SWG Strategy

Fact Extraction using Controlled English and the English Resource Grammar

David Mott, ETS, IBM UK
Stephen Poteet, Ping Xue, Anne Kao, Boeing
Ann Copestake, University of Cambridge

DELPH-IN Summit 31st July 2013

(C) Copyright IBM Corp. 2006, 2013. All Rights Reserved.

Agenda

- **The ITA programme**
- **Controlled English**
- **Previous research into fact extraction, reasoning and NL processing**
- **New research using the resources from DELPH-IN**
- **Questions (mine)**

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

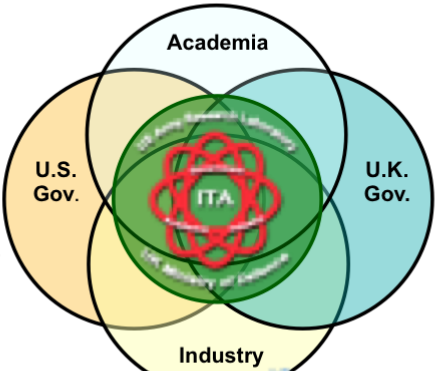
The ITA programme

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

ITA Programme

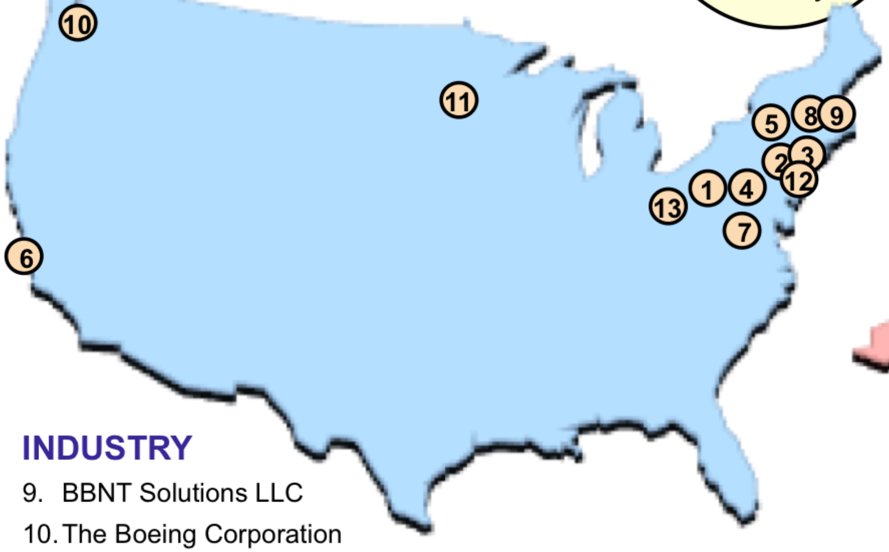
ACADEMIA

1. Carnegie Mellon University
2. City University of New York
3. Columbia University
4. Pennsylvania State University
5. Rensselaer Polytechnic Institute
6. University of California Los Angeles
7. University of Maryland
8. University of Massachusetts



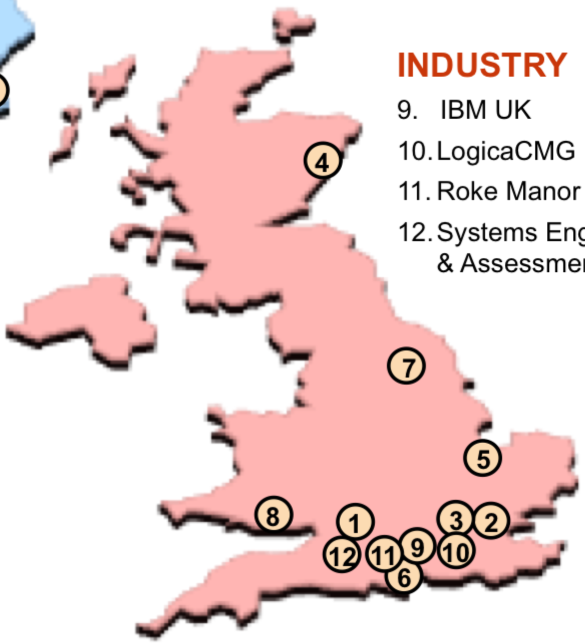
ACADEMIA

1. Cranfield University, Royal Military College of Science, Shrivenham
2. Imperial College, London
3. Royal Holloway University of London
4. University of Aberdeen
5. University of Cambridge
6. University of Southampton
7. University of York
8. University of Cardiff



INDUSTRY

9. IBM UK
10. LogicaCMG
11. Roke Manor Research Ltd.
12. Systems Engineering & Assessment Ltd.



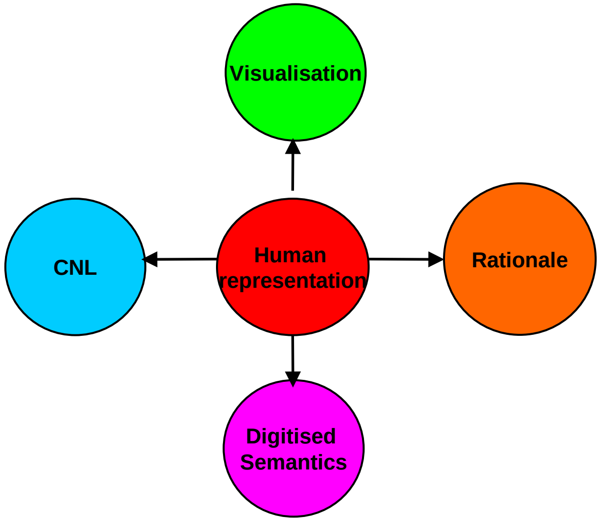
INDUSTRY

9. BBNT Solutions LLC
10. The Boeing Corporation
11. Honeywell Aerospace Electronic Systems
12. IBM Research
13. Klein Associates

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

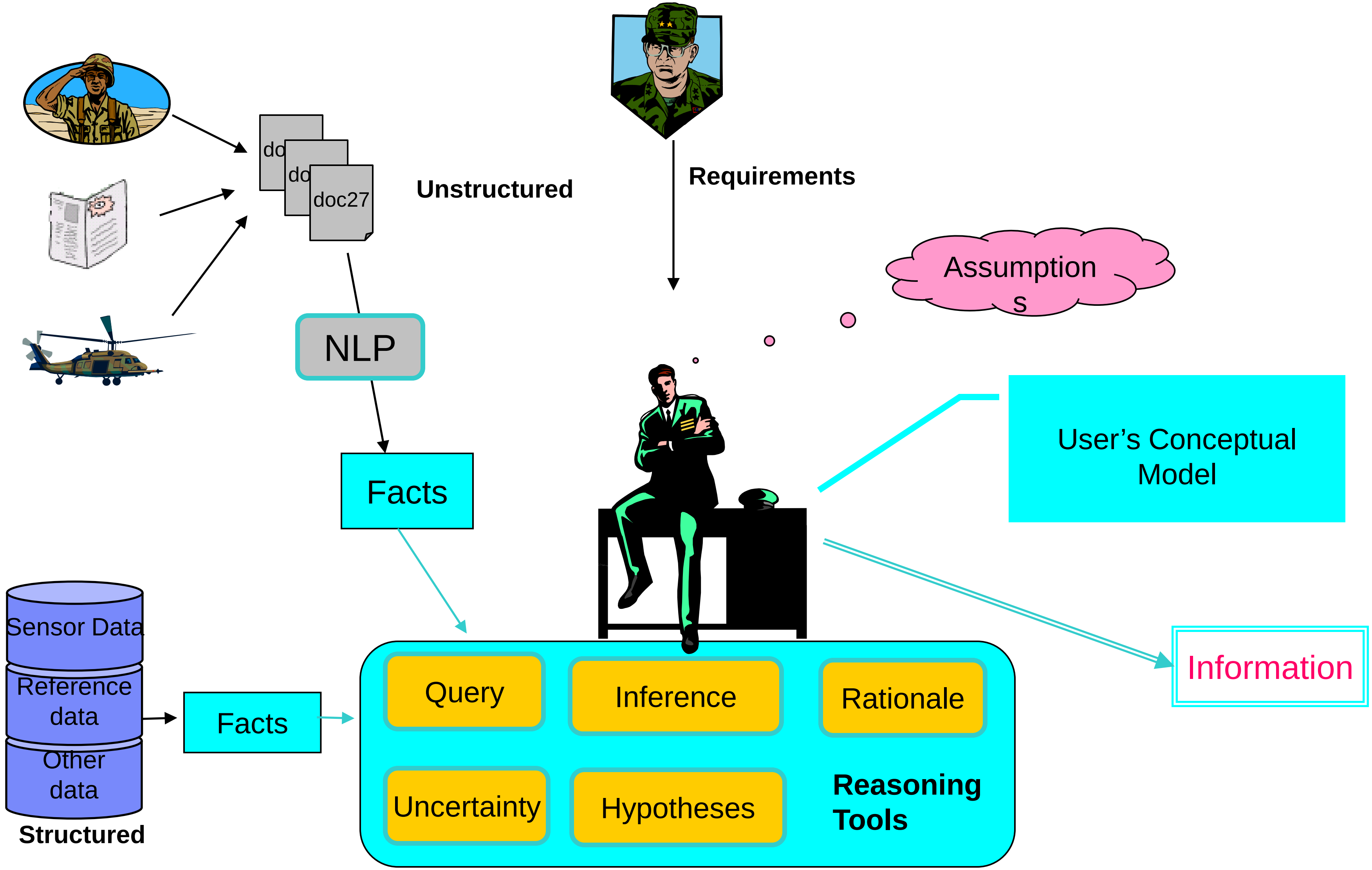
Some Research Issues

- **How do we assist people to create and use applications that reason?**
  - Modelling concepts, relationships and rules of inference
  - Grasping the basic logic of the model and rules
  - Understanding the reasoning performed by others
  - Sharing understanding across the human team
  - Sharing reasoning and results across different systems



(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Supporting the user



(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

User's "Conceptual Model"

- **User represents specialist knowledge as concepts, facts and rules for inference**
  - a conceptual model
  - a common set of concepts
- **The system must "understand" the conceptual model**
  - assist user to search for patterns, deduce information
- **A language to build the conceptual model**
  - user: easy to understand
  - system: readable, unambiguous and formal

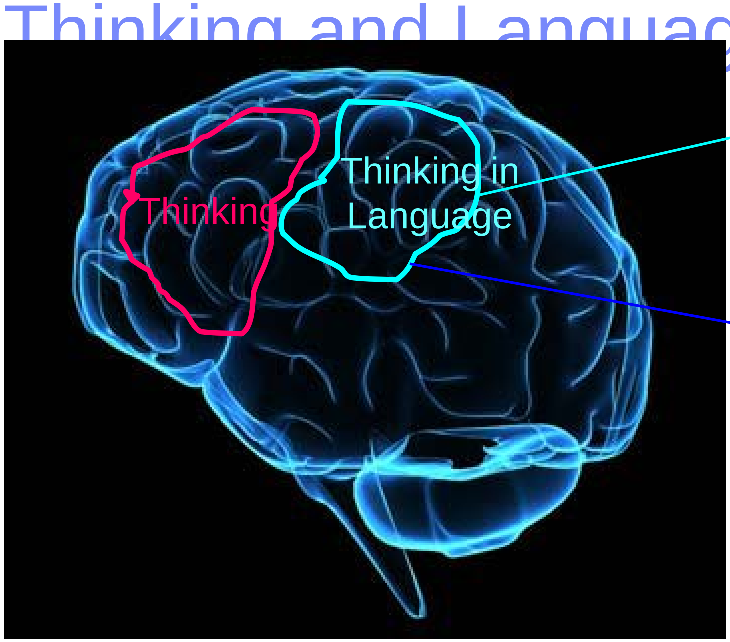
- **We use a Controlled Natural Language to express the model**

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Controlled English

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Thinking and Language



Photographer: Sebastian Kaulitzki | Agency: Dreamstime.com

http://health.howstuffworks.com/human-body/systems/nervous-system/brain-pictures.htm

Natural Language

Controlled Natural Language

“Thinking” in Language

Logic Prolog Java XML

Processing

We want thinking, communication and processing to be as integrated as possible

Based on work by John Sowa

***ITA Controlled English* is a Controlled Natural Language, being a subset of English that is both human readable and machine interpretable**

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Reasoning - How people might write facts

Family History	the man John is the parent of the woman Jean and is the sibling of the man James. the person James suffers from the disease migraine.
Patient	the woman Jean is a patient and presents with the symptom scotoma.
Medical Information	there is a visual symptom named scotoma. there is a neurological disease named migraine. there is a brain scan named mri. the disease XXX causes the symptom YYY.

