Natural Language Knowledge Graphs

Ido Dagan, Bar-Illan University
Joint work with Meni Adler, Ori Shapira, Vered Shwartz,
Gabriel Stanovksy, Rachel Wities, Iryna Gurevych

Work in progress
A Perspective: Structured Knowledge Graphs

Freebase knowledge graph

[Diagram showing relationships between entities such as Michelle Obama, Barack Obama, Chicago, Honolulu, United States, and event dates such as 1992.10.03 and 1961.08.04.]
Structured Knowledge Graphs (KG)

- Encode selected facts about the world
- as propositions - asserted relations (predicates) between entities
- Allow structured presentation and querying
- Consolidating many facts about the encoded entities
- Relations and entities are pre-specified
Our Motivation: NL-based KG (NLKG)

- How can we capture unbounded information in collections of text?
- How can we communicate it efficiently to users, or applications?

- An “open” NL-based representation
  - Vs. formal KG

- This talk:
  - Defining the graph structure & dataset annotation
  - Illustrating application potential
  - No algorithms...
Illustrating Example: Consolidating Event Tweets - Stand-alone Sentences
Turkey forces a plane from Russia to land in Ankara.

Syrian jet grounded by the Turkish military was carrying munitions.

Intercepted Syria-bound plane carried ammunition from Russia to Syria.

Russia angry at Turkey about grounded Syrian jet and Russian passengers

Turkey Escalates Confrontation with Syria, forcing Syrian plane flying from Russia to land in Turkey

Turkey forces down suspect Syrian plane

Turkish PM says plane from Moscow forced to land in Turkey was carrying ammunition for Syria government.

Turkish F16s force Syrian passenger plane enroute from Moscow land in Ankara.

...
Turkey forces a plane from Russia to land in Ankara.

Syrian jet grounded by the Turkish military was carrying munitions.

Intercepted Syria-bound plane carried ammunition from Russia to Syria.

Russia angry at Turkey about grounded Syrian jet and Russian passengers

Turkey Escalates Confrontation with Syria, forcing Syrian plane flying from Russia to land in Turkey

Turkey forces down suspect Syrian plane

Turkish PM says plane from Moscow forced to land in Turkey was carrying ammunition for Syria government.

Turkish F16s force Syrian passenger plane enroute from Moscow land in Ankara.
Turkey forces a plane from Russia to land in Ankara.

Syrian jet grounded by the Turkish military was carrying munitions.

Intercepted Syria-bound plane carried ammunition from Russia to Syria.

Russia angry at Turkey about grounded Syrian jet and Russian passengers

Turkey Escalates Confrontation with Syria, forcing Syrian plane flying from Russia to land in Turkey

Turkey forces down suspect Syrian plane

Turkish PM says plane from Moscow forced to land in Turkey was carrying ammunition for Syria government.

Turkish F16s force Syrian passenger plane enroute from Moscow land in Ankara.

...
Turkey forces a plane from Russia to land in Ankara.

Syrian jet grounded by the Turkish military was carrying munitions.

Intercepted Syria-bound plane carried ammunition from Russia to Syria.

Russia angry at Turkey about grounded Syrian jet and Russian passengers

Turkey Escalates Confrontation with Syria, forcing Syrian plane flying from Russia to land in Turkey

Turkey forces down suspect Syrian plane

Turkish PM says plane from Moscow forced to land in Turkey was carrying ammunition for Syria government.

Turkish F16s force Syrian passenger plane enroute from Moscow land in Ankara.

...
Characteristic 1: Redundancy

Turkey forces a plane from Russia to land in Ankara.

Syrian jet grounded by the Turkish military was carrying munitions.

Intercepted Syria-bound plane carried ammunition from Russia to Syria.

Russia angry at Turkey about grounded Syrian jet and Russian passengers

Turkey Escalates Confrontation with Syria, forcing Syrian plane flying from Russia to land in Turkey

Turkey forces down suspect Syrian plane

Turkish PM says plane from Moscow forced to land in Turkey was carrying ammunition for Syria government.

Turkish F16s force Syrian passenger plane enroute from Moscow land in Ankara.

...
Turkey forces a plane from Russia to land in Ankara.

Syrian jet grounded by the Turkish military was carrying munitions.

Intercepted Syria-bound plane carried ammunition from Russia to Syria.

