel{\text{def}}{=} \neg\exists x[x \lhd y] \quad (1)$$

$$\texttt{third}(z) \stackrel{\text{def}}{=} \exists x, y[\texttt{x} \wedge x \lhd y \wedge y \lhd z] \quad (2)$$

$$\texttt{same}(x, y) \stackrel{\text{def}}{=} (\texttt{p}(x) \wedge \texttt{p}(y)) \vee (\texttt{t}(x) \wedge \texttt{t}(y)) \vee \ldots \quad (3)$$

$$\texttt{RepeatedCs}(x, z) \stackrel{\text{def}}{=} \texttt{first}(x) \wedge \texttt{third}(z) \wedge \texttt{same}(x, z) \quad (4)$$

- Implement code in Python to check for these conditions or apply them

Plan for the year

- **Goal**: Apply formal logic as metalanguage for linguistics with Armenian as a case study
- **Output**: Develop software (Python, Prolog, Haskell) to run logical formula designed for phonology and morphology
- Timeline

Plan for the year

- **Goal**: Apply formal logic as metalanguage for linguistics with Armenian as a case study
- **Output**: Develop software (Python, Prolog, Haskell) to run logical formula designed for phonology and morphology
- Timeline
  1. **Sept-Jan**: Data collection and formalization into logic

Plan for the year

- **Goal**: Apply formal logic as metalanguage for linguistics with Armenian as a case study
- **Output**: Develop software (Python, Prolog, Haskell) to run logical formula designed for phonology and morphology
- Timeline
  1. **Sept-Jan**: Data collection and formalization into logic
  2. **Feb-June**: Develop software to implement (linguistic) logic in Python or Prolog

Plan for the year

- **Goal**: Apply formal logic as metalanguage for linguistics with Armenian as a case study
- **Output**: Develop software (Python, Prolog, Haskell) to run logical formula designed for phonology and morphology
- Timeline
  1. **Sept-Jan**: Data collection and formalization into logic
  2. **Feb-June**: Develop software to implement (linguistic) logic in Python or Prolog
- Future work
  1. Collect more data on Armenian and related languages
  2. Improve metalanguage for phonology, morphology, and other linguistic fields
  3. Given proper metalanguage, develop more computational resources for Armenian
  4. Develop instructional material on how to use and implement logic for linguistics

Courcelle, B. (1997). The expression of graph properties and graph transformations in monadic second-order logic. In G. Rozenberg (Ed.), Handbook of Graph Grammars and Computing by Graph Transformations, pp. 313–400.

Engelfriet, J. and H. J. Hoogeboom (2001, April). MSO definable string transductions and two-way finite-state transducers. ACM Trans. Comput. Logic 2(2), 216–254.

Rogers, J. (1998). A Descriptive Approach to Language-Theoretic Complexity. Stanford, CA: CSLI Publications.