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

<table>
<thead>
<tr>
<th>Cons.</th>
<th>Meaning</th>
<th>Spelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>‘money’</td>
<td>tram</td>
</tr>
<tr>
<td>2</td>
<td>‘low’</td>
<td>stor</td>
</tr>
<tr>
<td>3</td>
<td>‘grand’</td>
<td>hsga</td>
</tr>
<tr>
<td>4</td>
<td>‘to jabber’</td>
<td>pndrel</td>
</tr>
<tr>
<td>4</td>
<td>‘to find’</td>
<td>tntrel</td>
</tr>
<tr>
<td>6</td>
<td>‘to grumble’</td>
<td>krtmnjal</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Cons.</th>
<th>Meaning</th>
<th>Spelling</th>
<th>Pron.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>‘money’</td>
<td>tram</td>
<td>təram</td>
</tr>
<tr>
<td>2</td>
<td>‘low’</td>
<td>stor</td>
<td>østor</td>
</tr>
<tr>
<td>3</td>
<td>‘grand’</td>
<td>hsga</td>
<td>həsga</td>
</tr>
<tr>
<td>4</td>
<td>‘to jabber’</td>
<td>pndrel</td>
<td>pəndrel</td>
</tr>
<tr>
<td>4</td>
<td>‘to find’</td>
<td>tntrel</td>
<td>təntərel</td>
</tr>
<tr>
<td>6</td>
<td>‘to grumble’</td>
<td>krtmŋnal</td>
<td>kərtməŋnal</td>
</tr>
</tbody>
</table>

- 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
     - Pronunciation: pərpəral təntərel pəndrel
  2. sC:
     - Meaning: ‘money’ ‘low’
     - Spelling: tram stor
     - Pronunciation: təram østor
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 plus 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

<table>
<thead>
<tr>
<th>Meaning</th>
<th>‘to grumble’</th>
<th>‘to jabber’</th>
<th>‘to find’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spelling</td>
<td>krtmnjal</td>
<td>tntrel</td>
<td>pndrel</td>
</tr>
<tr>
<td>Process A→B</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Process B→A</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Meaning</th>
<th>‘to grumble’</th>
<th>‘to jabber’</th>
<th>‘to find’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spelling</td>
<td>krtmŋjal</td>
<td>tntrel</td>
<td>pndrel</td>
</tr>
<tr>
<td>Process A→B</td>
<td>kərtməŋjal ✓</td>
<td>təntərel ✓</td>
<td>pəndərel ✗</td>
</tr>
<tr>
<td>Process B→A</td>
<td>kərtməŋjal ✓</td>
<td>təntrel ✗</td>
<td>pəndrel ✓</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Meaning</th>
<th>‘to grumble’</th>
<th>‘to jabber’</th>
<th>‘to find’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spelling</td>
<td>krtmŋnal</td>
<td>tntrel</td>
<td>pndrel</td>
</tr>
<tr>
<td>Process A→B</td>
<td>kərtmæŋnal ✓</td>
<td>təntərel ✓</td>
<td>pəndərel ✓</td>
</tr>
<tr>
<td>Process B→A</td>
<td>kərtmæŋnal ✓</td>
<td>təntrel ✗</td>
<td>pəndrel ✓</td>
</tr>
</tbody>
</table>

- 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

<table>
<thead>
<tr>
<th>Meaning</th>
<th>‘to grumble’</th>
<th>‘to jabber’</th>
<th>‘to find’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spelling</td>
<td>krtmnjral</td>
<td>tntrel</td>
<td>pndrel</td>
</tr>
<tr>
<td>Process A→B</td>
<td>kərtməŋjral ✓</td>
<td>təntərel ✓</td>
<td>pəndərel ×</td>
</tr>
<tr>
<td>Process B→A</td>
<td>kərtməŋjral ✓</td>
<td>təntrel ×</td>
<td>pəndrel ✓</td>
</tr>
</tbody>
</table>

- 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

- What is Formal Logic
- Applying Formal Logic to linguistics
- Illustration with Armenian

<table>
<thead>
<tr>
<th>MSO</th>
</tr>
</thead>
<tbody>
<tr>
<td>FO (∀, ∃)</td>
</tr>
<tr>
<td>Prop (∨, ∧, →)</td>
</tr>
<tr>
<td>CNL</td>
</tr>
</tbody>
</table>
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
  
<table>
<thead>
<tr>
<th>Meaning</th>
<th>‘to be astonished’</th>
<th>‘to jabber’</th>
<th>‘to find’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spelling</td>
<td>prpral</td>
<td>tntrel</td>
<td>pndrel</td>
</tr>
<tr>
<td>Pron.</td>
<td>pərpəral</td>
<td>təntərel</td>
<td>pəndrel</td>
</tr>
</tbody>
</table>

- Add /ə/ after 1\textsuperscript{st} and 3\textsuperscript{rd} if they’re the same consonant; otherwise after just the 1\textsuperscript{st}
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 \textit{ksksel}?
- Look at \textit{ksksel} as an array-like graph

\begin{center}
\begin{tabular}{ccccccc}
  k & s & k & s & e & l \\
  cons & cons & cons & cons & vow & cons \\
  0 & $\triangle$ & 1 & $\triangle$ & 2 & $\triangle$ & 3 & $\triangle$ & 4 & $\triangle$ & 5
\end{tabular}
\end{center}

- Conditions can be formalized with

\begin{equation}
\text{first}(y) \overset{\text{def}}{=} \neg \exists x [x \triangle y]
\end{equation}
ARMENIAN ILLUSTRATION

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

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>s</td>
<td>k</td>
<td>s</td>
<td>e</td>
<td>l</td>
</tr>
<tr>
<td>cons</td>
<td>cons</td>
<td>cons</td>
<td>cons</td>
<td>vow</td>
<td>cons</td>
</tr>
<tr>
<td>0 &lt; 1 &lt; 2 &lt; 3 &lt; 4 &lt; 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Conditions can be formalized with

\[
\text{first}(y) \overset{\text{def}}{=} \neg \exists x [x \triangleleft y] \\
\text{third}(z) \overset{\text{def}}{=} \exists x, y [x \land x \triangleleft y \land y \triangleleft z]
\]
ARMENIAN ILLUSTRATION

- How should you pronounce \textit{ksksel}?
- Look at \textit{ksksel} as an array-like graph

\begin{verbatim}
  k  s  k  s  e  l
cons cons cons cons vow cons
0 \textless 1 \textless 2 \textless 3 \textless 4 \textless 5
\end{verbatim}

- Conditions can be formalized with

\[
\text{first}(y) \overset{\text{def}}{=} \neg \exists x[x \textless y] \quad (1)
\]
\[
\text{third}(z) \overset{\text{def}}{=} \exists x, y[x \land x \textless y \land y \textless z] \quad (2)
\]
\[
\text{same}(x, y) \overset{\text{def}}{=} (p(x) \land p(y)) \lor (t(x) \land t(y)) \lor \ldots \quad (3)
\]
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

\[
\text{first}(y) \overset{\text{def}}{=} \neg \exists x [x \triangleleft y] \quad (1) \\
\text{third}(z) \overset{\text{def}}{=} \exists x, y [x \land x \triangleleft y \land y \triangleleft z] \quad (2) \\
\text{same}(x, y) \overset{\text{def}}{=} (p(x) \land p(y)) \lor (t(x) \land t(y)) \lor \ldots \quad (3) \\
\text{RepeatedCs}(x, z) \overset{\text{def}}{=} \text{first}(x) \land \text{third}(z) \land \text{same}(x, z) \quad (4)
\]

- 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

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
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
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