This is how we input to the computer, and we want the answers in the same style

the brain scan mri is recommended for the woman Jean.

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Reasoning using logical rules

Family relations

```
if
   ( there is a person named ME ) and
   ( the person P is the parent of the person ME and
                    is the sibling of the man M )
then
   ( the man M is the uncle of the person ME ).
```

Medical relations

```
if
   ( the person P is the uncle of the patient PA )
then
   ( the person P is closely related to the patient PA ).
```

Disease - symptoms

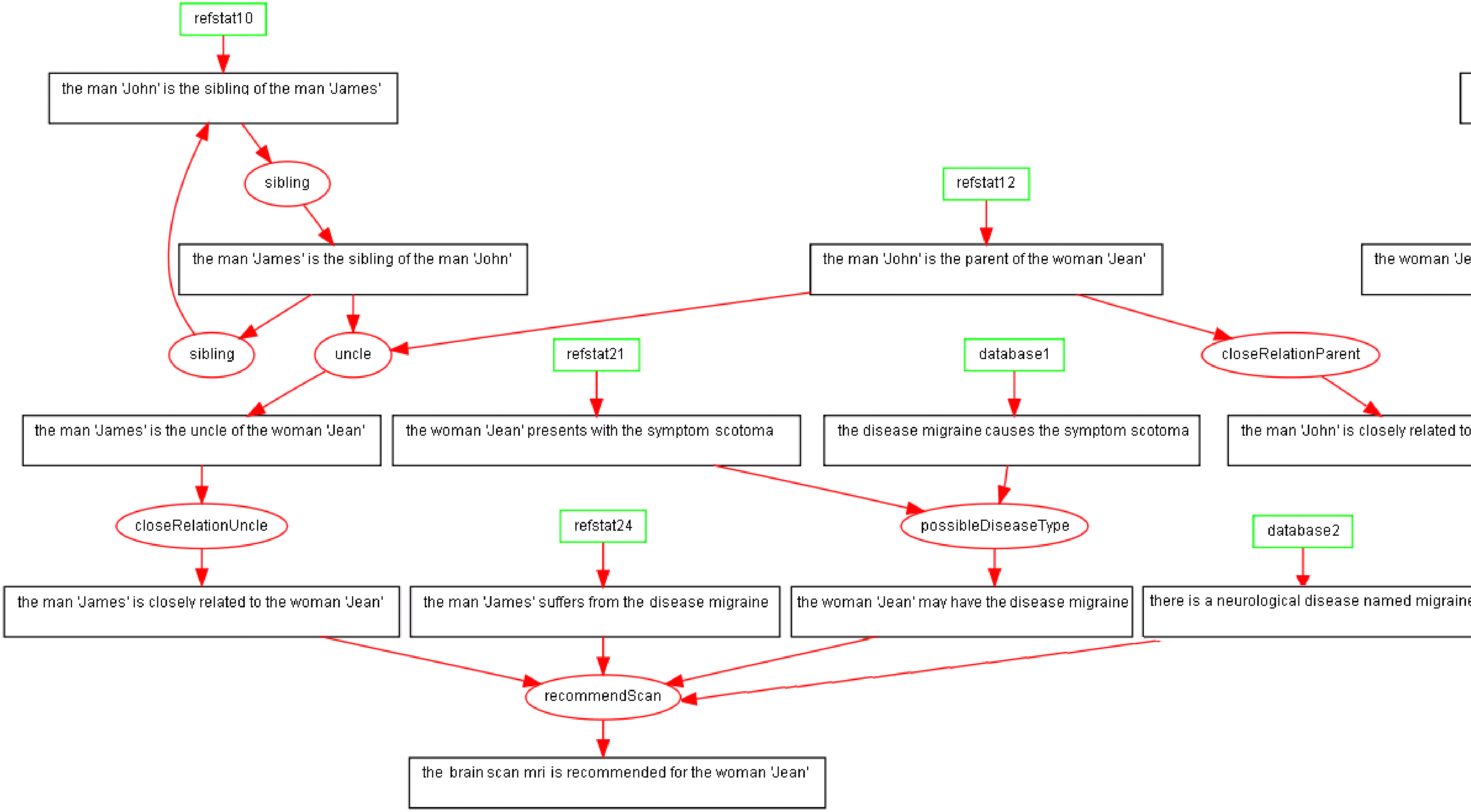
```
if ( the patient PA presents with the symptom S ) and
   ( there is a disease named C that causes the symptom S )
then
   ( the patient PA may have the disease C ).
```

Recommendations

```
if
   ( the patient PA may have the neurological disease C ) and
   ( the person R is closely related to the patient PA ) and
   ( the person R suffers from the neurological disease C ) and
   ( there is a brain scan named B )
then
   ( the brain scan B is recommended for the patient PA ).
```

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Explaining the reasoning



refstat10

the man 'John' is the sibling of the man 'James'

sibling

the man 'James' is the sibling of the man 'John'

sibling

uncle

refstat12

the man 'John' is the parent of the woman 'Jean'

the woman 'Je

closeRelationParent

refstat21

database1

the man 'James' is the uncle of the woman 'Jean'

the woman 'Jean' presents with the symptom scotoma

the disease migraine causes the symptom scotoma

the man 'John' is closely related to

closeRelationUncle

refstat24

possibleDiseaseType

database2

the man 'James' is closely related to the woman 'Jean'

the man 'James' suffers from the disease migraine

the woman 'Jean' may have the disease migraine

there is a neurological disease named migraine

recommendScan

the brain scan mri is recommended for the woman 'Jean'

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

More examples of facts

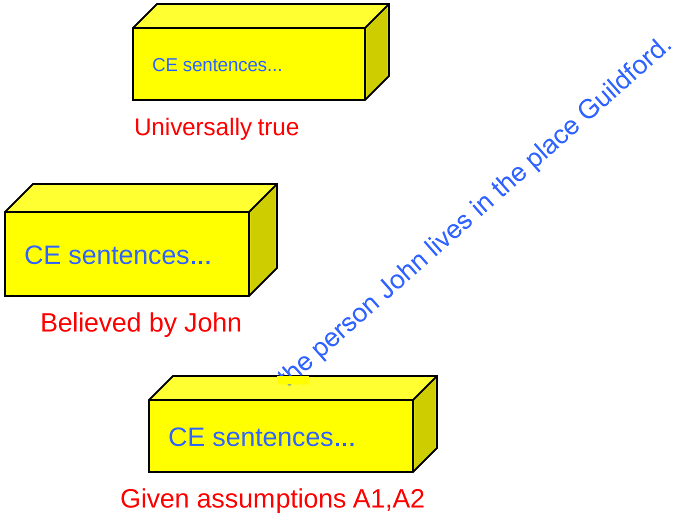
engineering	the oscillator osc2 connects to the filter f1.
planning	the task distribute_supplies is achieved after the task cross_bridge and has 8 as earliest start time. it is false that the water truck #10 is located at the bridge BR1.
influence analysis	it is assumed that the person 'John Smith' attends the meeting m1. the meeting m1 has the activity 'smuggle whisky' as topic.
crime data	the anti-social behaviour crime_002 is reported by the police force 'Hampshire Constabulary' and falls within the jurisdiction 'Hampshire Jurisdiction' and occurs during the month 2010-12.
Natural Language Processing	the NATO unit known as '|BCT patrol|' finds the facility #9. the facility #9 is located in the place known as |East Rashid|.

Also being researched for resource allocation, provenance, conversational interfaces, email toolkits

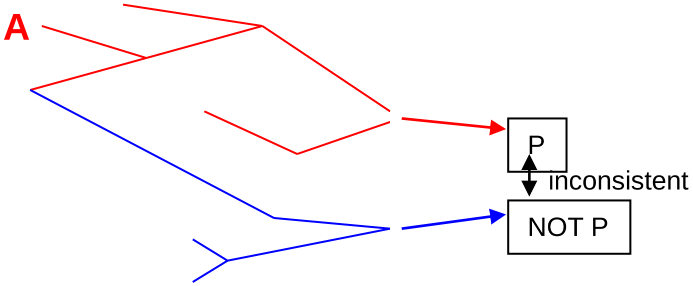
(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

CE Reasoning Capabilities

- **Reasoning with multiple views of truth:**
  - truth boxes
  - hypothetical reasoning based on assumptions
    - could be used to assist disambiguation of parses?



