Knowledge Representation from Text

Lecture Set 15

Professor Dan Moldovan

Outline

- **Knowledge Representation**
  - Layered approach
    - Davidsonian Logic Forms
    - Davidsonian Logic with Semantic Relations
  - Applications

Scenario

The train stood at the Amtrak station in Washington DC at 10:00 AM on March 15, 2005.
The name of the station was Union Station.
The train was an Acela Express, number 176.
The number 176 was scheduled to depart for New York City at 10:30 AM and arrive at 1:30 PM on March 15.
John arrived by taxi at Union Station at 10:15 AM.
John boarded the Acela at 10:20 AM and handed the train ticket to the conductor.
The conductor punched the ticket.
John sat by the window.
The train left the station on time.

Motivating Example

- Scenario:

  "…

  The train was an Acela Express, number 176.

  The 176 was scheduled to depart for New York City at 10:30 AM and arrive at 1:30 PM on March 15.

  …

  The train left the station on time."

- Question:

  "What time did the train depart?"

- **How do we represent and reason over this?**

General Approach

Keyword based approach is not good enough for many applications

Possible approach

- Knowledge representation of a text corpus KT1
- Knowledge representation of questions, or other smaller texts KT2
- Build ontologies for the domains of interest
- Reason by performing semantic similarity between KT1 and KT2. Use domain ontologies and methods from AI such as logic provers, subsumption, textual entailment, etc.

Approach 1: Davidsonian logic forms

Logic form representation

- Intermediate form between syntactic parse and deeper semantic meaning
- Preserves syntactic links between words in the sentence
  - subject and object
  - prepositional attachment
  - complex nominals
  - adjective and adverb connections

Logic Form Transformation

Logic form of a sentence is an intermediary step between syntactic parse and the deep semantic form.

Advantages of LFT over parsed format:

- LF codification acknowledges syntax-based relationships such as:
  - syntactic subjects
  - syntactic objects
  - prepositional attachments
  - complex nominals
  - adjectival/adverbial adjuncts
- LF is preferred when it comes to reasoning and other logic manipulations in knowledge bases.

Approach to Implement LFT

Criteria:

- Notation be as close as possible to English.
- Notation should be syntactically simple.

Approach:

- Derive LFT directly from the output of the syntactic parser. Parser resolves the structural and syntactic ambiguities.
- Ignore:
  - plurals and sets,
  - verb tenses
  - auxiliary verbs
  - comparatives
  - determiners
- By relaxing the logic formation we can have an effective representation closer to English.

LFT Definitions

Predicates:

- A predicate is generated for every noun, verb, adjective or adverb in the sentence. The name of a predicate includes the base form, and part-of-speech.

  Example:

  A learner is enrolled in an educational institution.

  has predicates: (learner:n, enroll:v, educational_institution:n)
- All verb predicates, as well as the nominalizations representing actions, events or states have three arguments:
- Action/static/event_predicate (e_i, x_1^i, x_2^i), where:
  - e_i represents the eventuality of the action, state or event i
  - x_1^i represents the syntactic subject of the action, event or state
  - x_2^i represents the syntactic direct object of the action, event or state.

LFT Definitions

A person who backs a politician.

person:n(x_1) & backs:v(e_1,x_1,x_2) & politician:n(x_2)

When the predicate is a bitransitive verb:

- verb(e_i,x_1^i,x_2^i,x_3^i)

Professor gives students the grades.

professor:n(x1)& give:v(e_1,x_1,x_2,x_3) & grade:v(x_2) & student:n(x_3)

- $\mathbf{x_3^i}$ represents the syntactic indirect object of the action, event, or state.

Slot allocation representation:

- The arguments of the verb predicates are always in the order: subject, direct object, indirect object.
- The position of the arguments is fixed - for the purpose of simpler notation.
- The arguments for the subject and direct objects are always present even when verb does not have these syntactic roles.
- Argument for indirect object occurs only if necessary.

LFT Definitions

Modifiers:

- Predicates generated from modifiers share the same arguments with the predicates corresponding to the phrase heads.
  - adjectives have same predicates as nouns

    a man-made object

    object:n(x_1)*& man-made:a*(x_1)

  - adverbial predicate is the eventuality of the verb it modifies.

    run quickly

    run:v(e_1,x_1,x_2)*& quickly:r*(e_1)

LFT Definitions

Conjunctions:

- Conjunctions are transformed in predicates.
- Conjunction predicates have a variable number of arguments.

An achievement demonstrating great skill or mastery

achievement:n(x_1)& demonstrate: v(e_1,x_1,x_2)& or(x_2,x_3,x_4)& skill:n(x_3)& great:a(x_3)& mastery:n(x_4)

Roll and turn skillfully

and(e_1,e_2,e_3)& roll:v(e_2,x_1,x_2)& turn:v(e_3,x_1,x_2)& skillfully:r(e_1)

An unintentional but embarrassing blunder

blunder:n(x_1)& but(x_1,x_2,x_3)& unintentional:a(x_2)& embarrasing:a(x_3)

LFT Definitions

Prepositions:

- Every preposition is a predicate with two arguments: the first argument corresponding to the predicate of the head of the phrase, to which PP attaches, and the second argument corresponds to prepositional object.