Russia angry at Turkey about grounded Syrian jet and Russian passengers

Turkey Escalates Confrontation with Syria, forcing Syrian plane flying from Russia to land in Turkey

Turkey forces down suspect Syrian plane

Turkish PM says plane from Moscow forced to land in Turkey was carrying ammunition for Syria government.

Turkish F16s force Syrian passenger plane enroute from Moscow land in Ankara.

...
Turkey intercepted a Syrian plane in Ankara

The plane carried ammunition from Russia to Syria

Turkey Escalates Confrontation with Syria

Russia angry at Turkey

Turkish PM did not apologize
Turkey intercepted a Syrian plane in Ankara. The plane carried ammunition from Russia to Syria.

Turkey Escalates Confrontation with Syria.

Russia angry at Turkey.

Turkish PM did not apologize.
Turkey intercepted a Syrian plane in Ankara.

The plane carried ammunition from Russia to Syria.

Turkey Escalates Confrontation with Syria.

Russia angry at Turkey.

Turkish PM did not apologize.
Turkey intercepted a Syrian plane in Ankara

The plane carried ammunition from Russia to Syria

Turkey Escalates Confrontation with Syria

Russia angry at Turkey

Turkish PM did not apologize
Turkey intercepted a Syrian passenger plane in Ankara.

The plane carried ammunition from Russia to Syria.

Turkey Escalates Confrontation with Syria.

Russia angry at Turkey.

Turkish PM did not apologize.
Turkey intercepted a Syrian passenger plane in Ankara.

The plane carried ammunition from Russia to Syria.

Turkey Escalates Confrontation with Syria.

Russia angry at Turkey.

Turkish PM did not apologize.
Background:
Build NLKG (mostly) from *known* component tasks
Single Sentence Representations (1)

- SRL - annotating predicate-argument structure of sentences.

  Danny arrived to the restaurant to meet Sarah

  ![SRL Diagram]

  * V. Punyakanok et al., 2008

- QA-SRL - an argument role is a natural language question which the argument answers

- Open IE - Isolates “atomic” propositions as predicate-argument tuples

  (Danny, arrived, at restaurant, to meet Sarah)
  (Danny, meet, Sarah)
Single Sentence Representation (2)

- **AMR** (Banarescu et al., 2013)
  - Represents sentence semantic structure as a labeled graph.
  - Assigns same AMR structure to sentences with the same meaning.
Cross-document Consolidation - Pre-defined IE

- **Rich ERE** (DEFT) - Entities, Relations, and Events

- Annotating a small predefined set of entity, relation and event types.
  - 5 Entity types: PER, ORG, GPE, LOC, FAC
  - 20 relation types: member, family, founder, ...
  - 38 event types: attack, marry, appeal, ...

- Consolidation via coreference between mentions of same entity/relation/event
Roth and Frank (2012), PARMA (2013), ECB+ (2014) - Cross-document alignment of events and entities

T1: On Monday Lindsay Lohan checked into rehab in Malibu, California after car crash.
T2: Ms. Lohan entered a treatment facility.
T3: The 26 year old actress checked into Betty Ford Center.

Each of these efforts is limited/partial in certain respects
No consolidation
Coreference-based inference

- Lexical substitution of co-referring elements
  - I saw a boy. He was running ⇒ A boy was running

- Used in question answering, RTE, etc.

- DRT (Kamp 1988, Kamp et al., 2011)
  - In formal semantics - a proposition applies to all co-referring (equal) entities
Our Positioning and Novelty

- Rely on extended QA-SRL extractions and entity/predicate coreference
- Merge pred-arg extractions into a consolidated structure
- Trace redundancies - through entailment relations
NLKG Scope & Requirements
Setting: Natural Language Propositions

- Assumption: textual information can be represented by NL propositions
  - Natural language predicate and arguments
  - Expressing an assertion, to which a truth value can be assigned
  - Typically specifies an event, relation or property

- Similar to Open-IE rationale
Scope of Represented Propositions

● “Verbatim” natural language propositions: only mentioned lexical elements
  ○ Lexical inference is left for a separate module
    ■ E.g. if only apple mentioned, avoid representing fruit

● Coreference-based inference: the basis for merging information
  ○ a plane from Russia + a Syrian jet => a jet from Russia.
  ○ The plane flew from London + The plane flew to Paris. => The plane flew from London to Paris.
Requirements

1. Represent the NL propositions conveyed jointly by the texts
   - **Completeness**: capture all propositions - verbatim + coref-based
   - **Soundness**: all represented propositions are indeed conveyed by the texts
   - Trace original mentions for each information component
     - Frequency - indicates salience & certainty

