AMR-to-Text Generation via GF

Normunds Grūzītis

University of Latvia, Institute of Mathematics and Computer Science
National information agency LETA

GF Summer School 2017, Rīga, Latvia

This work has received funding in part from the Latvian State research programs SOPHIS and NexIT, the EU Horizon 2020 project SUMMA (grant No. 688139), and the European Regional Development Fund (grant No. 1.1.1.1/16/A/219).
Agenda

• Frame semantics
  – FrameNet
  – PropBank

• AMR

• Text-to-AMR parsing, AMR-to-text generation
  – SemEval 2016
  – SemEval 2017
A fourth member, Jean-Marc Rouillan, remains behind bars.

FrameNet

TurboParser + SEMAFOR: http://demo.ark.cs.cmu.edu/parse

PropBank

LTH parser: http://barbar.cs.lth.se:8081/
FrameNet (https://framenet.icsi.berkeley.edu)

Remainder

Definition:

A **Resource** is depleted by some process, resulting in there being only a **Remainder** in existence some time into or after the process. The original **Resource** may be indicated metonymically via reference to an **Original_owner**. The **Location** where the **Resource** is found may be expressed.

There are only **two poems** **LEFT** from Quayle

After the storm, **only two houses** were **LEFT** **DNI**

**Not much** was **LEFT** **in the till**

**Nothing REMAINED** for the youngest brother **from the inheritance**

**Only this school** is **LEFT** **from the 1970s**; all the other ones are newer.

<table>
<thead>
<tr>
<th>Lexical Unit</th>
<th>LU Status</th>
<th>Lexical Entry Report</th>
<th>Annotation Report</th>
<th>Annotator ID</th>
<th>Created Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>left.a</td>
<td>Created</td>
<td>Lexical entry</td>
<td>Annotation</td>
<td>JKR</td>
<td>03/16/2004 12:23:35 PST Tue</td>
</tr>
<tr>
<td>remain.v</td>
<td>Finished_Initial</td>
<td>Lexical entry</td>
<td>Annotation</td>
<td>JKR</td>
<td>03/16/2004 12:24:22 PST Tue</td>
</tr>
<tr>
<td>remains.n</td>
<td>Created</td>
<td>Lexical entry</td>
<td>Annotation</td>
<td>KmG</td>
<td>03/14/2007 03:20:12 PDT Wed</td>
</tr>
<tr>
<td>remnant.n</td>
<td>Finished_Initial</td>
<td>Lexical entry</td>
<td>Annotation</td>
<td>JKR</td>
<td>03/16/2004 12:21:10 PST Tue</td>
</tr>
</tbody>
</table>
FrameNet (https://framenet.icsi.berkeley.edu)

State_continue

Definition:

Despite some implication that a State would be interrupted, the Entity remains in the specified State. Note that State includes locative relations. Online security remains elusive.

FEs:

Core:

Entity [ent] A concrete or abstract Entity.

Core Unexpressed:


<table>
<thead>
<tr>
<th>remain.v</th>
<th>Finished_Initial</th>
<th>Lexical entry</th>
<th>Annotation</th>
<th>JKR</th>
<th>06/30/2004 10:30:48 PDT Wed</th>
</tr>
</thead>
<tbody>
<tr>
<td>rest.v</td>
<td>Created</td>
<td>Lexical entry</td>
<td>Annotation</td>
<td>RLG</td>
<td>08/21/2006 02:26:58 PDT Mon</td>
</tr>
<tr>
<td>stay.v</td>
<td>Created</td>
<td>Lexical entry</td>
<td>Annotation</td>
<td>MJE</td>
<td>08/19/2004 05:38:49 PDT Thu</td>
</tr>
</tbody>
</table>
FrameNet (FN)

- A lexico-semantic resource based on the theory of frame semantics (Fillmore et al. 2003)
  - A semantic frame represents a cognitive, prototypical situation (scenario) characterized by frame elements (FE) – semantic valence
  - Frames are “evoked” in sentences by target words – lexical units (LU)
  - FEs are mapped based on the syntactic valence of the LU
    - The syntactic valence patterns are derived from FN-annotated corpora (for an increasing number of languages, incl. Latvian)
  - FEs are split into core and non-core ones
    - Core FEs uniquely characterize the frame and syntactically tend to correspond to verb arguments
    - Non-core FEs are not specific to the frame and typically are adjuncts
Berkeley FrameNet as Interlingua

| Definition: | An EXPERIENCER desires that an EVENT occur. In some cases, the EXPERIENCER is an active participant in the EVENT, and in such cases the EVENT itself is often not mentioned, but rather some FOCAL_PARTICIPANT which is subordinately involved. |
| Core FEs: | EVENT, EXPERIENCER, FOCAL_PARTICIPANT, LOCATION_OF_EVENT |
| Non-core FEs: | CAUSE, DEGREE, DURATION, MANNER, PLACE, PURPOSE_OF_EVENT, REASON, ROLE_OF_FOCAL_PARTICIPANT, TIME, TIME_OF_EVENT |