  Deprive of value for payment

  deprive:v(e_1,x_1,x_2) & *of*(e_1,x_3) & *value:n*(x_3)& *for*(x_3,x_4) & *payment:n*(x_4)

  Playing the position of pitcher on a baseball team

  playing:v(e_1,x_1,x_2) & *position:n*(x_2) & *of*(x_2,x_3)& *pitcher:n*(x_3) & *on*(e_1,x_4) & *baseball_team: n*(x_4)

LFT Definitions

- **Complex nominals**:
  - A new predicate nn is introduced to link together the collocating nouns.
  - nn has a variable number of arguments, the first representing the result of aggregation of the nouns, the rest one for each noun.

An organization created for business ventures.

organization:n(x_2) & *create*(e_1,x_1,x_2) & *for*(e_1,x_3) & *nn*(x_3,x_4,x_5) & *business:n*(x_4) & *venture:n*(x_5)

Government income credited to taxation.

nn(x_2,x_3,x_4) & *government:n*(x_3) & *income:n*(x_4) & *credit:v*(e_2,x_1,x_2) & *to*(e_1,x_5) & *taxation:n*(x_5)

LFT Inter-Phrase Rulers

Implementation is done with transformation rules that take the parser output into a LF. Every parser rule translates into a LFT rule. IntraPhrase transformation rules generate predicates for every noun, verb, adjective or adverb. They assign the variables that describe the local dependencies.

$ART\ N \rightarrow n(x_1)$

a human being $human\_being{:}n(x_1)$

$ART\ ADJ_1 ADJ_2\ N \rightarrow n(x_1)\ \&\ adj1(x_1)\ \&\ adj2(x_1)$

A hard straight return

$return{:}n(x_1)\ \&\ hard{:}a(x_1)\ \&\ straight{:}a(x_1)$

$ART\ ADJ1\ AND\ ADJ2N \rightarrow n(x_1)\ \&\ adj_1(x_1)\ \&\ adj_2(x_1)$

A week and tremulous light

$Light{:}n(x_1)\ \&\ weak{:}a(x_1)\ \&\ tremulous{:}a(x_1)$

$V\ ADV \rightarrow v(e_1,x_1,x_2)\ \&\ adv(e_1)$

Cut open $cut{:}v(e_1,x_1,x_2)\ \&\ open{:}r(e_1)$

$ART\ N_1\text{'}sN_2 \rightarrow n_2(x_1)\ \&\ n_1(x_2)\ \&\ pos(x_1,x_2)$

A person's body

$body{:}n(x_1)\ \&\ person{:}n(x_2)\ \&\ pos(x_1,x_2)$

An Example

A game played with rackets by two or four players who hit a ball back and forth over a net that divides a tennis court

$game{:}n(x_2)$ & $play{:}v(e_1,x_1,x_2)$ & $with(e_1,x_3)$ &

$racket{:}n(x_3)$ & $by(e_1,x_1)$ & $or(x_1,x_3,x_4)$ & $two{:}n(x_3)$ &

$four{:}n(x_4)$ & $player{:}n(x_1)$ & $hit{:}v(e_2,x_1,x_5)$ &

$ball{:}n(x_5)$ & $back\_and\_forth{:}r(e_2)$ & $over(e_2,x_6)$ &

$net{:}n(x_6)$ & $divide{:}v(e_3,x_6,x_7)$ & $tennis\_court{:}n(x_7)$

Approach 2: Davidsonian Logic with Semantics

- Idea:

  Expand Davidsonian Logic Representation with Semantic Relations to create a deeper logic representation of text.

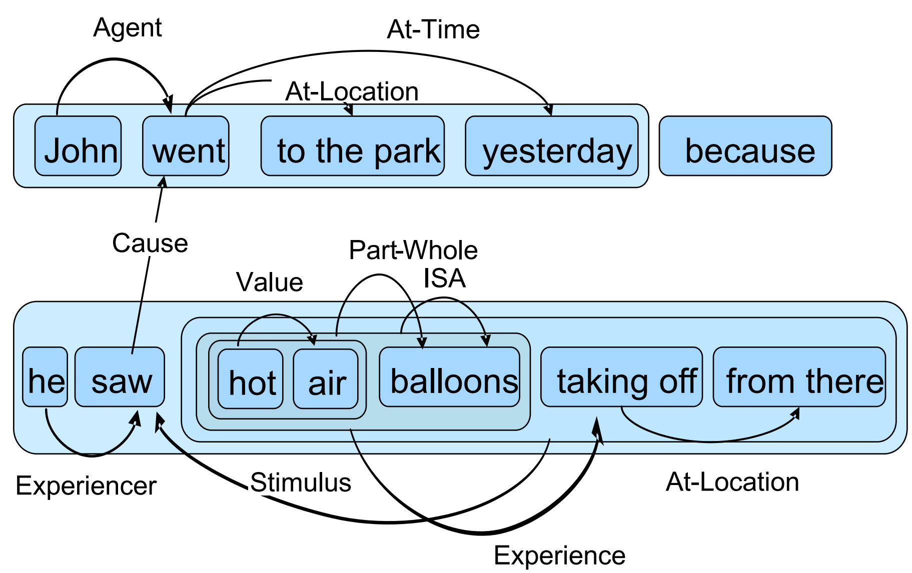
Semantic Relations in text

Semantic Relations are the interconnections between words or concepts that define the meaning of text. They are used as elements of knowledge bases.