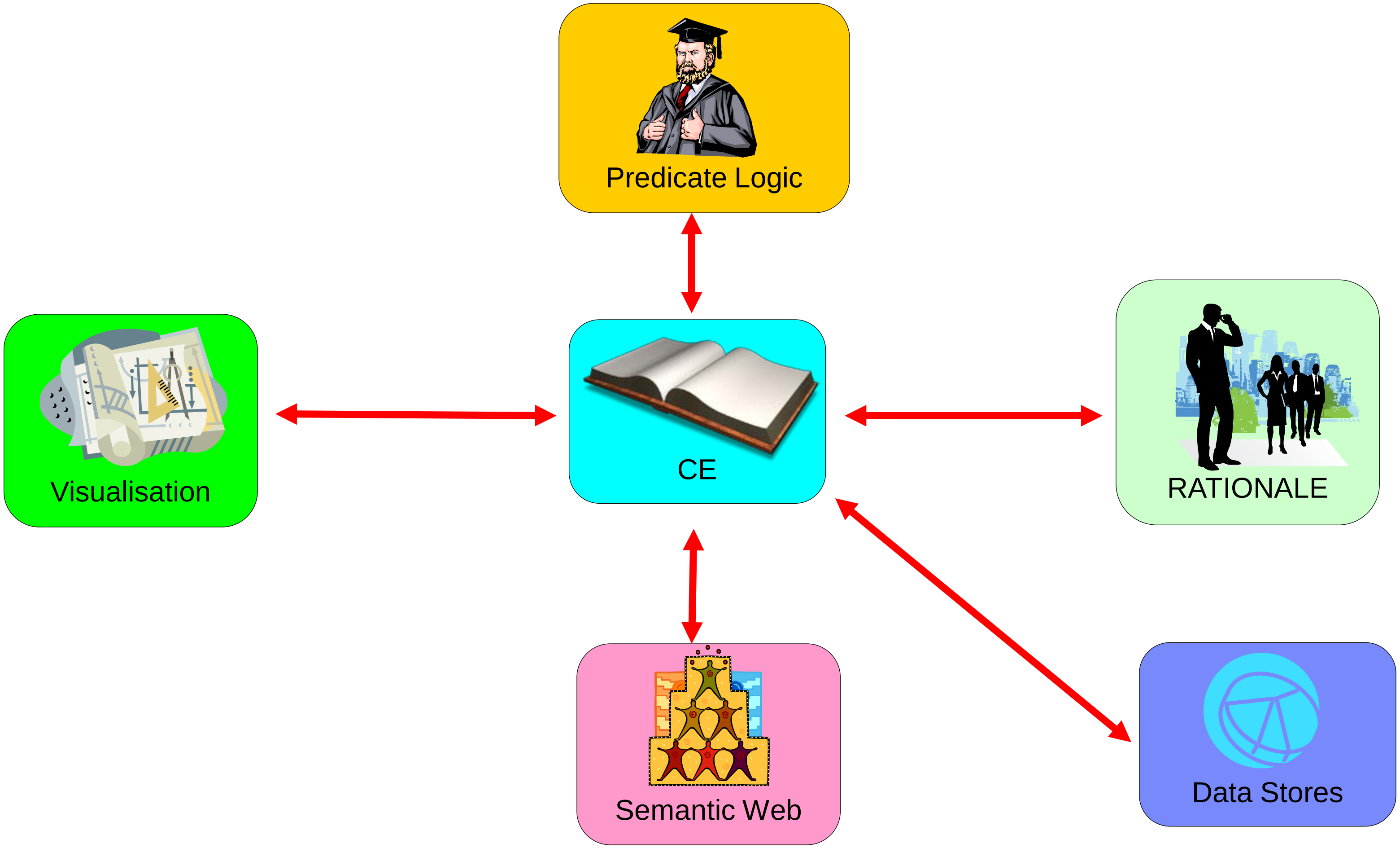
- **Reasoning with uncertainty**
  - propagation of uncertainty values through the rationale graphs
    - used to represent uncertain parsing and uncertain analysts reasoning



- **Interpretation of generic logical structures**
  - linguistic frames
    - more abstract view of linguistic reasoning?

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

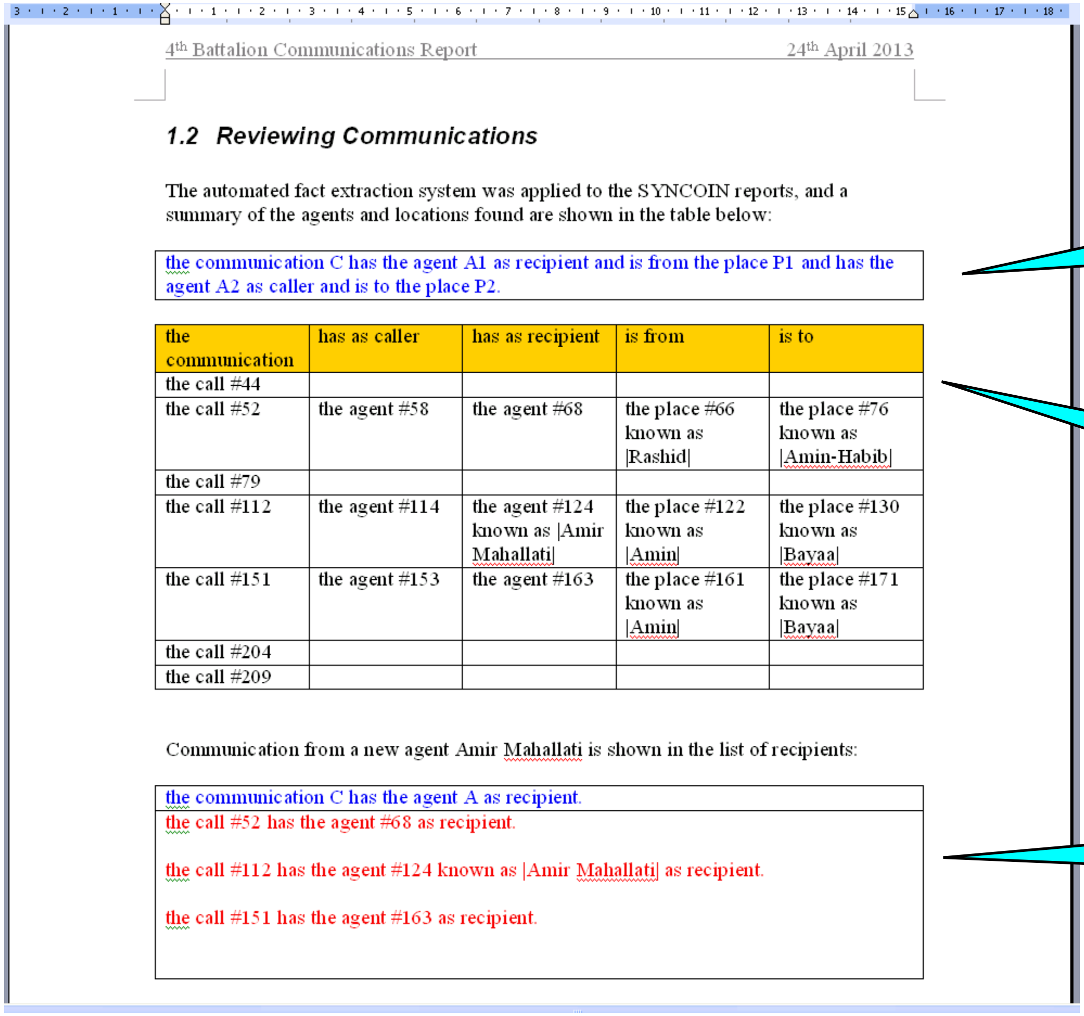
Formalising CE in other languages



Integration of different representations by sharing same semantics

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Embedding CE into Word documents



4th Battalion Communications Report 24th April 2013

1.2 Reviewing Communications

The automated fact extraction system was applied to the SYNCOIN reports, and a summary of the agents and locations found are shown in the table below:

the communication C has the agent A1 as recipient and is from the place P1 and has the agent A2 as caller and is to the place P2.

the communication	has as caller	has as recipient	is from	is to
the call #44				
the call #52	the agent #58	the agent #68	the place #66 known as |Rashid|	the place #76 known as |Amin-Habib|
the call #79				
the call #112	the agent #114	the agent #124 known as |Amir Mahallati|	the place #122 known as |Amin|	the place #130 known as |Bayaa|
the call #151	the agent #153	the agent #163	the place #161 known as |Amin|	the place #171 known as |Bayaa|
the call #204				
the call #209				

Communication from a new agent Amir Mahallati is shown in the list of recipients:

the communication C has the agent A as recipient.

the call #52 has the agent #68 as recipient.

the call #112 has the agent #124 known as |Amir Mahallati| as recipient.

the call #151 has the agent #163 as recipient.

Embedded Query

Result in tabular form

Result in sentence form

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

ITA Controlled English – what does it give us?

- **A Controlled Natural Language, being a subset of English**
  - limited syntax, but readable and writeable by humans
  - a formal semantics, so processable by machine
- **Provides a means of defining semantics**
  - general and domain specific
  - concepts and logical rules
- **A “virtual machine” for reasoning with the semantics**
  - storal, retrieval, querying
  - inference, simple and more complex

More than just a language

- **We are aiming to use ITA Controlled English as pervasively as possible**

But does need extending

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

2011-2013: Building the ideas for fact extraction, NL processing and reasoning

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Fact Extraction using Controlled Natural Language

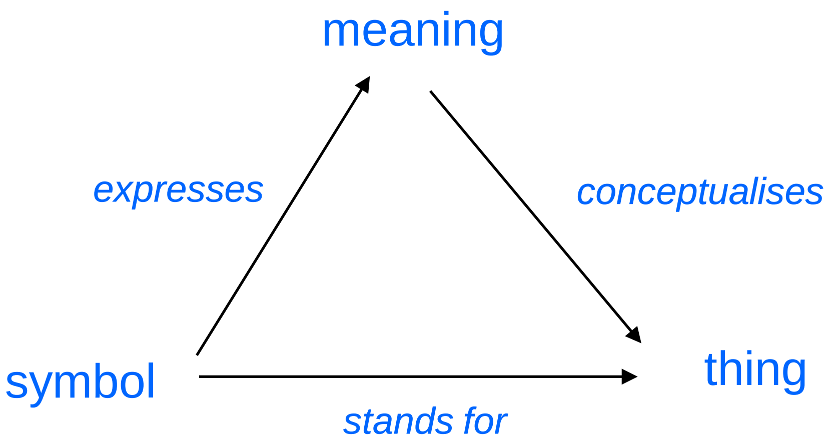
- **As the target of the NL processing**
  - facts in documents can be used for further reasoning
- **As a means of describing the NL processing**
  - to allow the user to understand the linguistic processing
  - to help configure NL tooling to the user’s specific domain

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Conceptual Model(s)

Meta Model	**Concept, Entity Concept, Relation Concept, Conceptual Model**	**belongs to, has as domain**
Semiotic Triangle	**Thing, Meaning, Symbol**	**stands for, expresses**
General	**Agent, Spatial Entity, Temporal Entity, Situation, Container**	**has as agent role, is contained in**
Linguistic	**Sentence, Phrase, Word, Noun, Linguistic Category, Linguistic Frame**	**has as dependent, is parsed from**
ACM	**Place, Church, Person, Village, IED, Facility, ...**	**is located in**

Want a common model of language in Controlled English



meaning

expresses

conceptualises

symbol

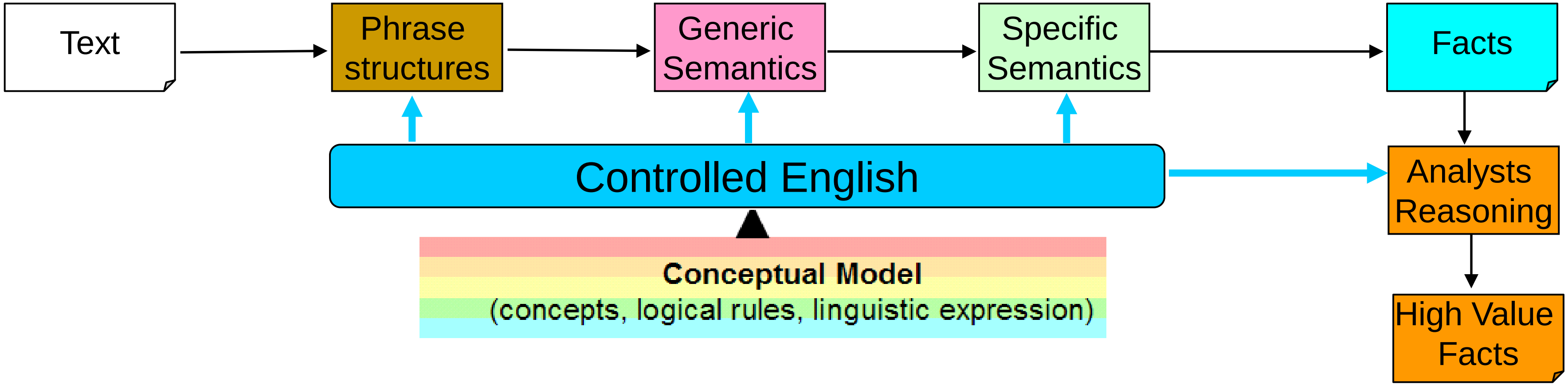
thing

stands for

"Our" Semiotic Triangle, based on the original [Ogden, C. K. and Richards, I. A. (1923).]