Introduced in BFN, reused in SweFN

| Examples | Valence patterns |
| 40 | EVENT EXPERIENCER |
| 14 | EXPERIENCER FOCAL_PARTICIPANT |

Some valence patterns found in BFN

| Examples | Valence patterns |
| 1 | EVENT EXPERIENCER |
| 2 | EXPERIENCER FOCAL_PARTICIPANT |

Some valence patterns found in SweFN

"I\_\text{Experiencer} \text{do\ 't WANT [to deceive anyone]}\text{EVENT}"

"Jag\_\text{Experiencer KÄNNER FÖR [en tur på landet]}\text{Focal\_participant}"

want.v..6412

känna\_för.vb..1

an embedded frame
FrameNet and GF

• Existing FNs are not entirely formal and computational
  – A limited but computational FN-based GF grammar and lexicon

• Grammatical Framework:
  – Separates between an abstract syntax and concrete syntaxes
  – Provides a general-purpose resource grammar library (RGL)

• The language-independent layer of FrameNet (frames and FEs) – the abstract syntax
  – The language-specific layers (surface realization of frames and FEs; LUs) – concrete syntaxes

• RGL can be used for unifying the syntactic types used in different FNs and for the concrete implementation of frames
  – FrameNet allows for abstracting over RGL
Use case (1)

- Provide a **semantic** API on top of RGL to facilitate the development of GF application grammars
  - In combination with the **syntactic** API of RGL
  - Hiding the comparatively complex construction of **verb phrases**

```
mkC1 person (mkVP (mkVP live_V) (mkAdv in_Prep place))
  -- mkC1  : NP -> VP -> Cl
  -- mkVP  : V -> VP
  -- mkVP  : VP -> Adv -> VP
  -- mkAdv : Prep -> NP -> Adv

Residence
  person
  (mkAdv in_Prep place)
  live_V_Residence
  -- Residence : NP -> Adv -> V -> Cl
  -- NP (Resident)
  -- Adv (Location)
  -- V (LU)
```
Use case (2)

- FN-annotated knowledge bases → multilingual verbalization

<table>
<thead>
<tr>
<th>Being_born</th>
<th>Time</th>
<th>Place</th>
<th>Relatives</th>
<th>Child</th>
</tr>
</thead>
<tbody>
<tr>
<td>dzimt.v</td>
<td>1933. gada 3. maijs</td>
<td>Slokas pagasts</td>
<td>zvejnieka ģimene</td>
<td>Imants Ziedonis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Education Teaching</th>
<th>Institution</th>
<th>Subject</th>
<th>Time</th>
<th>Place</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>absolvēt.v</td>
<td>Tukuma 1. vidusskola</td>
<td></td>
<td>1952. gads</td>
<td>Tukums</td>
<td>Imants Ziedonis</td>
</tr>
<tr>
<td>beigt.v</td>
<td>Latvijas Universitāte vēsture un filoloģija</td>
<td>1959. gads</td>
<td></td>
<td></td>
<td>Imants Ziedonis</td>
</tr>
<tr>
<td>beigt.v</td>
<td>Augstākais literārais [..]</td>
<td>1964. gads</td>
<td></td>
<td></td>
<td>Imants Ziedonis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Being_employed</th>
<th>Employer</th>
<th>Place_of_employment</th>
<th>Position</th>
<th>Time</th>
<th>Employee</th>
</tr>
</thead>
<tbody>
<tr>
<td>redaktors.n</td>
<td>izdevniecība Liesma</td>
<td></td>
<td>&gt; redaktors</td>
<td></td>
<td>Imants Ziedonis</td>
</tr>
<tr>
<td>sekretārs.n</td>
<td>Latvijas rakstnieku [..]</td>
<td></td>
<td>&gt; sekretārs</td>
<td></td>
<td>Imants Ziedonis</td>
</tr>
<tr>
<td>loceklis.n</td>
<td>AP tautas izglītības [..]</td>
<td></td>
<td>&gt; loceklis</td>
<td></td>
<td>Imants Ziedonis</td>
</tr>
<tr>
<td>loceklis.n</td>
<td>Latvijas Institūts</td>
<td></td>
<td></td>
<td></td>
<td>Imants Ziedonis</td>
</tr>
<tr>
<td>padomnieks.n</td>
<td></td>
<td></td>
<td>&gt; padomnieks</td>
<td>1997. gads</td>
<td>Imants Ziedonis</td>
</tr>
<tr>
<td>skolotājs.n</td>
<td>Jūrmalas 1. vidusskola</td>
<td></td>
<td>&gt; skolotājs</td>
<td></td>
<td>Imants Ziedonis</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Win_prize</th>
<th>Time</th>
<th>Prize</th>
<th>Rank</th>
<th>Competition</th>
<th>Competitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>apbalvot.v</td>
<td>1983. gads</td>
<td>Tautu draudzības [..]</td>
<td></td>
<td></td>
<td>Imants Ziedonis</td>
</tr>
<tr>
<td>piešķirt.v</td>
<td>1972. gads</td>
<td>Nopelniem bagātais [..]</td>
<td></td>
<td></td>
<td>Imants Ziedonis</td>
</tr>
<tr>
<td>piešķirt.v</td>
<td>1977. gads</td>
<td>Tautas dzejnieka goda [..]</td>
<td></td>
<td></td>
<td>Imants Ziedonis</td>
</tr>
<tr>
<td>sanemt.v</td>
<td>1991. gada barikāzu [..]</td>
<td></td>
<td></td>
<td></td>
<td>Imants Ziedonis</td>
</tr>
</tbody>
</table>