2. Redundancy modeling
   - Model redundancy between the represented NL propositions, such that -
   - Support incremental communication
     - Use cases: hierarchical presentation, updates
     - A communicated proposition should add new information
Graph Specification & Construction
Overview

*Knowledge graph of consolidated propositions: a *novel structure* based (mostly) on *known NLP tasks*

1. Consolidating *entities & propositions*
   - Via coreference chains

2. Modeling redundancy via entailments
Entities

- **Entity mention**
  - A word or multi-word term in a sentence, referring to a real world object or concept

- **Entity**
  - A coreference chain of entity mentions, all referring to the same object/concept
    - A cross-document “discourse entity”
1. Intercepted plane carried ammunition to Moscow.

2. Syrian jet grounded by the Turkish military was carrying munitions from Russia to Syria.

3. Turkey uses military to force a plane from Russia to land in Ankara.

Entity mentions & co-reference chains:

- E1 (plane) -> {plane, jet}
- E2 (ammunition) -> {ammunition, munitions}
- E3 (Ankara) -> {Ankara}
1. Intercepted plane carried ammunition to Moscow.

2. Syrian jet grounded by the Turkish military was carrying munitions from Russia to Syria.

3. Turkey uses military to force a plane from Russia to land in Ankara.
Propositions

- **Proposition mention**
  - A fragment in the text, referring to a single event, relation or property.

- **Proposition**
  - A coreference chain of proposition mentions that refer to the same real-world fact.
1. Intercepted plane *carried* ammunition from Moscow.

2. Syrian jet grounded by Turkey was carrying munitions from Russia to Syria.

![Graph](image-url)
1. Intercepted plane carried ammunition from Moscow.

2. Syrian jet grounded by Turkey was carrying munitions from Russia to Syria.
1. Intercepted plane carried ammunition from Moscow.

2. Syrian jet grounded by Turkey was carrying munitions from Russia to Syria.
1. Intercepted plane **carried** ammunition **from** Moscow.

2. Syrian jet grounded by Turkey was carrying munitions from Russia to Syria.
1. Intercepted plane carried ammunition from Moscow.

2. Syrian jet grounded by Turkey was carrying munitions from Russia to Syria.
1. Intercepted plane carried ammunition from Moscow.

2. Syrian jet grounded by Turkey was carrying munitions from Russia to Syria.
Turkey uses military to force a plane from Russia to land in Ankara.
Turkey uses military to force a plane from Russia to land in Ankara.
Syrian jet grounded by the Turkish military was carrying munitions from Russia to Syria.
Entailment

- Textual Entailment (between propositions):

  Proposition A entails proposition B if a human reading A would infer that B is true.

- In our graph, we encode entailment between terms or predicate templates:

  Term/predicate entailment holds iff textual entailment holds when substituting one phrase for the other within the given contexts in the graph.

  - jet $\Rightarrow$ plane
    - “Jet landed in Ankara” $\Rightarrow$ “plane landed in Ankara”
    - “Jet carried ammunition” $\Rightarrow$ “plane carried ammunition”
    - ...

  - A1 grounded A2 $\Rightarrow$ A1 intercept A2
    - “Turkey grounded plane” $\Rightarrow$ “Turkey intercepts plane”
    - ...

Intercepted plane carried ammunition to Moscow.

Syrian jet grounded by the Turkish military was carrying munitions from Russia to Syria.

Turkey uses military to force a plane from Russia to land in Ankara.
Intercepted plane carried ammunition to Moscow.

Syrian jet grounded by the Turkish military was carrying munitions from Russia to Syria.

Turkey uses military to force a plane from Russia to land in Ankara.
Intercepted plane carried ammunition to Moscow.

Syrian jet grounded by the Turkish military was carrying munitions from Russia to Syria.