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Logical flow of information



- **Phrase structures, based upon upon a conceptual model of linguistics**
- **Generic semantics, based on a model of situations and agents with roles**
- **Specific semantics based, upon a domain conceptual model**

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

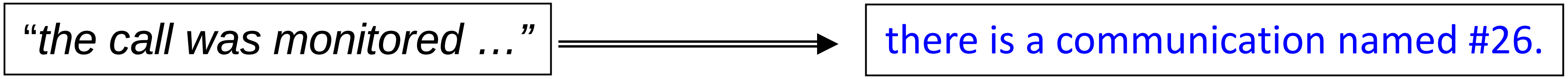
Making our “intuitions about language” accessible

“Nouns stand for things”

```
if   ( there is a noun phrase named PH )
then
    ( the noun phrase PH stands for the thing T ).
```

“Nouns tells us what type of thing”

```
if  ( the noun phrase NP has the noun N as head and stands for the thing T ) and
    ( the noun N expresses the concept C )
then
    ( the thing T is a C ).
```



“the call was monitored …”

there is a communication named #26.

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Verbs refer to “situations”

- **A situation is “something happening in the world”:**
  - an event, action, state *(from verb phrases)*
  - things *(from noun phrases)*
  - roles that these things play in the situation *(from phrase structure)*
  - location, time *(from prepositional phrases).*

Using VerbNet

- **For example:**

there is a communications monitoring situation named #39 that has
the call #15 as patient role and
has the thing #17 as source role and
has the thing #27 as destination role.

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Domain Semantics

- **e.g modelling “Communications”**
  - Reports speak about monitoring communications between people together with the things that were said

```
conceptualise
 a “communication” C that
  has the agent A as “caller” and
  has the agent B as “recipient” and
  has the value D as “date” and
  has the value T as “time” and
  has the value V1 as “caller utterance” and
  has the value V2 as “recipient utterance” and
  “is from” the place FROM and
  “is to” the place TO.
```

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Domain intuitions

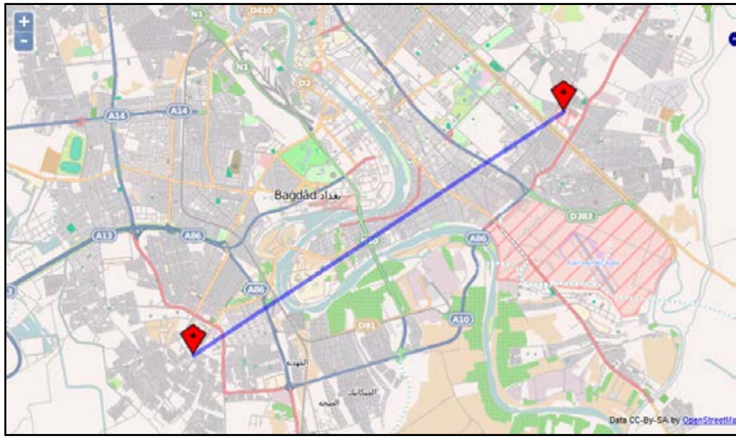
“the thing being ‘done to’ in a communications monitoring is a communication”

if (the situation S is a communications monitoring situation and has the thing T as patient role)
then
(the communications monitoring S monitors the communication T).

“the communication comes from the place where the caller is”

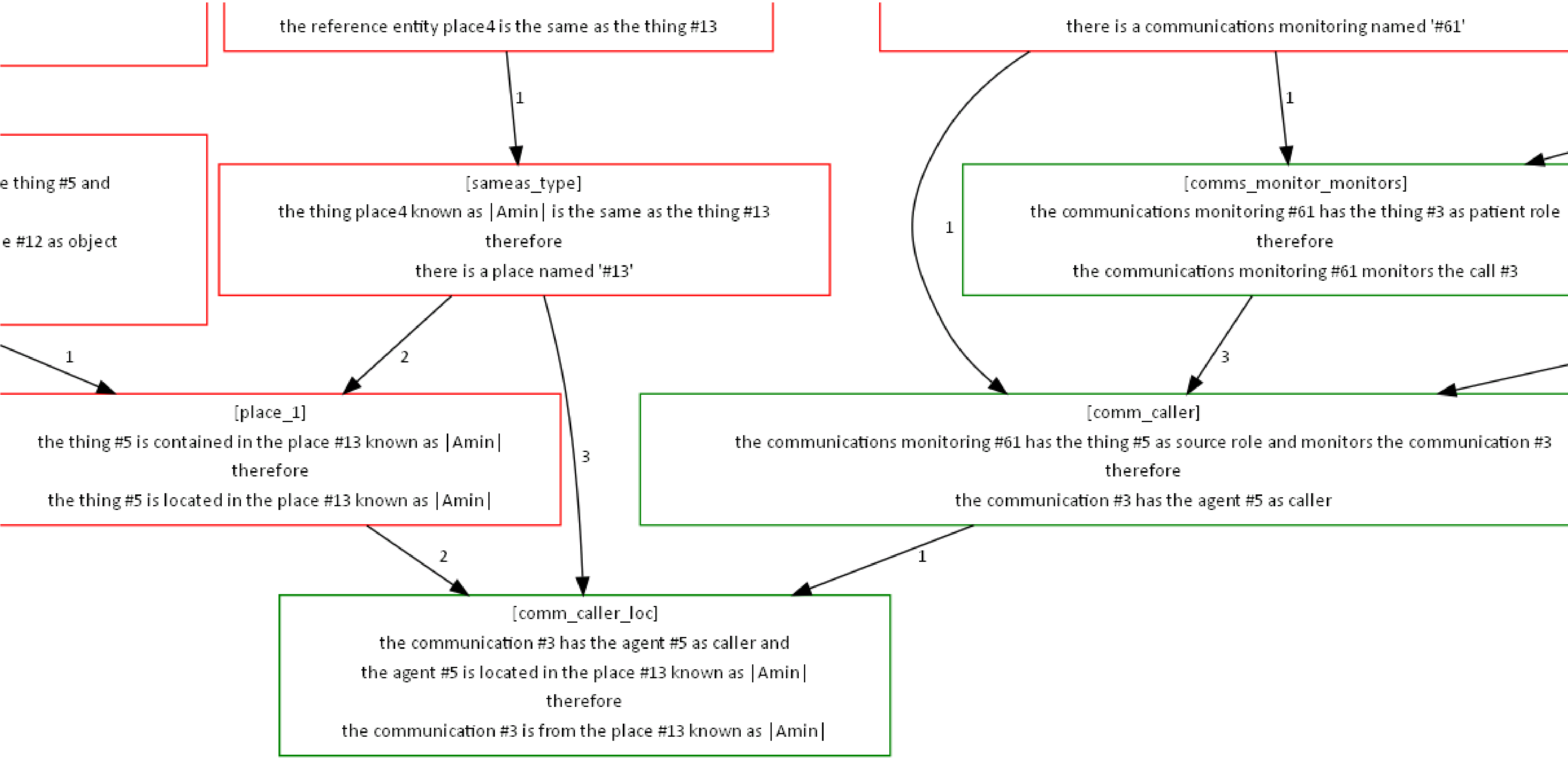
if
(the communication C has the agent A as caller) and
(the agent A is located in the place P)
then
(the communication C is from the place P).

No mention of syntax or phrase structure, you don’t need to be a linguist!



(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Rationale shows the steps leading to a fact



the reference entity place4 is the same as the thing #13

there is a communications monitoring named '#61'

e thing #5 and

e #12 as object

[sameas_type]
the thing place4 known as |Amin| is the same as the thing #13
therefore
there is a place named '#13'

[comms_monitor_monitors]
the communications monitoring #61 has the thing #3 as patient role
therefore
the communications monitoring #61 monitors the call #3

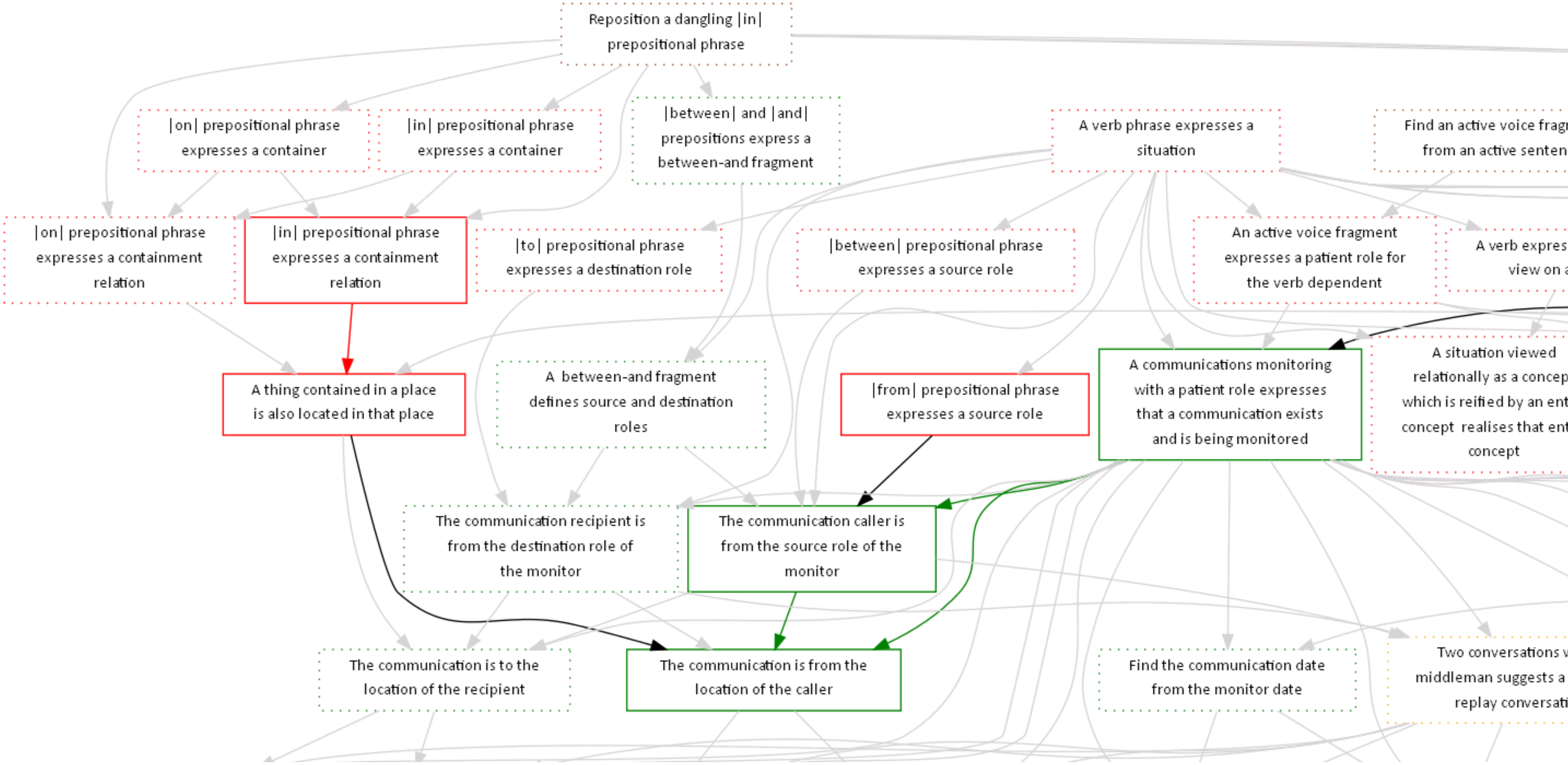
[place_1]
the thing #5 is contained in the place #13 known as |Amin|
therefore
the thing #5 is located in the place #13 known as |Amin|