*Imants Ziedonis ir dzimis 1933. gada 3. maijā Slokas pagastā.*

*Imants Ziedonis was born in Sloka parish on 3 May 1933.*
FrameNet-based grammar: abstract

- **Frame** valence patterns are represented by functions
  - Taking one or more core FEs (A-Z) and one LU as arguments
  - Returning an object of type *Clause* whose linearization type is 
    \{np: NP; vp: VP\}

```
fun Desiring_V : Experiencer_NP -> Focal_participant_Adv -> V -> Clause
fun Desiring_V2 : Experiencer_NP -> Focal_participant_NP -> V2 -> Clause
fun Desiring_V2_Pass : Experiencer_NP -> Focal_participant_NP -> V2 -> Clause
fun Desiring_VV : Event_VP -> Experiencer_NP -> VV -> Clause
```

- **FEs** are declared as semantic categories subcategorized by the syntactic RGL types
  - *NP, VP, Adv* (includes prepositional objects), *S* (embedded sentences), *QS*

```
cat Event_VP
cat Experiencer_NP
cat Focal_participant_NP
cat Focal_participant_Adv
```
FrameNet-based grammar: concrete

- The mapping from the semantic FrameNet types to the syntactic RGL types is shared for all languages
  - Linearization types are of type *Maybe* to allow for optional (empty) FEs
    
    \[
    \text{lin} \text{cat } \text{Focal\_participant\_NP} = \text{Maybe} \ NP \\
    \text{lin} \text{cat } \text{Focal\_participant\_Adv} = \text{Maybe} \ Adv
    \]

- To implement the frame functions, RGL constructors are applied to the arguments depending on their types and syntactic roles, and the voice

  \[
  \text{lin } \text{Desiring\_V2} \text{ experiencer focal\_participant } v2 = \{
  np = \text{fromMaybe} \ NP \text{ experiencer} ; \\
  vp = \text{mkVP} \ v2 \ (\text{fromMaybe} \ NP \text{ focal\_participant})
  \}
  \]

  \[
  \text{lin } \text{Desiring\_V2\_Pass} \text{ experiencer focal\_participant } v2 = \{
  np = \text{fromMaybe} \ NP \text{ focal\_participant} ; \\
  vp = \text{mkVP} \ (\text{passiveVP} \ v2) \ (\text{mkAdv} \ by8agent\_Prep \ (\text{fromMaybe} \ NP \text{ experiencer}))
  \}
  \]
FrameNet-based API to GF Resource Grammar Library
A tool for cross-lingual comparison of FrameNet-annotated corpora

<table>
<thead>
<tr>
<th>Frames</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Court_examination</td>
<td>(1)</td>
</tr>
<tr>
<td>Create_physical_artwork</td>
<td>(1)</td>
</tr>
<tr>
<td>Create_representation</td>
<td>(2)</td>
</tr>
<tr>
<td>Creating</td>
<td>(2)</td>
</tr>
<tr>
<td>Criminal_investigation</td>
<td>(1)</td>
</tr>
<tr>
<td>Cutting</td>
<td>(2)</td>
</tr>
<tr>
<td>Damaging</td>
<td>(2)</td>
</tr>
<tr>
<td>Daring</td>
<td>(2)</td>
</tr>
<tr>
<td>Death</td>
<td>(1)</td>
</tr>
<tr>
<td>Deciding</td>
<td>(1)</td>
</tr>
<tr>
<td>Delimitation_of_diversity</td>
<td>(1)</td>
</tr>
<tr>
<td>Delivery</td>
<td>(3)</td>
</tr>
<tr>
<td>Deny_permission</td>
<td>(1)</td>
</tr>
<tr>
<td>Departing</td>
<td>(1)</td>
</tr>
<tr>
<td>Deserving</td>
<td>(1)</td>
</tr>
<tr>
<td>Desiring</td>
<td>(3)</td>
</tr>
<tr>
<td>Destroying</td>
<td>(4)</td>
</tr>
<tr>
<td>Detaching</td>
<td>(2)</td>
</tr>
<tr>
<td>Detaining</td>
<td>(2)</td>
</tr>
<tr>
<td>Differentiation</td>
<td>(3)</td>
</tr>
<tr>
<td>Discussion</td>
<td>(1)</td>
</tr>
<tr>
<td>Dispersal</td>
<td>(1)</td>
</tr>
<tr>
<td>Distinctiveness</td>
<td>(1)</td>
</tr>
<tr>
<td>Dodging</td>
<td>(1)</td>
</tr>
<tr>
<td>Dominate_competitor</td>
<td>(1)</td>
</tr>
<tr>
<td>Dominate_situation</td>
<td>(1)</td>
</tr>
<tr>
<td>Dressing</td>
<td>(1)</td>
</tr>
<tr>
<td>Drop_in_on</td>
<td>(1)</td>
</tr>
<tr>
<td>Dunking</td>
<td>(1)</td>
</tr>
<tr>
<td>Duplication</td>
<td>(2)</td>
</tr>
<tr>
<td>Duration_relation</td>
<td>(2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Verbs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Desiring_V</td>
<td></td>
</tr>
<tr>
<td>Desiring_V2</td>
<td></td>
</tr>
<tr>
<td>Desiring_VV</td>
<td></td>
</tr>
</tbody>
</table>