Example:

John went to the park yesterday because he saw hot air balloons taking off from there

Agent(John, went)
At-Location(went, to the park)
At-Time(went, yesterday)
Cause(saw, went)
Experiencer(He, saw)
Stimulus(hot air balloons taking off from there, saw)
Value(hot, air)
Part-Whole(hot air, balloons)
Is-A(hot air balloons, balloons)
Experiencer(hot air balloons, taking off)
At-Loc(taking off, from there)



A Comprehensive list of Semantic Relations

Code	Relation	Definition	Example
POS	**Possession**	X is a possession of Y, Y owns/has X	*[YX] [John] owns [a Porsche];* *[YX] [John] has [4 acres]*
PW	**Part-Whole/ Meronymy**	X is a part of Y	*[XY] [The engine] is the most important part of [the car];* *[XY] [steel][cage];* *[YX] [faculty] [professor];* *[XY] [door] of the [car]*
KIN	**Kinship**	X is a kinship of Y; X is related to Y by blood or by marriage	*[XY] [John]'s [uncle]*
ASO	**Association**	X is associated with Y; X and Y can be people or groups	*[XY] [John] and [Mary] are friends for 20 years.* *[XY] [John] talked to [Mary] about her catering service.*
SRC	**Source/Origin**	X is the origin or previous location of Y	*[XY] [Chilean] [Sea Bass];* *[YX] [Student] from [Russia]*
ISA	**ISA**	X is a (kind of) Y	*[XY] [John] is a [person].*
SYN	**Synonymy/Name**	X is a synonym/name/equal for/to Y	*[XY] [FBI] ([Federal Bureau of Investigation])* *[YX] [This car] is called ["Johann"]*
PRO	**Property Type**	X is a property type of Y	*[XY] [The color] of [the car] is blue.*
VAL	**Property/ Attribute/ Value**	X is a property/attribute/value of Y	*[YX] [The car] is [blue]* *[YX] [The color] of the car is [blue].*
QNT	**Quantification/ Extent**	X is a quantification of Y; Y can be an entity or event	*[XY] [XY] John saw [three] [hurricanes] in the last [two] [years].* *[Y X]The budget [increased] with [10%]*

A Comprehensive list of Semantic Relations Cont.

Code	Relation	Definition	Example
AGT	**Agent**	X is the agent for Y; X is prototypically a person.	*[XY] [John] [eats] eggs and ham*
EXP	**Experiencer**	X is an experiencer of Y; involves cognition and senses; X is a person	*[XY] [John] [feels] bad*
INS	**Instrument**	X is an instrument in Y	*[YX] John [broke] the window with [a hammer].* *[XY] [The hammer] [broke] the window.* *[YX] John [played] the Brandenburg Concerto on [the harmonica]*
THM	**Theme/Patient/ Result/ Consumed**	X is the theme/patient/result/consumed in/from/of Y	*[YX] John [painted] [his truck].* *[YX] John [baked] [a cake].*
RCP	**Recipient/ Receiver**	X is the recipient of Y; X is an animated entity. The theme of received can be both positive and negative.	*[YX] John [gave] [Mary] roses.* *[YX] John [stole] [Mary]'s car.*
TPC	**Topic/ Content**	X is the topic/focus of cognitive communication Y	*[YX] John [talked] about [politics] with Mary.* *[YX] John [said] [he likes the other party].*
INT	**Intent**	X is the intent/goal/reason of Y	*[YX][YX] [John] wants to [finish the paper] so [he] can [go on vacation].*
STI	**Stimulus**	X is the stimulus of Y; Perceived thorough senses	*[YX] [YX] Mary [heard] [the train] while [smelling] [the roses].*

A Comprehensive list of Semantic Relations Cont.

Code	Relation	Definition	Example
MNR	**Manner**	X is the manner in which Y happens	*[YX] John [read] [carefully]; [ran] [quickly]; [spoke] [hastily]*
LOC	**At-Location/ Space/ Direction/ Source/ Path/ Goal**	X is at location y	*[XY] There is [a cat] on [the roof]* *[XY] The hurricane [passes] through [Galveston].*
TMP	**At-Time**	X is at time Y	*[XY] John [woke up] at [noon]*
CAU	**Cause**	X causes Y; X and Y are events, states	*[XY] [Drinking] causes [accidents].*
MAK	**Make-Produce**	X is a product of Y	*[YX] [GM] manufactures [cars].*
JST	**Reason/ Justification**	X is the reason/motive/justification for Y	*[XY] [The severity of the crime] justifies [the harsh sentence]; [YX] [He is innocent] by reason of [insanity]*
PRP	**Purpose**	X is the purpose for Y; Y did something because this person wanted X	*[YX] John [swims] for [fun]; Mary [works] part-time [to earn some extra money]*
IFL	**Influence**	X caused something to happen to Y	*[XY] [The war] had an impact on [the Economy]*