[comm_caller]
the communications monitoring #61 has the thing #5 as source role and monitors the communication #3
therefore
the communication #3 has the agent #5 as caller

[comm_caller_loc]
the communication #3 has the agent #5 as caller and
the agent #5 is located in the place #13 known as |Amin|
therefore
the communication #3 is from the place #13 known as |Amin|

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Rationale for facts extracted



Reposition a dangling |in| prepositional phrase

|on| prepositional phrase expresses a container

|in| prepositional phrase expresses a container

|between| and |and| prepositions express a between-and fragment

A verb phrase expresses a situation

Find an active voice frag from an active senten

|on| prepositional phrase expresses a containment relation

|in| prepositional phrase expresses a containment relation

|to| prepositional phrase expresses a destination role

|between| prepositional phrase expresses a source role

An active voice fragment expresses a patient role for the verb dependent

A verb expres view on a

A thing contained in a place is also located in that place

A between-and fragment defines source and destination roles

|from| prepositional phrase expresses a source role

A communications monitoring with a patient role expresses that a communication exists and is being monitored

A situation viewed relationally as a concep which is reified by an ent concept realises that ent concept

The communication recipient is from the destination role of the monitor

The communication caller is from the source role of the monitor

The communication is to the location of the recipient

The communication is from the location of the caller

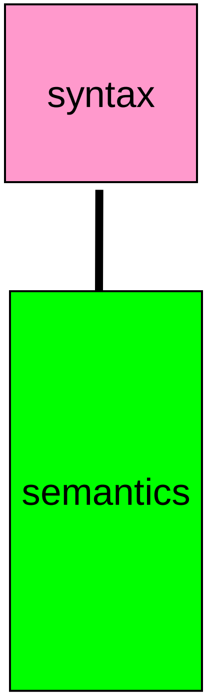
Find the communication date from the monitor date

Two conversations w middleman suggests a replay conversati

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

“Linguistic Frames” for capturing syntax and semantics

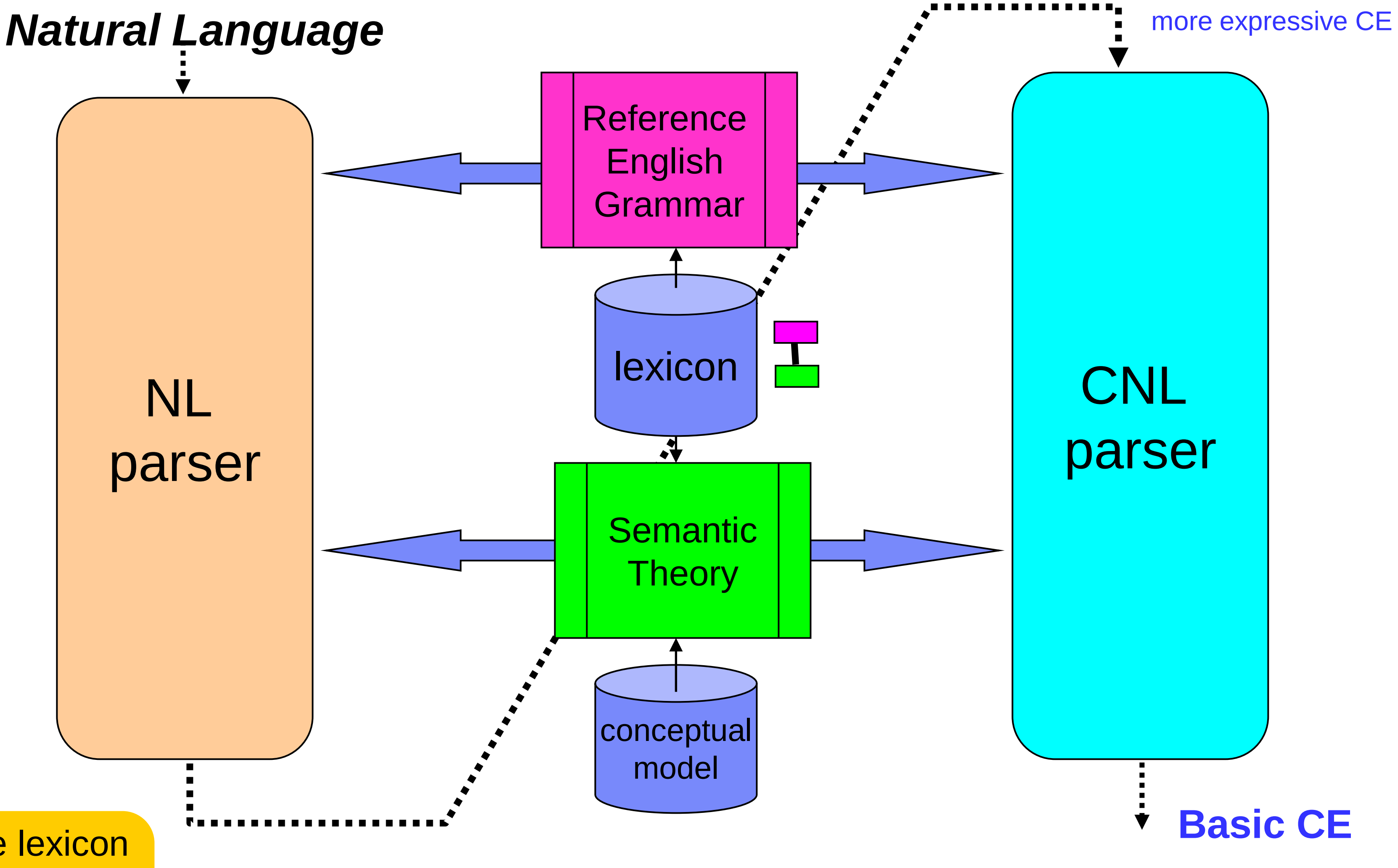
```
there is a linguistic frame named np3 that
   has 'a person' as example and

   defines the noun phrase NP and

   has the sequence
     ( the determiner DET and the noun COMMON )
   as syntactic pattern and

   has the statement that
     ( the noun COMMON expresses the entity concept EC )
   as preconditions and

   has the statement that
     ( the noun phrase NP stands for the thing X ) and
     ( the thing X is an EC )
  as semantic statement.
```



We have used this for extending the syntax of CE

2013. All Rights Reserved.

Converging NL and CNL parsers



Is the lexicon just a set of linguistic frames?

← *Better understanding of linguistics*

Increase stylistic expressibility of CE →

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

2013-2015: Objectives

- **Integration to DELPH-IN linguistic resources to provide better fact extraction**
  - **Linking to our CE-based architecture**
  - **“Deeper semantics”: integrate general semantics with domain semantics**
  - **Expressing grammatical knowledge in CE**
- Extension of Controlled English for greater expressiveness
  - learning to build a CNL from understanding NL
- Improved reasoning capabilities
  - constraints
  - assumptions and hypotheses

Now working with Prof Ann Copestake

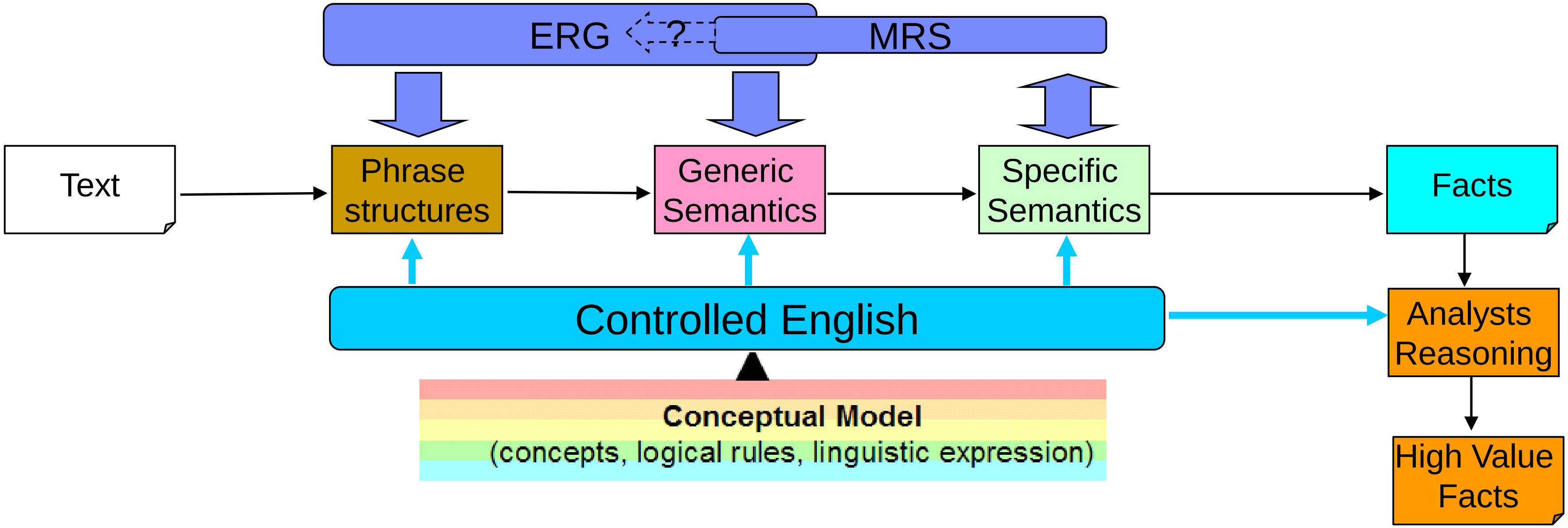
Chosen to use the ERG

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Linking to the CE-based architecture