**Desiring**

**Desiring_V** : Experiencer_NP \rightarrow Focal_participant_Adv \rightarrow V \rightarrow Clause

- **Eng**: [They] Experiencer [ASPIRED] [towards the Chelsea shore, where, in the early 1960s many thousands lived with sensible occupations and adequate amounts of money] Focal_participant

- **Swe**: [Roberte] Experiencer [LÄNGTADE] [hem till Tyskland] Focal_participant

**Desiring_V2** : Experiencer_NP \rightarrow Focal_participant_NP \rightarrow V2 \rightarrow Clause

- **Eng**:
  - covet_V2_Desiring : V2
  - crave_V2_Desiring : V2
  - desire_V2_Desiring : V2

- **Swe**: [Jag] Experiencer [KÄNNER FÖR] [en tur på landet] Focal_participant

**Desiring_VV** : Event_VP \rightarrow Experiencer_NP \rightarrow VV \rightarrow Clause

- **Eng**: [He] Experiencer ground his teeth together, [LUSTING] [to tear the alien apart and eat of its lurid vitals, so as to comprehend something of its strange nature] Event

[http://grammaticalframework.org/framenet/](http://grammaticalframework.org/framenet/)
Semantic Role Labelling (SRL)

A fourth member, Jean-Marc Rouillan, remains behind bars.

FrameNet
TurboParser + SEMAFOR: http://demo.ark.cs.cmu.edu/parse

<table>
<thead>
<tr>
<th>A</th>
<th>fourth</th>
<th>member</th>
<th>,</th>
<th>Jean Marc</th>
<th>Rouillan</th>
<th>,</th>
<th>remains</th>
<th>behind</th>
<th>bars</th>
</tr>
</thead>
<tbody>
<tr>
<td>member.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>remain.01</td>
<td></td>
<td>A1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A3</td>
<td></td>
</tr>
</tbody>
</table>

PropBank
LTH parser: http://barbar.cs.lth.se:8081/
PropBank (http://propbank.github.io)

remain-v; 3 senses

**Sense Number 1: remain in a certain state**

**Examples:**
The President is expected to remain firm on his veto of the bill. In this game you must remain down to the count of ten.

**Mappings:**
FrameNet: State_continue
PropBank: remain.01
WordNet 3.0 Sense Numbers: 1
WordNet Verb Particle Constructions, Multiword Expressions: remain_down 1
remain_firm 1

**Predicate: remain**

Roles:
Arg1-PPT: Thing remaining (vnrole: 47.1-1-Theme)
Arg2-GOL: benefactive, entity who gets the remainder
Arg3-PRD: attribute of arg1

**Example: All alone**

Only one local ringer remains.

Arg1: Only one local ringer
Rel: remains

**Example: With PP**

Exports will remain under a government target.

Arg1: Experts
Rel: remain
Arg3: under a government target

**Example: remain on**

Five countries will remain on the so-called priority watch list.

Arg1: Five countries
Rel: remain
Arg3: on the so-called priority watch list

**Example: With AdjP**

Their influence will remain subordinate to Japan's.

Arg1: Their influence
Rel: remain
Arg3: subordinate to Japan's

**Sense Number 2: be left; stay behind; continue in a place, position, or situation**

**Examples:**
She remained with the child until he fell asleep. If you subtract ten from twelve, two remain.

**Mappings:**
FrameNet: Existence,Remainder
PropBank: remain.01
WordNet 3.0 Sense Numbers: 2, 3, 4
AMR (Abstract Meaning Representation)

• From SRL to whole-sentence meaning representation
  – Incl. PropBank SRL, NER and NEL, treatment of modality, negation, etc.