Example of semantic relations annotation

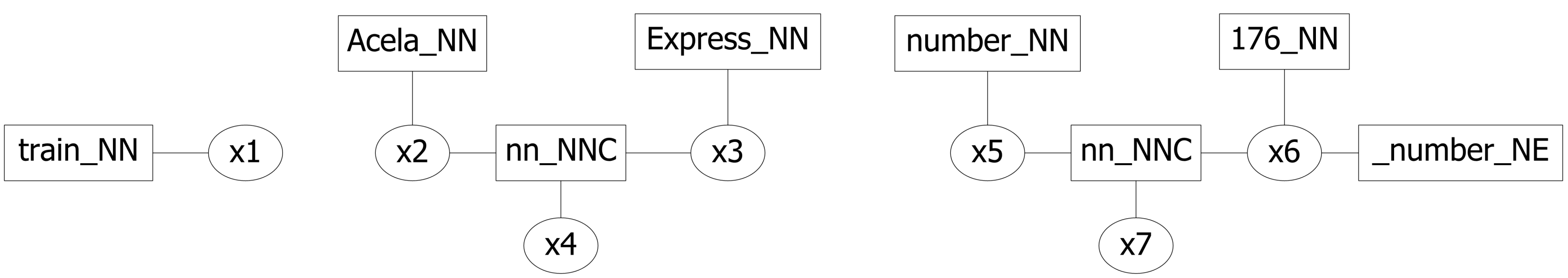
The president is concerned that the strike will undermine efforts by Palestinian authorities and others to bring an end to terrorist attacks and does not contribute to the security of Israel, Bush's spokesman, Ari Fleischer, told reporters.

EXP(The president, concerned)
TPC(the strike will undermine efforts, concerned)
EXP(the strike, undermine)
THM(efforts, undermine)
IFL(the strike, efforts by Palestinian authorities and others to bring an end to terrorist attacks)
PRP(bring an end to terrorist attacks, efforts)
SRC(Palestinian, authorities)
EXP(the strike, contribute)
THM(security of Israel, contribute)
IFL(the strike, the security of Israel)
ASO(Bush, spokesman)
SYN(Ari Fleischer, Bush's spokesman)
ISA(Ari Fleischer, spokesman)
AGT(Ari Fleischer, told)
RCP(reporters, told)
TPC(The president is concerned, told).

Logic form + semantics: Example 1

- Scenario:
  "…
  The train was an Acela Express, number 176.
  The 176 was scheduled to depart for New York City at 10:30 AM and arrive at 1:30 PM on March 15.
  …
  The train left the station on time."

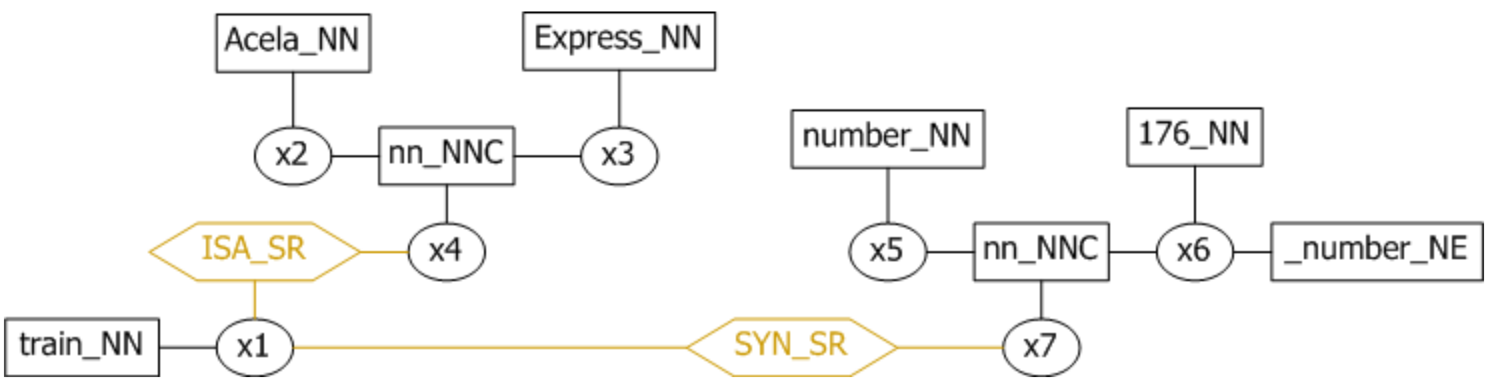
- Logic Form Representation:
  train_NN(x1) & Acela_NN(x2) & Express_NN(x3) & nn_NNC(x4,x2,x3) & number_NN(x5) & 176_NN(x6) & _number_NE(x6) & nn_NNC(x7,x5,x6)



Logic form + semantics: Example 1

- Scenario:
  "…
  The train was an Acela Express, number 176.
  The 176 was scheduled to depart for New York City at 10:30 AM and arrive at 1:30 PM on March 15.
  …
  The train left the station on time."

- Logic Form + Semantic Relations:
  train_NN(x1) & Acela_NN(x2) & Express_NN(x3) & nn_NNC(x4,x2,x3) & number_NN(x5) & 176_NN(x6) & _number_NE(x6) & nn_NNC(x7,x5,x6) & ISA_SR(x1,x4) & SYN_SR(x7,x1)