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

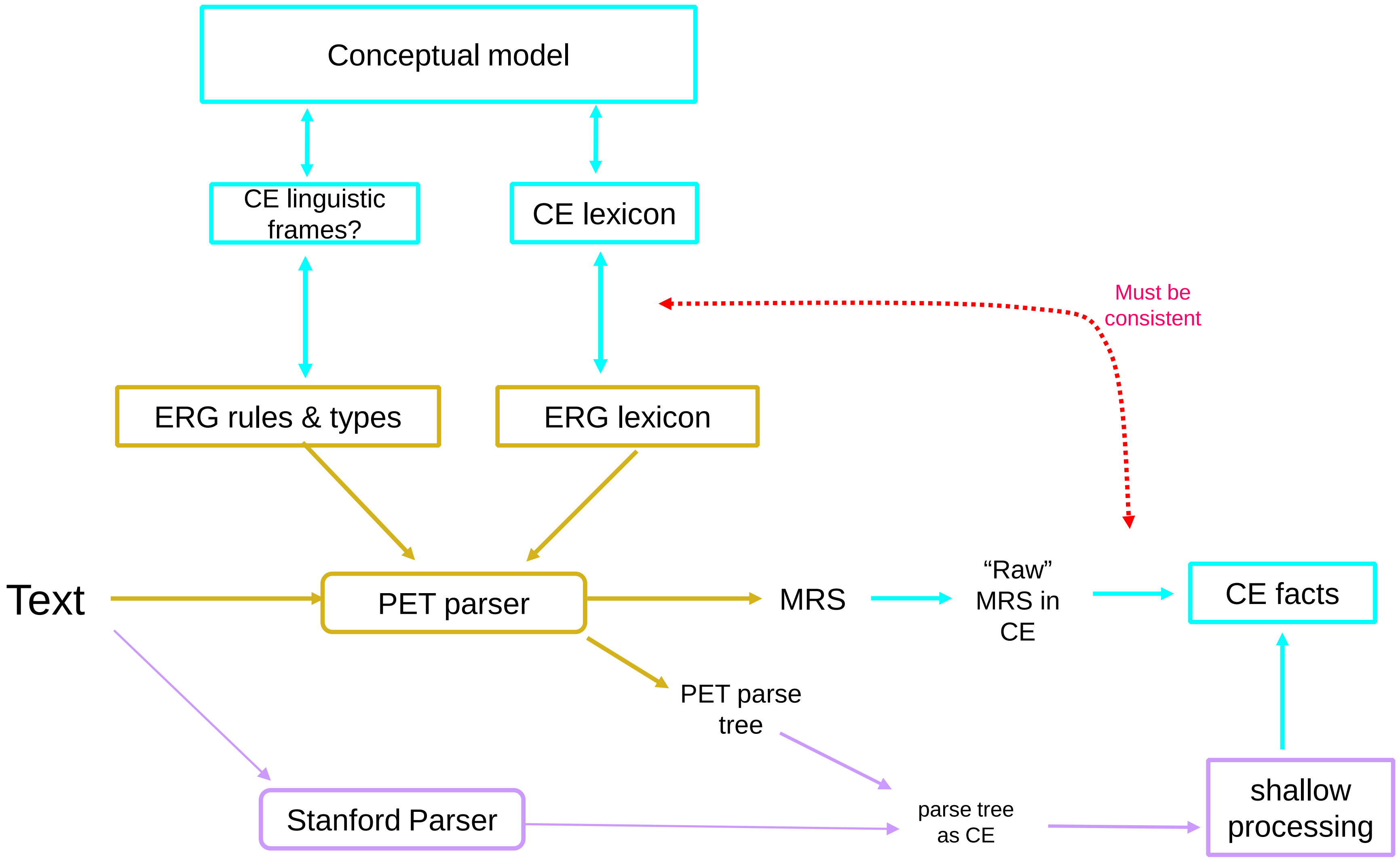
Linking to the CE-based Functional Architecture



- Use the ERG to parse sentences and provide the phrase structure
- Use MRS to express generic semantics
- Integrate the the domain semantics in the conceptual model, MRS and generic semantics
- Feedback domain semantics (via MRS) to affect the parser?

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Integration of ERG and CE ?



(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Tasks for integrating Controlled English and the ERG

- adding domain specific words to the lexicon
- generating the parse tree
  - for applications that work off a parse tree
- representing grammar rules
  - for updating domain specific rules
  - for understanding the linguistic reasoning
- integrating MRS and domain semantics
  - for output of CE facts
  - guiding the parsing rules by domain semantics?
- determining rationale for a specific linguistic conclusion
  - to review the reasons for an important conclusion

All of this must be “understandable” and in CE

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Integrating to the lexicon

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Lexicon in CE

```
person_n1 := n_-_mc_le &
 [ ORTH < "person" >,
   SYNSEM [ LKEYS.KEYREL.PRED "_person_n_1_rel",
            PHON.ONSET con ] ].
```

The user has to define this link

there is a singular noun named |person_NN| that
 is written as the word ‘person’ and
 is a form of the noun sense ‘person_n1’.

there is a mass or count noun sense named person_n1 that
 expresses the entity concept ‘person’.

- **Translation principles:**
  - An entry is equivalent to a CE word sense
  - Word senses are subclassed into a hierarchy of CE generic subtypes with specific ENTRYs at the leaf nodes.
  - The orthography is represented a word (simple or compound)
  - The word sense expresses an entity concept in the conceptual model, defined by the user

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Grammar rules

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Phrase “the man (20)”

```
(137 np_frg_c 0 0 3 [root_inffrag]
   (122 hdn-np_app-r-pr_c 0 0 3
     (50 sp-hd_n_c 0 0 2
       (16 the_1/d_-_the_le -0.8038 0 1 []
         (1 "the" 0 0 1 <0:1>))
       (34 n_ms-cnt_ilr 0 1 2
         (23 man1/n_-_mc_le 0.1576 1 2 []
           (2 "man" 0 1 2 <1:2>))))
     (90 hdn_bnp-num_c 0 2 3
       (86 hdn_np-num_c 0 2 3
         (80 w_lparen_plr 0 2 3
           (77 w_rparen_plr 0 2 3 [w_lparen_plr]
             (32 twenty_num/aj_-_i-crd-two_le 0 2 3
                      [w_rparen_plr w_lparen_plr]
               (3 "(20)" 0 2 3 <2:3>))))))))
```

the head phrase #p_137 has the nominal head nominal phrase phrase #p_122 as head.

the nominal head nominal phrase phrase #p_122 has the determiner phrase #p_50 as head
and has the adjective phrase #p_90 as dependent.

the determiner phrase #p_50 has the determiner |the_DT| as head
and has the noun phrase #p_34 as dependent.

the noun phrase #p_34 has the noun |man_NNS| as head.

the adjective phrase #p_90 has the adjective phrase #p_86 as head.

the adjective phrase #p_86 has the adjective phrase #p_80 as head.

the adjective phrase #p_80 has the adjective phrase #p_77 as head.

the adjective phrase #p_77 has the adjective ‘|(20)_JJ|’ as head.

the noun |man_NNS| is a plural noun and has 'n_-_mc_le' as erg type.

the noun phrase #p_34 is a head phrase and has 'n_ms-cnt_ilr' as erg type.

the determiner phrase #p_50 is a specifier head phrase and has 'sp-hd_n_c' as erg type.

the adjective phrase #p_77 is a head phrase and has 'w_rparen_plr' as erg type
and has the thing w_lparen_plr as feature.

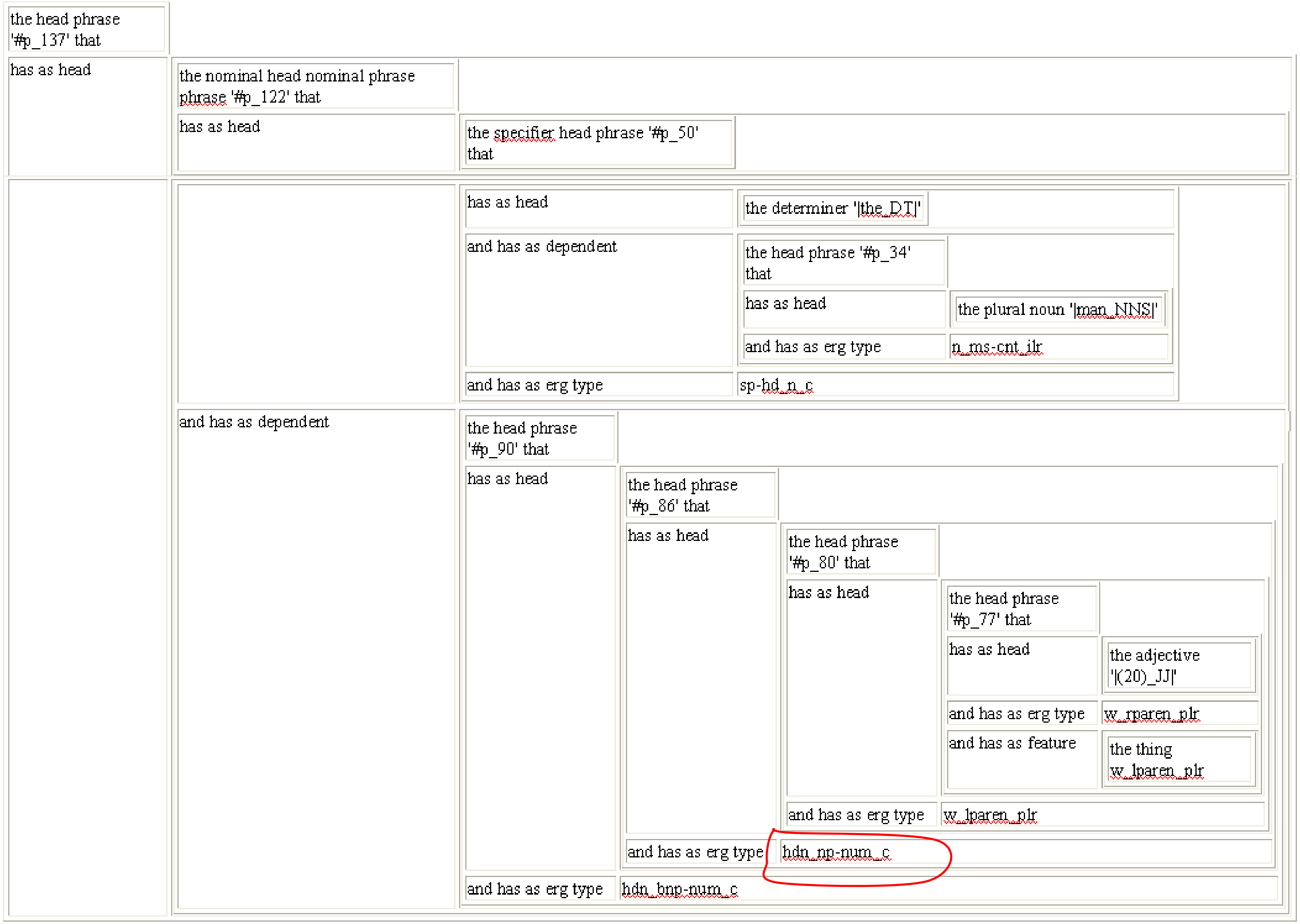
the adjective phrase #p_80 is a head phrase and has 'w_lparen_plr' as erg type.

the adjective phrase #p_86 is a head phrase and has 'hdn_np-num_c' as erg type.

the adjective phrase #p_90 is a head phrase and has 'hdn_bnp-num_c' as erg type.