• Simple and compact data structure
  – PENMAN notation: directed labeled graph encoded in a tree-like form
  – Easy to read and write (for a human), and traverse (for a program)

• Aimed at large-scale human annotation and semantic parsing
  – Practical, replicable amount of abstraction
  – An actual sembank of 40K+ sentences

• Captures many aspects of meaning
  – Aims to abstract away from (English) syntax
**AMR (Abstract Meaning Representation)**

- Nodes are **variables** labelled by **concepts**
  - Entities, events, states, properties
  - \( s / \text{soldier} \): \( s \) is an instance of \( \text{soldier} \)
- Edges are semantic **relations**
- AMR abstracts in numerous ways by assigning the same conceptual structure to different surface realizations

\[
(f / \text{fear-01} \\
: \text{polarity} "-" \\
: \text{ARG0} ( s / \text{soldier} ) \\
: \text{ARG1} ( d / \text{die-01} \\
: \text{ARG1} s )
\]

The soldier was not afraid of dying. The soldier was not afraid to die. The soldier did not fear death.

(Pust et al., 2015)
AMR (Abstract Meaning Representation)

• AMR is still **biased** towards English or other source languages

• Meanwhile, AMR is **agnostic** about how to derive meanings from strings, and vice versa


Schneider N., Flanigan J., O’Gorman T. AMR Tutorial at NAACL 2015
[https://github.com/nschneid/amr-tutorial/](https://github.com/nschneid/amr-tutorial/)
Text-to-AMR: human annotation

(j / join-01
 :ARG0 (p / person :name (p2 / name :op1 "Pierre" :op2 "Vinken")
 :age (t / temporal-quantity :quant 61
 :unit (y / year)))
 :ARG1 (b / board)
 :prep-as (d2 / director
 :mod (e / executive))
 :time (d / date-entity :month 11 :day 29))

Enter text command: e :polarity -

Last command: d2 :mod executive

Or select an action template: top add add-ne replace delete move undo exit/load prop

Workset wsj100-sent 1/100 nw.wsj_0001.1 Save and load next Discard and load next

Figure 1: Screenshot of the AMR Editor when entering a text command, showing the core portion of the main window.
AMR-to-text: human evaluation

(c / claim-01
 :ARG0 (h / he)
 :ARG1 (e / expose-01
 :ARG0 (p / person
 :ARG0-of (s / sing-01
 :age (t / temporal-quantity :quant 28
 :unit (y / year)))
 :ARG1 p
 :ARG2 h
 :ARG1-of (r / repeat-01)))

— Source

he claims the 28-year-old singer repeatedly exposed herself to him.

— Reference

He claims that a person exposes a singing person of 28 years repeated.

he claims to have been exposed to singers, 28 years old and has repeatedly

he claimed repeatedly that the 28-year-old singing has exposed.
A fourth member, Jean-Marc Rouillan, remains behind bars.
A fourth member, Jean-Marc Rouillan, remains behind bars.

Remaining members person 4 jean-marc rouillan – behind bar.

Jean-Marc Rouillan, that is the 4th member, is remained behind a bar.
A fourth member, Jean-Marc Rouillan, remains behind bars.

(r / remain-01
  :ARG1 (p / person
    :wiki -
    :name (n / name
      :op1 "Jean-Marc" :op2 "Rouillan")
    :mod (p2 / person
      :ARG0-of (h / have-org-role-91
        :ARG2 (m / member))
      :ord (o / ordinal-entity
        :value 4)))
  :ARG3 (b / behind
    :op1 (b2 / bar)))
A fourth member, Jean-Marc Rouillan, remains behind bars.

Remaining members person 4 jean-marc rouillan – behind bar.

Jean-Marc Rouillan, that is the 4th member, is remained behind a bar.
They should have been expelled from school at a minimum.

It is recommended that they are expelled to a school at a minimum.

**expel-01**

- **ARG0=PAG** (prototypical agent)
- **ARG1=PPT** (prototypical patient)
- **ARG2=DIR** (direction) $\rightarrow$ DIR Prep $\rightarrow$ to Prep

**ToDo:** based on statistics from PropBank and FrameNet corpora, “reconstruct” Prep-s, depending on frame/verb valency, ARG role, or NP head
Texas criminal courts and prosecutors do not coddle anyone.

No Texas Texas criminal court and prosecutors coddle anyone.

A criminal court in Texas and a person that prosecutes do not coddle anyone.

person that prosecutes → prosecutor
organization that governs → government
Sample AMR (4)

# ::snt How Long are We Going to Tolerate Japan?

(t / tolerate-01
 :ARG0 (w / we)
 :ARG1 (c / country
       :wiki "Japan"
       :name (n / name :op1 "Japan")
       :duration (a / amr-unknown))

We have tolerated the japan amr-unknown.