Logic form + semantics: Example 2

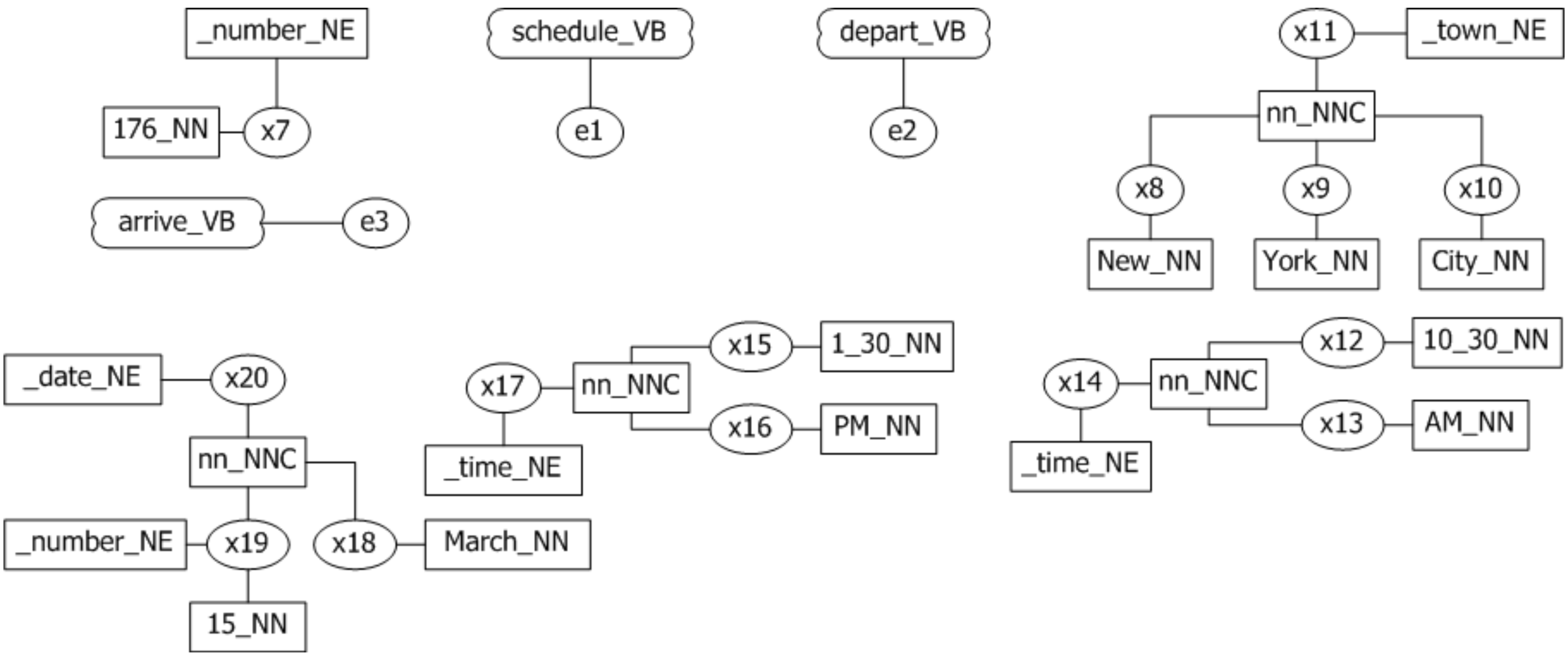
- Scenario:
  "…
  The train was an Acela Express, number 176.
  The 176 was scheduled to depart for New York City at 10:30 AM and arrive at 1:30 PM on March 15.
  …
  The train left the station on time."

- Logic Form Representation

  176_NN(x7) & _number_NE(x7) & schedule_VB(e1) & depart_VB(e2) & New_NN(x8) & York_NN(x9) & City_NN(x10) & nn_NNC(x11,x8,x9,x10) & _town_NE(x11) & 10_30_NN(x12) & AM_NN(x13) & nn_NNC(x14,x12,x13) & _time_NE(x14) & arrive_VB(e3) & 1_30_NN(x15) & PM_NN(x16) & nn_NNC(x17,x15,x16) & _time_NE(x17) & march_NN(x18) & 15_NN(x19) & _number_NE(x19) & nn_NNC(x20,18,19) & _date_NE(x20)

Logic form + semantics: Example 2

The 176 was scheduled to depart for New York City at 10:30 AM and arrive at 1:30 PM on March 15.