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

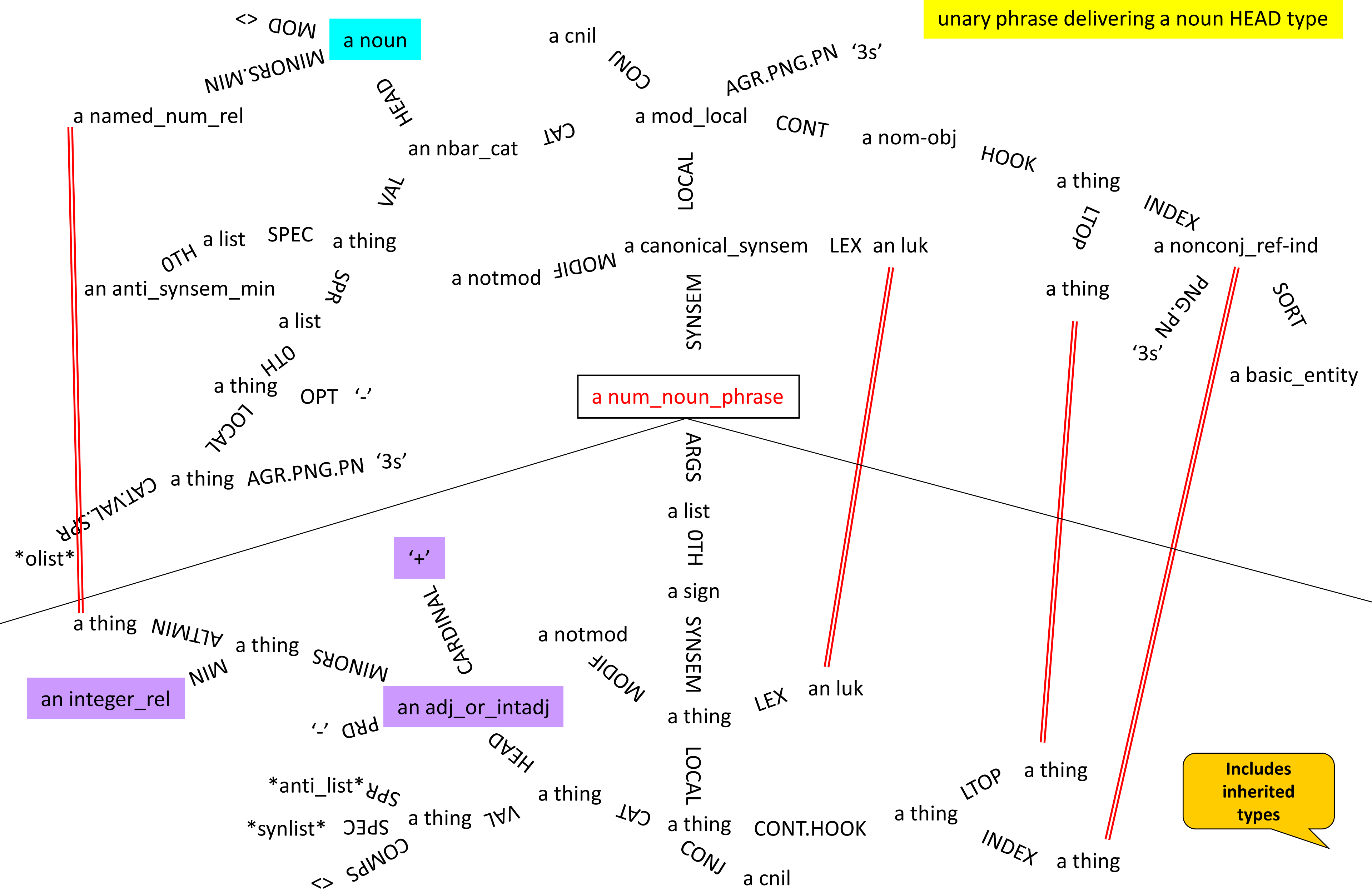
Parse tree in CE tabular form



- the head phrase '#p_137' that
- has as head
  - the nominal head nominal phrase phrase '#p_122' that
  - has as head
    - the specifier head phrase '#p_50' that
    - has as head
      - the determiner '|the_DT|'
    - and has as dependent
      - the head phrase '#p_34' that
      - has as head
        - the plural noun '|man_NNS|'
      - and has as erg type
        - n_ms-cnt_ilr
    - and has as erg type
      - sp-hd_n_c
  - and has as dependent
    - the head phrase '#p_90' that
    - has as head
      - the head phrase '#p_86' that
      - has as head
        - the head phrase '#p_80' that
        - has as head
          - the head phrase '#p_77' that
          - has as head
            - the adjective '|(20)_JJ|'
          - and has as erg type
            - w_rparen_plr
          - and has as feature
            - the thing w_lparen_plr
        - and has as erg type
          - w_lparen_plr
      - and has as erg type
        - hdn_np-num_c
    - and has as erg type
      - hdn_bnp-num_c

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

unary phrase delivering a noun HEAD type



a num_noun_phrase

SYNSEM

a canonical_synsem

LOCAL

a mod_local

CAT

an nbar_cat

HEAD

a noun

MOD

<>

MINORS.MIN

a named_num_rel

VAL

a thing

SPEC

a list

0TH

an anti_synsem_min

SPR

a list

0TH

a thing

OPT

'-'

LOCAL

a thing

AGR.PNG.PN

'3s'

CAT.VAL.SPR

olist

MODIF

a notmod

CONJ

a cnil

AGR.PNG.PN

'3s'

CONT

a nom-obj

HOOK

a thing

LTOP

a thing

INDEX

a nonconj_ref-ind

PNG.PN

'3s'

SORT

a basic_entity

LEX

an luk

ARGS

a list

0TH

a sign

SYNSEM

a thing

LEX

an luk

MODIF

a notmod

LOCAL

a thing

CAT

a thing

HEAD

an adj_or_intadj

CARDINAL

'+'

MINORS

a thing

ALTMIN

a thing

MIN

an integer_rel

PRD

'-'

VAL

a thing

SPR

anti_list

SPEC

synlist

COMPS

<>

CONJ

a cnil

CONT.HOOK

a thing

LTOP

a thing

INDEX

a thing

Includes inherited types

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Intuition about ERG phrase rules

“Its all about combining substructures into superstructures”

Can we use linguistic frames?

there is a linguistic frame named f1 that
defines the PHRASETYPE PH and

has the sequence
(the sign SUB1 , the sign SUB2 …)
as subcomponents and

has the statement that
(the sign SUB1 has …) and
(the sign SUB2 has …)
as precondition and

has the statement that
(the phrase PH has …)
as semantics.

unification

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Possible linguistic frame for num_noun_phrase

```
there is a linguistic frame named f1 that
  defines the num_noun_phrase PH and

  has the sequence ( the sign SUB1 ) as subcomponents and

  has the statement that
    ( the thing C1 is headed by the adj_or_intadj ADJ ) and
    ( the adj_or_intadj ADJ is a CARDINAL and is not a PRD and
        has an integer_rel as the MIN of the MINORS ) and
    ( the thing C1 has the list <> as the complements and
        has an *anti_list* as the specifier )
  as precondition and

  has the statement that
    ( the phrase PH has the nbar_cat C as syntactic category and
        has '3s' as the PN of the PNG of the AGR of the LOCAL of the SYNSEM ) and
    ( the nbar_cat C is headed by the noun N and
        has '3s' as the PN of the PNG of the AGR of the LOCAL of the 0th element of the specifier ) and
    ( the noun N has the ALTMIN of the MINORS of the adj_or_intadj ADJ as the MIN of the MINORS ) and
  as semantics.
```

I can begin to see …

It’s a unary phrase, turning a sign into a num_noun_phrase

only applies when the sign is a CARDINAL and not a PRD and has an integer relation

an adj_or_intadj HEAD is turned into a noun HEAD

some form of third singular agreement is being constructed

a relation is passed from the sign to the phrase via MIN and ALTMIN ??

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Integrating MRS and domain semantics

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Three stage approach to integrating MRS and CE

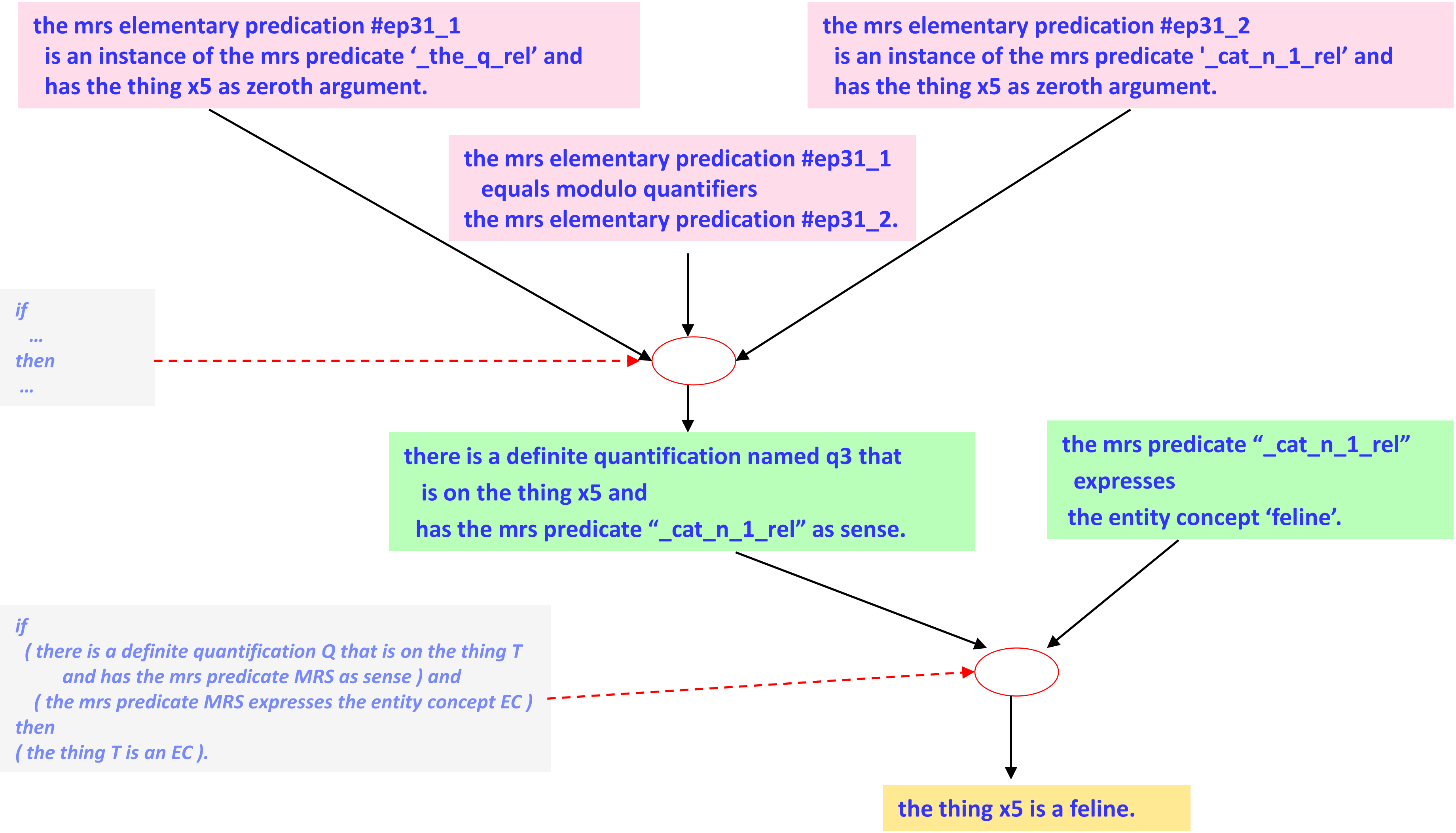
1. **Generate a raw representation of :**
   1. the elementary predications (EPs) as objects with predicate and arguments
   2. the scope information between EPs

2. **Extract intermediate, but generic, concepts describing the raw MRS:**
   1. patterns of quantification

3. **Turn the raw and intermediate representation into domain specific CE facts:**
   1. using the links between the predicate and the CE concept.
   2. taking account of selectional restrictions?
   3. …

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Three level example “the cat”



the mrs elementary predication #ep31_1
is an instance of the mrs predicate ‘_the_q_rel’ and
has the thing x5 as zeroth argument.

the mrs elementary predication #ep31_2
is an instance of the mrs predicate '_cat_n_1_rel’ and
has the thing x5 as zeroth argument.

the mrs elementary predication #ep31_1
equals modulo quantifiers
the mrs elementary predication #ep31_2.

if
…
then
…

there is a definite quantification named q3 that
is on the thing x5 and
has the mrs predicate “_cat_n_1_rel” as sense.

the mrs predicate “_cat_n_1_rel”
expresses
the entity concept ‘feline’.

if
(there is a definite quantification Q that is on the thing T
and has the mrs predicate MRS as sense) and
(the mrs predicate MRS expresses the entity concept EC)
then
(the thing T is an EC).

the thing x5 is a feline.