How long do we tolerate Japan?

if ':mode expressive' in amr: amr = amr.replace(':mode expressive', ' ') + '!
if ':mode imperative' in amr: amr = amr.replace(':mode imperative', ' ') + '!
if ':mode interrogative' in amr: amr = amr.replace(':mode interrogative', ' ') + '?
if 'cause-01:ARG0(amr-unknown)' in amr: amr = 'why ' + amr.replace('cause-01:ARG0(amr-unknown)', ' ') + '?
if ':location(amr-unknown)' in amr: amr = 'where ' + amr.replace(':location(amr-unknown)', ' ') + '?
if ':ARG1(amr-unknown)' in amr: amr = 'who ' + amr.replace(':ARG1(amr-unknown)', ' ') + '?
if ':mod(amr-unknown)' in amr: amr = 'what ' + amr.replace(':mod(amr-unknown)', ' ') + '?
if ':duration(amr-unknown)' in amr: amr = 'how ' + amr.replace(':duration(amr-unknown)', ' ') + '?
if 'amr-unknown'

JAMR

if ':mode expressive'
if ':mode imperative'
if ':mode interrogative'
if 'cause-01:ARG0(amr-unknown)'
if ':location(amr-unknown)'
if ':ARG1(amr-unknown)'
if ':mod(amr-unknown)'
if ':duration(amr-unknown)'
if 'amr-unknown'

GF
Xinhua News Agency has reported Yiguo Yu byline in a Tokyo 19.

Xinhua News Agency by Yiguo Yu on 1 September in Tokyo.
Alliot-Marie arrived on Sunday.

Sunday's arrival of alliot-marie michèle_alliot-marie.

unknown qualified constant L.arrive_V2

Alliot-marie michèle_alliot-marie arrived sunday.

Alliot-Marie arrives on Sunday.
SemEval 2017: Task 9

- Subtask 1: Parsing Biomedical Data
- Subtask 2: AMR-to-English Generation

<table>
<thead>
<tr>
<th>Approaches:</th>
</tr>
</thead>
<tbody>
<tr>
<td>“grammar-based”</td>
</tr>
<tr>
<td>SMT/NMT</td>
</tr>
<tr>
<td>end-to-end</td>
</tr>
</tbody>
</table>

Table 3: Main generation results: The three manually-derived metrics agree on the systems’ relative rankings.

<table>
<thead>
<tr>
<th></th>
<th>Win</th>
<th>Win+Tie</th>
<th>Trueskill</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIGOTRIO</td>
<td>54.91</td>
<td>81.49</td>
<td>1.07</td>
<td>18.82</td>
</tr>
<tr>
<td>CMU</td>
<td>50.36</td>
<td>72.48</td>
<td>0.85</td>
<td>19.01</td>
</tr>
<tr>
<td>FORGe</td>
<td>43.64</td>
<td>57.43</td>
<td>0.45</td>
<td>4.74</td>
</tr>
<tr>
<td>ISI</td>
<td>26.05</td>
<td>38.39</td>
<td>-1.19</td>
<td>10.92</td>
</tr>
<tr>
<td>Sheffield</td>
<td>8.38</td>
<td>21.16</td>
<td>-2.20</td>
<td>3.32</td>
</tr>
</tbody>
</table>

Table 4: Human judgments of generation results after self-judgments are removed: The results are fundamentally the same.

<table>
<thead>
<tr>
<th></th>
<th>Win</th>
<th>Win+Tie</th>
<th>Trueskill</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIGOTRIO</td>
<td>53.00</td>
<td>79.98</td>
<td>1.03</td>
</tr>
<tr>
<td>CMU</td>
<td>50.02</td>
<td>71.91</td>
<td>0.819</td>
</tr>
<tr>
<td>FORGe</td>
<td>44.49</td>
<td>58.57</td>
<td>0.458</td>
</tr>
<tr>
<td>ISI</td>
<td>26.40</td>
<td>38.60</td>
<td>-1.172</td>
</tr>
<tr>
<td>Sheffield</td>
<td>9.46</td>
<td>22.84</td>
<td>-2.132</td>
</tr>
</tbody>
</table>
RIG-GOT-RIO → Trio from Riga with regards to GOT & RIO ;)

RIGOTRIO at SemEval-2017 Task 9: Combining Machine Learning and Grammar Engineering for AMR Parsing and Generation

Normunds Gruzitis¹, Didzis Gosko² and Guntis Barzdins¹

¹University of Latvia, IMCS / Rainis blvd. 29, Riga, Latvia
normunds.gruzitis@lumii.lv, guntis.barzdins@lumii.lv
²LETA / Marija street 2, Riga, Latvia
didzis.gosko@leta.lv

Abstract

By addressing both text-to-AMR parsing and AMR-to-text generation, SemEval-2017 Task 9 established AMR as a powerful semantic interlingua. We strengthen the interlingual aspect of AMR by applying the multilingual Grammatical Framework (GF) for AMR-to-text generation. Our current rule-based GF approach completely covered only 12.3% of the test AMRs, therefore we combined it with state-of-the-art JAMR Generator to see if the combination increases or decreases the overall performance. The combined system achieved the automatic BLEU score of 18.82 and the human Trueskill score of 107.2, to be compared to the plain JAMR Generator results. As for AMR common NER tools that are often restricted to types “person”, “organization”, “location”, etc.