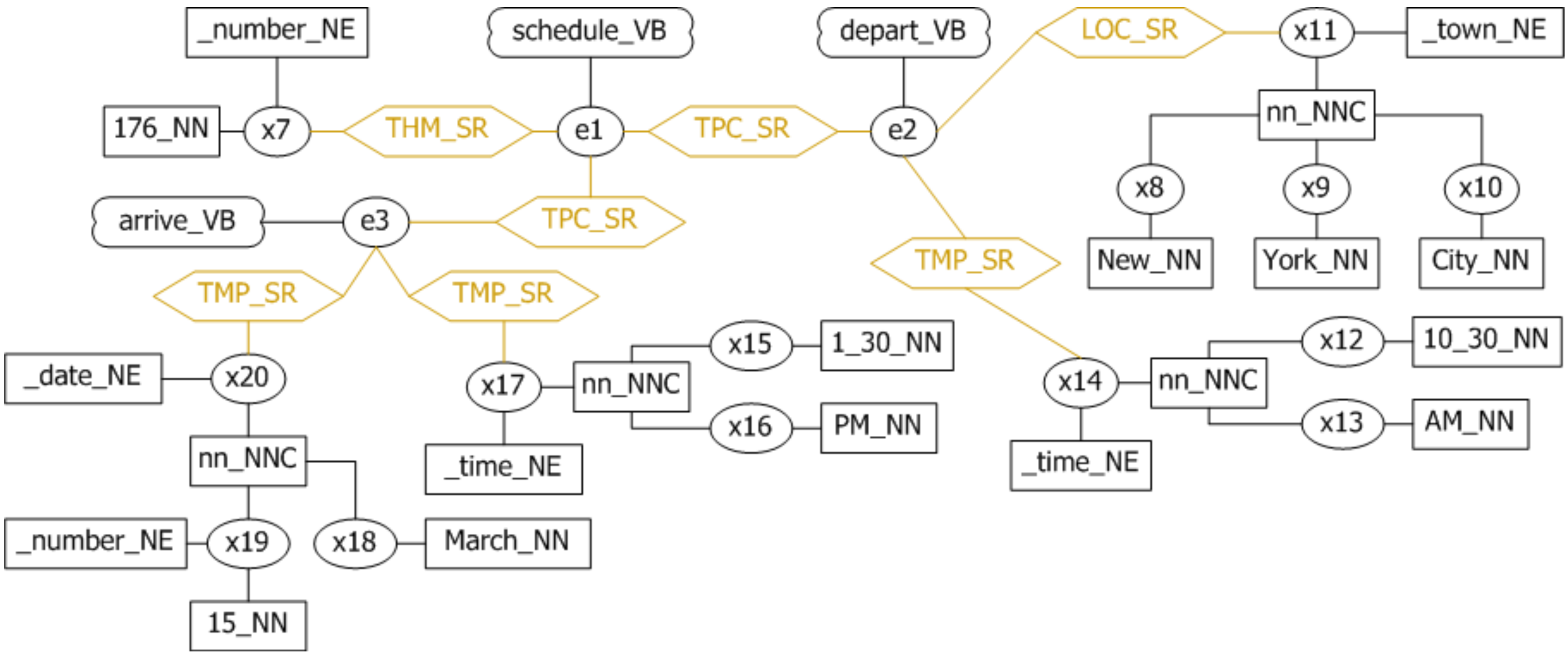
Logic form + semantics: Example 2

- Scenario:
  "…
  The train was an Acela Express, number 176.
  The 176 was scheduled to depart for New York City at 10:30 AM and arrive at 1:30 PM on March 15.
  …
  The train left the station on time."

- Logic Form + Semantic Relations:
  176_NN(x7) & _number_NE(x7) & schedule_VB(e1) & depart_VB(e2) & New_NN(x8) & York_NN(x9) & City_NN(x10) & nn_NNC(x11,x8,x9,x10) & _town_NE(x11) & 10_30_NN(x12) & AM_NN(x13) & nn_NNC(x14,x12,x13) & _time_NE(x14) & arrive_VB(e3) & 1_30_NN(x15) & PM_NN(x16) & nn_NNC(x17,x15,x16) & _time_NE(x17) & march_NN(x18) & 15_NN(x19) & nn_NNC(x20,x18,x19) & _date_NE(x20) & THM_SR(x7,e1) & TPC_SR(e2,e1) & LOC_SR(x11,e2) & TMP_SR(x14,e2) & TPC_SR(e3,e1) & TMP_SR(x17,e3) & TMP_SR(x20,e3)

Logic form + semantics: Example 2

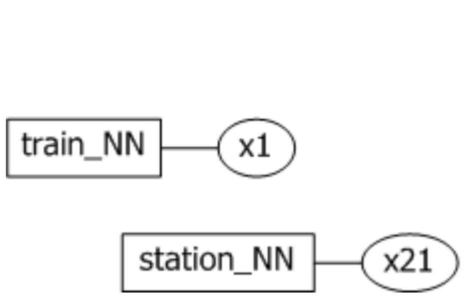
The 176 was scheduled to depart for New York City at 10:30 AM and arrive at 1:30 PM on March 15.



Logic form + semantics: Example 3

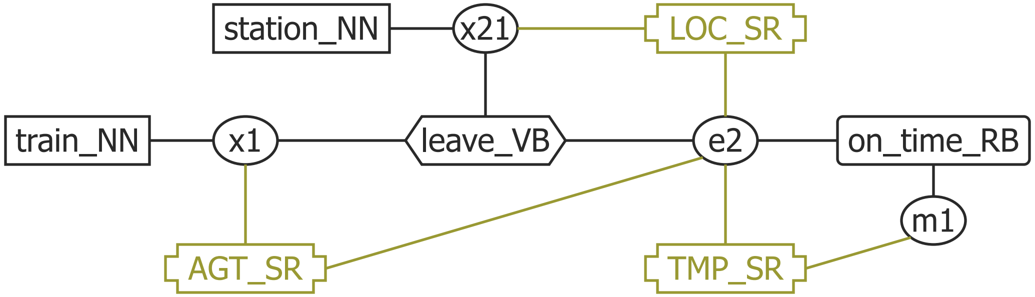
- Scenario:
  "…
  The train was an Acela Express, number 176.
  The 176 was scheduled to depart for New York City at 10:30 AM and arrive at 1:30 PM on March 15.
  …
  The train left the station on time."