Turkey uses military to force a plane from Russia to land in Ankara.
Putting it all together:

[p1] E1 (plane) [a1] carried [a2] from [a3] to [a4]

[a1] land in [a2]  

[a1] intercepted [a2]  
[a2] grounded  
[a1] forced [a2] to [a3]

[a2] land in [a3]

[a2] land in [a4]

[a1] land in [a5]
Putting it all together:

1. **E1 (plane)**
   - **a1**
   - **E1 (plane)**
   - **p1**
   - [a1] carried [a2] from [a3] to [a4]
   - **E2 (ammunition)**
   - **E3 (ankara)**

2. **E6 (Turkey)**
   - **E6 (Turkey)**
   - **E7 (Turkish Military)**
   - **E5 (Syria)**
   - **E7 (Moscow)**
   - **E4 (Russia)**

3. **p3**
   - [a1] intercepted [a2]
   - [a2] grounded
   - [a1] forced [a2] to [a3]

4. **p2**
   - [a1] land in [a2]
   - **E1 (plane)**
   - **E2 (ammunition)**
   - **E1 (plane)**
   - **E3 (ankara)**

5. **E1 (plane)**
   - **p2**
   - [a1] land in [a2]
   - **E1 (plane)**
   - **E1 (plane)**
   - **E3 (ankara)**
Summary: building on (mostly) known tasks

1. Extracting extended **predicate-argument structures**
2. **Coreference chains** for entities and predicates
3. **Merging** co-referring propositions
4. **Entailment graphs** for entities/arguments/predicates

⇒ *Can be seen as a cross-text semantic representation* - extending Open-IE / QA-SRL to the cross-text level
Fulfilling the Requirements
Fulfilling requirements - merging

Original propositions:

1. Plane carried ammunition from Russia
2. Jet carried ammunition to Syria

Generated propositions:

- Plane carried ammunition from Russia
- Jet carried ammunition to Syria
- Plane carried ammunition to Syria
- Jet carried ammunition from Russia
- Jet carried ammunition from Russia to Syria
Fulfilling requirements - modeling redundancy

Generated sentences:
1. Turkish military forces jet to land
2. Plane intercepted by Turkey

Entailments:
- plane $\Rightarrow$ jet
- Turkish military $\Rightarrow$ Turkey
- Forces to land $\Rightarrow$ intercepted

$>>$ Turkish military forces jet to land $\Rightarrow$
Plane intercepted by Turkey
Turkey intercepted a Syrian plane in Ankara.

The plane carried ammunition from Russia to Syria.

Turkey Escalates Confrontation with Syria.

Russia angry at Turkey.
Potential Applications - Content Presentation

- **Interactive text exploration** - “What is said here?”
  - Dynamic - salience, detail level, length
    - F-16 vs. Turkish army vs. Turkey
  - Hierarchical
  - Incremental
  - Bullet style vs. continuous text

- Improving abstractive multi-document summarization
Potential Applications - Querying

- Querying applications
  - Complex questions, requiring aggregated information from multiple texts
  - E.g.: *Grounding Syrian planes*

- Combining queries and content presentation: show all information about
  - An entity
  - Relations between entities
What happened between Turkey and Syria?

- **E1 (plane)**
  - **E5 (Syria)**
    - **p1**
      - [a1] carried [a2] from [a3] to [a4]
    - **a1**
    - **E2 (ammunition)**
    - **E3 (Moscow)**
      - **E4 (Russia)**
    - **a2**
    - **a3**
    - **a4**
  - **a1**
  - **E7 (Turkish Military)**
    - **E6 (Turkey)**
      - **p2**
        - [a1] land in [a2]
      - **a1**
      - **a2**
    - **E1 (plane)**
      - **E6 (Turkey)**
        - **p3**
          - [a1] intercepted [a2]
          - [a2] grounded to [a3]
        - **a1**
        - **a2**
        - **a3**
  - **E1 (plane)**
    - **E6 (Turkey)**
      - **E7 (Turkish Military)**
        - **E5 (Syria)**
Graphical Tools
Manual Annotation Tool - Entities & Propositions
Manual Annotation Tool - Entailments
Future Enhancements & Conclusion
Lots of things to add

- Contradiction & factuality (partly in 1st version)
- Modeling sub-events - partial argument alignment
  - ISIS attacks Syria, 5 killed - shared agent for attacking and killing
- Non-verbatim lexical inference
  - For generalization - in presentation and in queries
  - E.g. Russia criticized the grounding, Which countries criticized the grounding?
- Inter-proposition relations
  - Causal, temporal, ...
- Tying to general knowledge bases
- ...
Conclusion

- A consolidated representation of propositions across texts
- Modeling:
  - Coreference-based inference (merging)
  - Information redundancy/overlap via textual entailment
- Yielding a cross-text extension of Open-IE / QA-SRL representations
- Driving application - interactive text exploration
- Potential for semantic applications over aggregated content
- Lots of ways to go...  
  thank you!
References


- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. **Abstract meaning representation for sembanking.** ACL 2013.