The paper starts with NER extensions used for the Biomedical AMR parsing subtask, followed by a novel approach of using Grammatical Framework for AMR generation, and concludes with a brief analysis of our SemEval results.

2 Text-to-AMR parsing

Only two adaptations to the AMR parser from SemEval-2016 (Barzdins and Gosko, 2016) were implemented: it was retrained on the union of LDC2015E86, LDC2016E25, LDC2016E33 and Bio AMR Corpus, and a gazetteer was added to extend the NER coverage to organic compound names found in the Bio AMR Corpus (e.g. “B-Raf enzyme”, “dabrafenib small-molecule”, etc.). The gazetteer was generalized w.r.t. numbers used
Under the hood

The boys want to go to New York City.

Text-to-AMR parser

AMR-to-AST transformer

(mkText (mkUtt (mkS (mkCl (mkNP a_Quant (mkCN boy_N)) (mkVP want_VV) (mkVP go_V)) (mkAdv to_PreP (mkNP (mkPN "New York City")) ))))) fullStopPunct)

(w / want-01
  :ARG0 (b / boy)
  :ARG1 (g / go-02
    :ARG0 b
    :ARG4 (c / city
      :name (n / name
        :op1 "New"
        :op2 "York"
        :op3 "City"
        :wiki "New_York_City"))))

Spanish resource grammar

Un niño quiere ir a New York City.

English resource grammar

A boy wants to go to New York City.

German resource grammar

Ein Junge will nach New York City gehen.

# mkVP : VV -> VP -> VP
# (frame1 (:ARG1 (var frame2))) => (frame1 (mkVP frame2))
/VV_FRAME/ < (/:ARG1/=vp < (/VAR/=var < /FRAME/=v))>
[move v >1 vp] [relabel vp /\.+$/mkVP/] [delete var]  # Tregex
# Tsurgeon

```
(EVENT-3 / see
   :ARG0 (PERSONS-1 / girls)
   :ARG1 (PERSON-2 / boy))
```
Multilingual AMR-to-Text: experiment

TestTrees: t01_girls_see_a_boy
TestTreesEng: a girl sees a boy.
TestTreesLav: meitene redz zēnu.
TestTreesRus: девушка видит мальчика.

TestTrees: t04_two_pretty_girls_see_a_boy
TestTreesEng: 2 pretty girls see a boy.
TestTreesLav: 2 jaukas meitenes redz zēnu.
TestTreesRus: 2 хорошенькие девочки видят мальчика.

TestTrees: t21_girls_who_see_the_game_like_the_boys_who_play
TestTreesEng: a girl that sees a game likes a boy that plays.
TestTreesLav: meitenei, kas redz spēli, patīk zēns, kas spēlē.
TestTreesRus: девушка, которая видит игру нравдит мальчика, которого играет.

TestTrees: t27_they_are_thugs_and_deserve_a_bullet
TestTreesEng: they are a thug and it deserves a bullet.
TestTreesLav: viņi ir slepkava, un pelna lodi.
Under the hood

The overall AMR-to-text process:

1. The input AMR is rewritten from the PENMAN notation to the the LISP-like bracketing tree syntax.

2. In case of a multi-sentence AMR, the graph is split into two or more graphs to be processed separately.

3. For each AMR, a sequence of tree pattern-matching transformation rules is applied (Tregex + Tsurgeon), acquiring a fully or partially converted GF abstract syntax tree (AST).

4. In case of a partially converted AST, the pending subtrees are pruned.**

5. The resulting ASTs are passed to the GF interpreter for RGL-based linearization.

6. Since RGL supports many more languages (30+), this approach can be extended to multilingual AMR-to-text generation, given a large translation lexicon (15+).
**Our SemEval submission:**

Because the coverage of our hand-crafted AMR-to-AST transformation rules is currently far from complete, we used JAMR Generator (Flanigan et al., 2016) as a “fall-back” option for AMRs that are not fully covered by the current rule set (~200).

However, we applied heuristic post-processing rules to the JAMR output, which might have influenced the human judgements:

- Adding a full-stop, or question mark, or exclamation mark at the end of the sentence, or a wh-word at the beginning, based on the AMR constructs.
- Removing the remaining (unresolved) AMR constructs and concepts.
- Converting large numbers into words, adding some prepositions, etc.
AMR-to-text generation via GF abstract syntax trees

The initial proposal is concisely described in a position paper by Grūžītis & Bārzdīņš (2016). Details of the current implementation can be found in a system description paper by Grūžītis et al. (2017). See the Publication section below.