- Logic Form Representation:
  train_NN(x1) & leave_VB(e4) & station_NN(x21) & on_time_RB(x22)



Semantic Relations Example

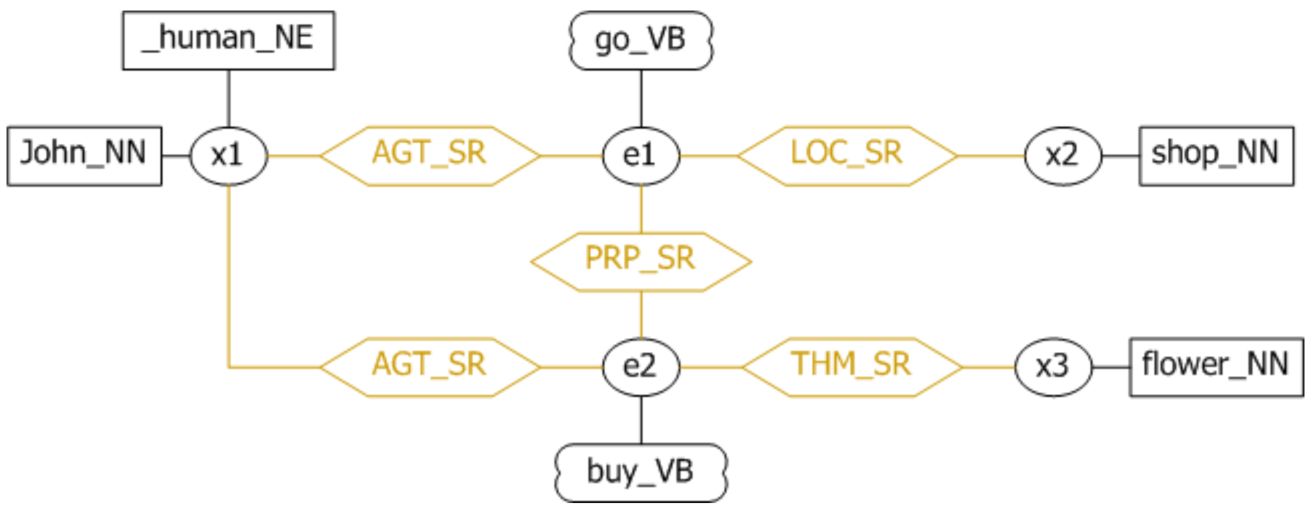
- Scenario:
  "…
  The train was an Acela Express, number 176.
  The 176 was scheduled to depart for New York City at 10:30 AM and arrive at 1:30 PM on March 15.
  …
  The train left the station on time."
- Logic Form + Semantic Relations:
  train_NN(x1) & leave_VB(e2,x1,x21) & station_NN(x21) & on_time_RB(m1,e2) & AGT_SR(x1,e2) & LOC_SR(x21,e2) & TMP_SR(m1,e2).



Logic form + semantics: Example 4

- John went to the shop to buy flowers.

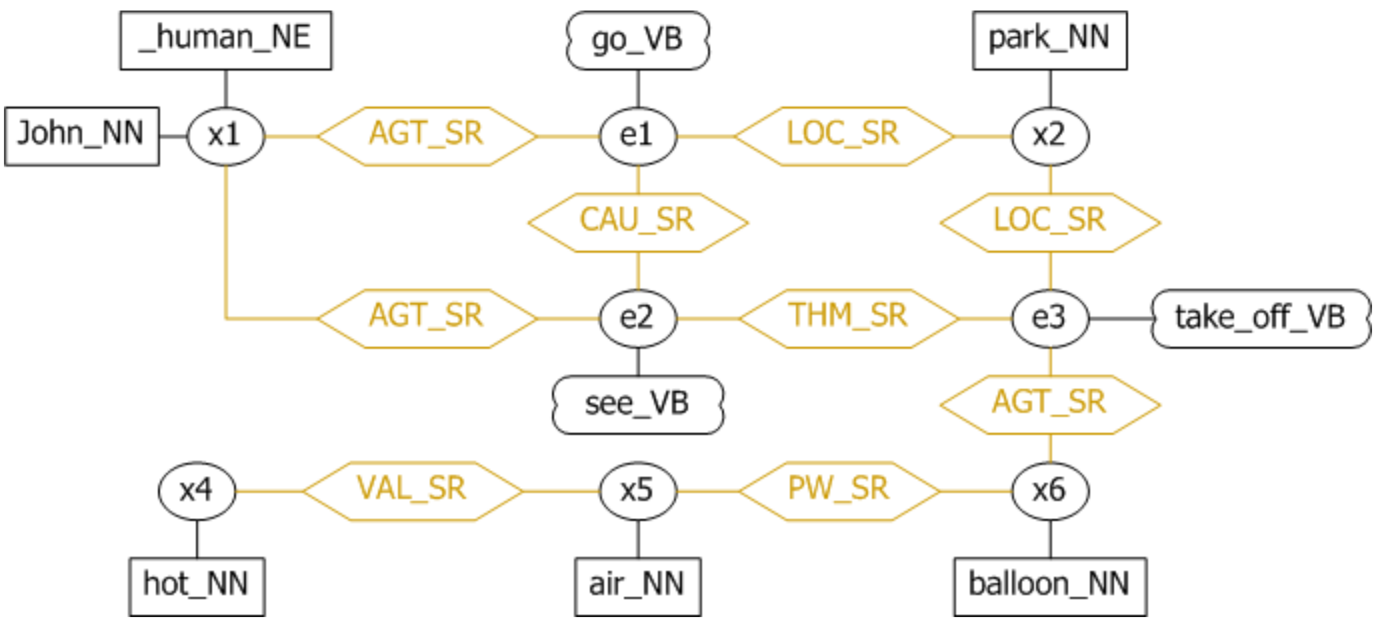
➡ Logic Form + Semantic Relations
John_NN(x1) & _human_NE(x1) & go_VB(e1) & shop_NN(x2) & buy_VB(e2) & flower_NN(x3) & AGT_SR(x1,e1) & LOC_SR(x2,e1) & AGT_SR(x1,e2) & PRP_SR(e2,e1) & THM_SR(x3,e2)



Logic form + semantics: Example 5

- John went to the park yesterday because he saw hot air balloons taking off from there.