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Other simple rules for turning MRS into domain concepts

adjectives

```
if
 ( the mrs elementary predication P
    is an instance of the mrs predicate MRS and
    has the situation S as zeroth argument and
       has the thing T as first argument ) and
   ( the mrs predicate MRS expresses the entity concept EC )
then
 ( the  thing T is an EC ).
```

proper names

```
if
( the mrs elementary predication P
  is an instance of the mrs predicate ‘named_rel’ and
  has the thing T as zeroth argument and
  has the value C as c argument )
then
 ( the thing T has the value C as common name ).
```

```
if
 ( the mrs elementary predication EP
   is an instance of the mrs predicate '_in_p_rel'
   and has the thing T as first argument
   and has the thing C as second argument )
then
 ( the thing T is contained in the container C ).
```

prepositions

```
if
 ( the mrs elementary predication P
    is an instance of the mrs predicate MRS and
    has the situation S as zeroth argument and
    has the thing T1 as first argument and
    has the thing T2 as second argument ) and
 ( the mrs predicate MRS expresses the relation concept RC )
then
 ( the thing T1 RC the thing T2).
```

verbs

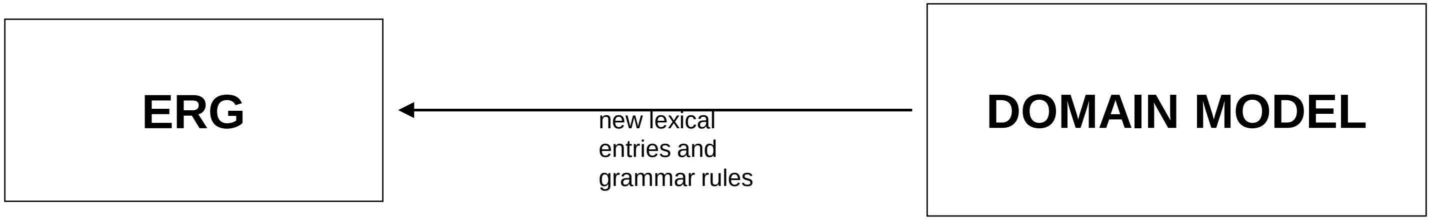
Should we infer roles?

Need to understand the theory of semantics

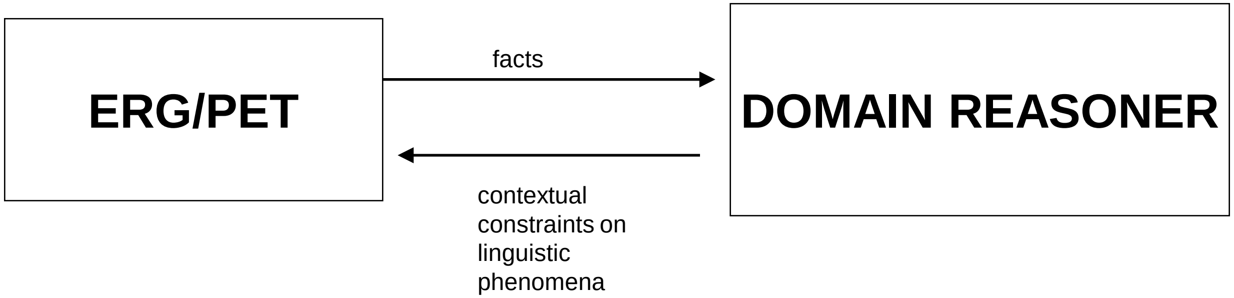
(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Feedback of domain reasoning to the parsing?

- **We want the domain to affect the parse, eg:**
  - creating new lexical entries and grammar rules prior to parsing



- **But we also want arbitrary domain reasoning to affect the parse at runtime**



- **Could this:**
  - rule out inconsistent parses
  - provide disambiguations, and dialog context?

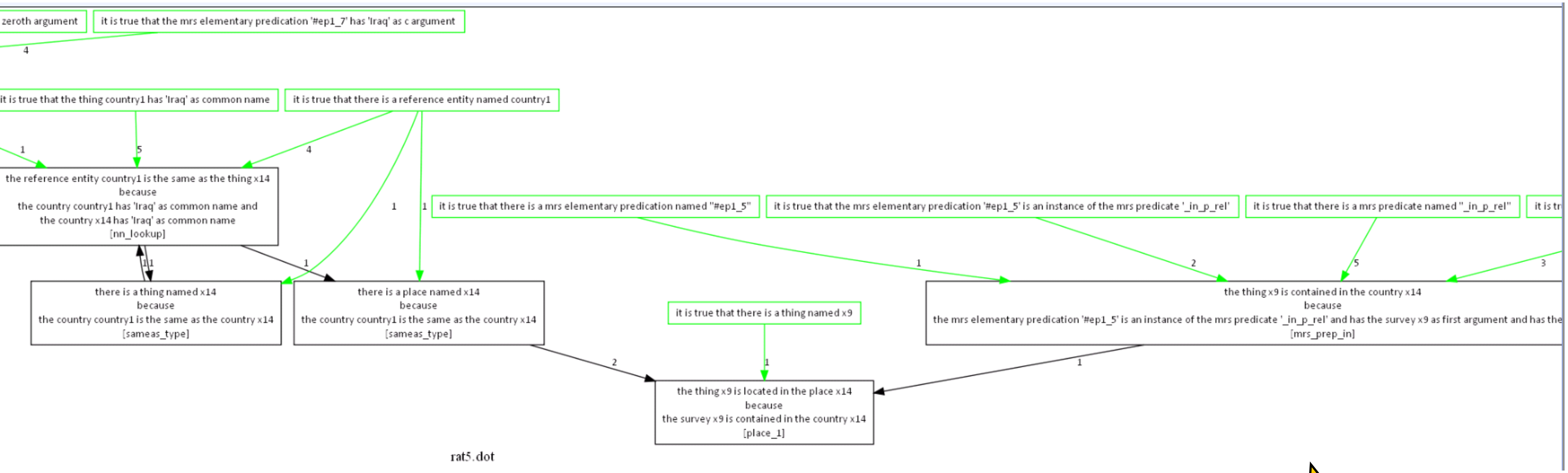
(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Rationale

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Rationale for the semantic reasoning

the survey x9 is located in the country x14 known as Iraq.



zeroth argument

it is true that the mrs elementary predication '#ep1_7' has 'Iraq' as c argument

it is true that the thing country1 has 'Iraq' as common name

it is true that there is a reference entity named country1

the reference entity country1 is the same as the thing x14
because
the country country1 has 'Iraq' as common name and
the country x14 has 'Iraq' as common name
[nn_lookup]

it is true that there is a mrs elementary predication named "#ep1_5"

it is true that the mrs elementary predication '#ep1_5' is an instance of the mrs predicate '_in_p_rel'

it is true that there is a mrs predicate named "_in_p_rel"

it is tr

there is a thing named x14
because
the country country1 is the same as the country x14
[sameas_type]

there is a place named x14
because
the country country1 is the same as the country x14
[sameas_type]

it is true that there is a thing named x9

the thing x9 is contained in the country x14
because
the mrs elementary predication '#ep1_5' is an instance of the mrs predicate '_in_p_rel' and has the survey x9 as first argument and has the
[mrs_prep_in]

the thing x9 is located in the place x14
because
the survey x9 is contained in the country x14
[place_1]

rat5.dot

Original Sentence: htt conduct the survey in iraq

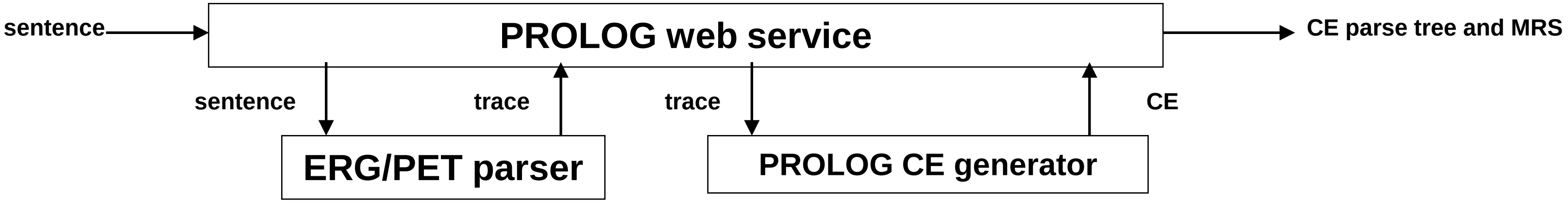
Can we get the rationale for the EPS?

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Physical Integration

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Running ERG under PET



- **PET is run under Linux (DEBIAN) in an ORACLE VirtualBox image**
  - run with the flags:
    - –verbose=3 –mrs –nsolutions english
  - output is in the form of a text trace, which must be “scraped” to obtain the data

Is there a better way?

- **A Prolog program provides a web service for parsing sentences and turning the result into CE**
  - On initialisation, the program:
    - starts PET as a process, with an input and output pipe
  - On request for a sentence to be parsed, the program:
    - puts the sentence into the input pipe and grabs the output from the output pipe
    - parses and analyses the output into CE
    - returns the CE as the result of the web service call

Avoids the long startup

- **Aiming to integrate to our CE Store**

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Objectives of research

- To better understand the complexities of natural language, to link to the external NL research community and to increase our capabilities
- To offer common models of language processing to cover a range of techniques
- To extend the research in semantics as applied to linguistic processing and to allow guidance of language parsing via domain models
- To provide better tools for allowing users to configure NL processing and to integrate fact extraction and reasoning to generate high-value information

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

Some questions

- Is this of any interest?
- How do I get the PET system better integrated as a service?
  - or should I use ACE?
- How do we link between phrases and entities in the MRS?
- How do we get the rationale?
- How do we feedback domain semantics to the parsing?

Thank you

mottdh@googlemail.com

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.

References

- Mott, D., Giammanco, C., Dorneich, M.C., Patel, J., & Braines, D. (2010). “Hybrid Rationale and Controlled Natural Language for Shared Understanding”. *Proceedings of the 6th Knowledge Systems for Coalition Operations*, Vancouver, Canada.
- Mott, D. (2010). Summary of Controlled English, ITACS, https://www.usukita.org/papers/5658/details.html.
- Mott, D., Braines, D., Poteet, S., Kao, A., and Xue, P. (2012). Controlled Natural Language to Facilitate Information Extraction. In Proceedings of the Sixth Annual Conference of the International Technology Alliance, London, UK.
- Xue, P., Poteet, S., Kao, A., Mott, D., Braines, D., (2013) Constructing Controlled English for Both Human Usage and Machine Processing, RuleML 2013

(C) Copyright IBM Corp. 2011,2013. All Rights Reserved.