- Agata Cybulska and Piek Vossen. **Using a sledgehammer to crack a nut? Lexical diversity and event coreference resolution.** LREC 2014.

- Jeffrey Flanigan, Sam Thomson, Jaime Carbonell, Chris Dyer, and Noah A. Smith. **A discriminative graph-based parser for the abstract meaning representation.** ACL 2014.


Unused slides
What happened between Turkey and Syria?

- **E1**: Plane
- **E6**: Turkey
- **E5**: Syria
- **E2**: Ammunition
- **E3**: Ankara
- **E4**: Moscow
- **E7**: Turkish Military
- **E4**: Russia

- **p1**: [a1] carried [a2] from [a3] to [a4]
- **p2**: [a1] land in [a2]
- **p3**: [a1] intercepted [a2]
- [a1] forced [a2] to [a3]
- [a2] grounded
- [a1] forced [a2] to [a3]
- [a1] land in [a2]
- [a1] land in [a2]
- [a2] in [a3]
- [a1] intercepted [a2]

- **E1** (plane)
- **E6** (Turkey)
- **E5** (Syria)
E1 (plane) [a1] carried [a2] from [a3] to [a4]

E2 (ammunition) [a1] carried [a2] from [a3] to [a4]

E5 (Syria) [a1] intercepted [a2] [a2] grounded [a1] forced [a2] to [a3]

E1 (plane) [a1] land in [a2]

E7 (Moscow) [a1] land in [a2]

E4 (Russia) [a1] land in [a2]

E6 (Turkey) [a1] land in [a2]

E3 (ankara) [a1] land in [a2]

E7 (Turkish Military)
Expressiveness of semantic structure

Level of multi-text consolidation

Comprehensive predicate-argument representation

AMR

QA- SRL

SRL

Open IE

Limited predicate-argument structure

PARMA

Roth and Frank

NLKG (our proposal)

ERE

ECB+

EECB

ECB

Entity coreference

One sentence

2 documents alignment

multi document coreference

Level of multi-text consolidation
Takeout: Characteristics of Textual Information

- Identifying and modeling **redundancy**
  - Plane to Syria ⇔ Syria-bound plane
  - forced plane to land ⇒ intercepted plane
  - jet ⇒ plane

- Merging **complementary scattered info** about the same entity or event
  - “Turkey forced a jet from Russia to land”
    - “Turkish military intercepted a Syria-bound plane”
    ⇒ “Turkish military forced a Syria-bound jet from Russia to land”
Multi-document Summarization

- Focusing only on salient sentences
- Identifying redundant sentences
  - Typically selecting one representative sentence
- Abstractive summarization
  - Aims to merge some parts of different sentences - hasn’t matured
- Heuristic - Vs. systematic semantic representations
Fulfilling requirements - merging

Original sentences:

1. Plane carried ammunition from Russia
2. Jet carried ammunition to Syria

Generated sentences:
Fulfilling requirements - merging

Original sentences:
1. Plane carried ammunition from Russia
2. Jet carried ammunition to Syria

Generated sentences:
❖ Plane carried ammunition from Russia
Fulfilling requirements - merging

Original sentences:
1. Plane carried ammunition from Russia
2. Jet carried ammunition to Syria

Generated sentences:
❖ Plane carried ammunition from Russia
❖ Jet carried ammunition to Syria
❖ Plane carried ammunition to Syria
Fulfilling requirements - merging

Original sentences:
1. Plane carried ammunition from Russia
2. Jet carried ammunition to Syria

Generated sentences:
1. Plane carried ammunition from Russia
2. Jet carried ammunition to Syria
Fulfilling requirements - merging

Original sentences:
1. Plane carried ammunition from Russia
2. Jet carried ammunition to Syria

Generated sentences:
❖ Plane carried ammunition from Russia
❖ Jet carried ammunition to Syria
❖ Plane carried ammunition to Syria
❖ Jet carried ammunition from Russia

Diagram:
- Plane (E1) carried ammunition (a1) to Syria (E5)
- Jet (E1) carried ammunition (a1) from Russia (E4)
- E4 (Russia) to E5 (Syria)
- E2 (ammunition) carried from E3 (plane) to E4 (Russia)
- E3 (plane) carried to E1 (plane)
Fulfilling requirements - Supporting Mentions

- Support Tracking & Frequency
  - Salience
    - “Turkey attacks Syria” vs “Turkish PM did not apologize”
  - Certainty
    - Frequently mentioned facts are more likely to be true.