➡ Logic Form + Semantic Relations
John_NN(x1) & _human_NE(x1) & go_VB(e1) & park_NN(x2) & yesterday_NN(x3) & see_VB(e2) & hot_JJ(x4) & air_NN(x5) & balloon_NN(x6) & take_off_VB(e3) & AGT_SR(x1,e1) & LOC_SR(x2,e1) & TMP_SR(x3,e1) & CAU_SR(e2,e1) & AGT_SR(x1,e2) & THM(e3,e2) & PW_SR(x5,x6) & VAL_SR(x4,x5) & AGT_SR(x6,e3) & LOC_SR(x2,e3)



Textual Entailment

- T: Toyota Ireland, through its Eurocare program, provides owners of both new and up to 3 year old Toyota models with the peace of mind of a comprehensive roadside assistance service.

  H: Eurocare is a program of roadside assistance service.

- T: Competition between IBM and Oracle is certainly nothing new, and the customer does ultimately benefit.

  H: IBM is a partner of Oracle.

- T: Canadian Nation Defense has been using virtual reality to train pilots and ground soldiers.

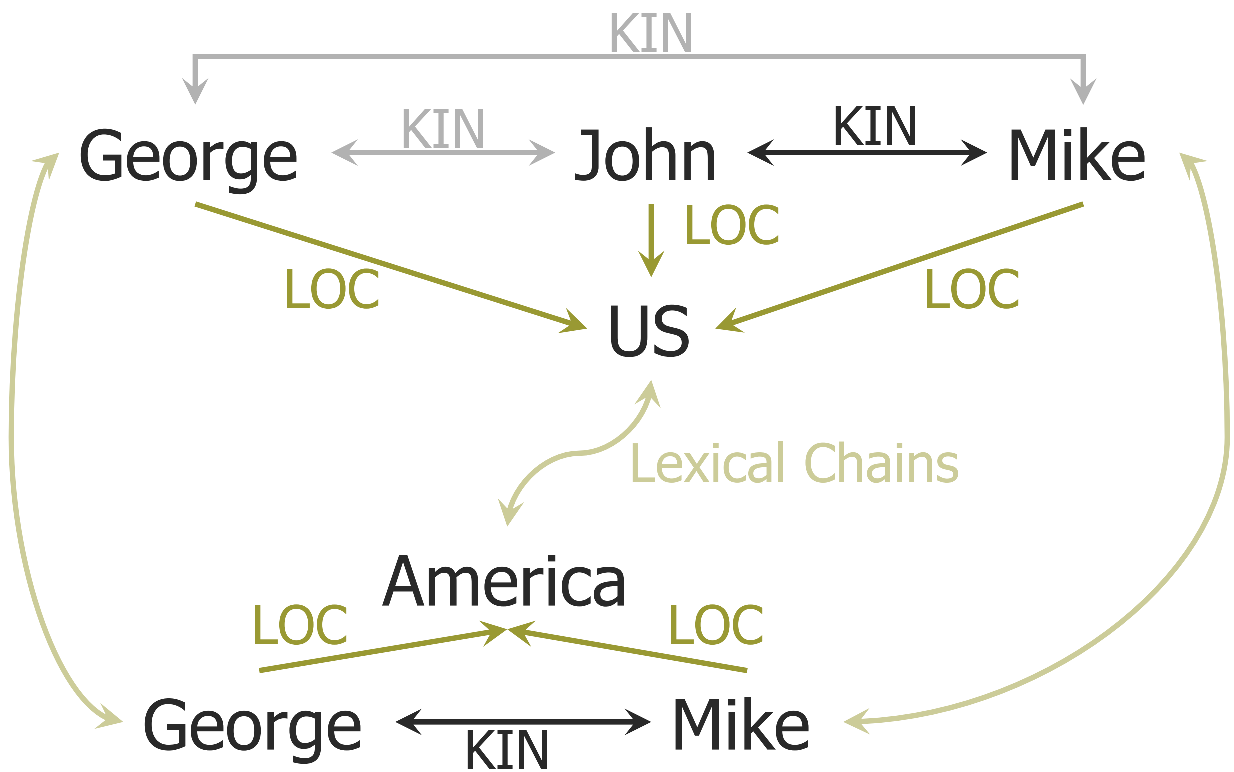
  H: Soldiers have been trained using virtual reality.

Semantic-based Logic Approach

- Textual Entailment
  - Task definition: *T entails H*, denoted by $T \rightarrow H$, if the meaning of *H* can be inferred from the meaning of *T*
  - inferred means logic (theorem prover + natural language axioms)
  - meaning means semantics (semantically enhanced knowledge representation)

Proof Sketch

T: John and his son, George, emigrated with Mike, John's uncle, to US in 1969.



H: George and his relative, Mike, came to America.