Outline: for a given AMR graph, represented as a tree in the PENMAN notation, transform it to a GF abstract syntax tree (AST), and linearize the AST in the target language. The output sentence is in general a paraphrase of the input sentence represented by the AMR graph.

Structure

1. lexicons: monolingual and multilingual GF lexicons – extensions to the wide coverage lexicons provided by the GF resource grammar library (RGL).
The Role of CNL and AMR in Scalable Abstractive Summarization for Multilingual Media Monitoring
BBC monitoring journalists translate from 30 languages into English, follow 400 social media accounts every day.

A monitoring journalist typically monitors 4 TV channels and several online sources simultaneously. This is about the maximum that any person can cope with mentally and physically. The required human effort thus scales linearly with the number of monitored sources.

Monitoring journalists constantly need to be on the lookout for more sources and follow important stories—but as it is, they are tied down with mundane, routine monitoring tasks.

Monitoring 250 video channels results in a daily buffer of 2.5TB, a weekly buffer of 19Tb, and an annual buffer of 1Pb.
Identify **people, places, events** of interest

Discover **trends**, emerging **events**, crucial new **stories**
### Storyline highlights

| Article 1: | ⋄ An ongoing battle in Aleppo eventually terminated when the rebels took over the city. ⋄ President Assad gave a speech, denouncing the death of Syrian soldiers. ⋄ |
| Article 2: | ⋄ Syrian rebels took control of Aleppo ⋄ |
| Article 3: | ⋄ The Syrian opposition forces won the battle over Aleppo city. ⋄ Syrian president announced on Syrian television that such insurgency will not be tolerated. ⋄ |
| Blog 1: | ⋄ As described in [this news story](link to article 1) our Syrian brothers are starting to make progress in their opposition to the tyrannic rule of Assad. ⋄ |

### Output Summary:

- Syrian rebels took over Aleppo
  - Article1 Article2 Article3
  - Sentiment: 70% positive, Variance: High

- Assad gave a speech about the battle
  - Article1 Article3
  - Sentiment: 35% positive, Variance: Low

- Event-based multi-document summarization
- Storyline highlights across a set of related stories
Abstractive text summarization

- Extractive summarization selects representative sentences from the input documents.

- Abstractive summarization builds a semantic representation from which a summary is generated.

- What semantic representation?
  - PropBank / FrameNet
  - AMR

---

Sentence A: I saw Joe’s dog, which was running in the garden.
Sentence B: The dog was chasing a cat.
Summary: Joe’s dog was chasing a cat in the garden.

SemEval 2016 Task 8 on AMR parsing

1. **Riga** (U Latvia, IMCS / LETA): 0.6196
2. **CAMR** (U Brandeis / Boulder Learning Inc. / Rensselaer Polytechnic Institute): 0.6195
3. **ICL-HD** (Ruprecht-Karls-Universität Heidelberg): 0.6005
4. **UCL+Sheffield** (University College London / U Sheffield): 0.5983
5. **M2L** (Kyoto University): 0.5952
6. **CMU** (Carnegie Mellon University / U Washington): 0.5636
7. **CU-NLP** (OK Robot Go Ltd. / U Colorado): 0.5566
8. **UofR** (U Rochester): 0.4985
9. **MeaningFactory** (U Groningen): 0.4702*
10. **CLIP@UMD** (U Maryland): 0.4370
11. **DynamicPower** (National Institute for Japanese Language and Linguistics): 0.3706*

* Rule/grammar-based; did not use AMR training data
Conclusion

• Unrestricted large-scale **NLU** is difficult for grammars
  – SemEval 2016: **few** grammar-based systems
  – SemEval 2017: **no** grammar-based systems (Boxer gave up...)

• **For NLG, grammar-based systems are very competitive!**

• Scaling up AMR-to-AST:
  – Add more Tregex/Tsurgeon rules
  – A more flexible and systematic graph/tree-transducer (like UD2GF)
  – Learning transformation rules (C6.0; training data?)
  – Seq-to-seq deep learning?
Publications


• Dana Dannélls, Normunds Grūzītis. Extracting a bilingual semantic grammar from FrameNet-annotated corpora. LREC 2014

• Dana Dannélls, Normunds Grūzītis. Controlled natural language generation from a multilingual FrameNet-based grammar. CNL 2014

• Normunds Grūzītis, Dana Dannélls, Benjamin Lyngfelt, Aarne Ranta. Formalising the Swedish Constructicon in Grammatical Framework. GEAF 2015

• Normunds Grūzītis, Guntis Bārzdiņš. The role of CNL and AMR in scalable abstractive summarization for multilingual media monitoring. CNL 2016

• Normunds Grūzītis, Dana Dannélls. A Multilingual FrameNet-based Grammar and Lexicon for Controlled Natural Language. Language Resources and Evaluation, 51(1), 2017