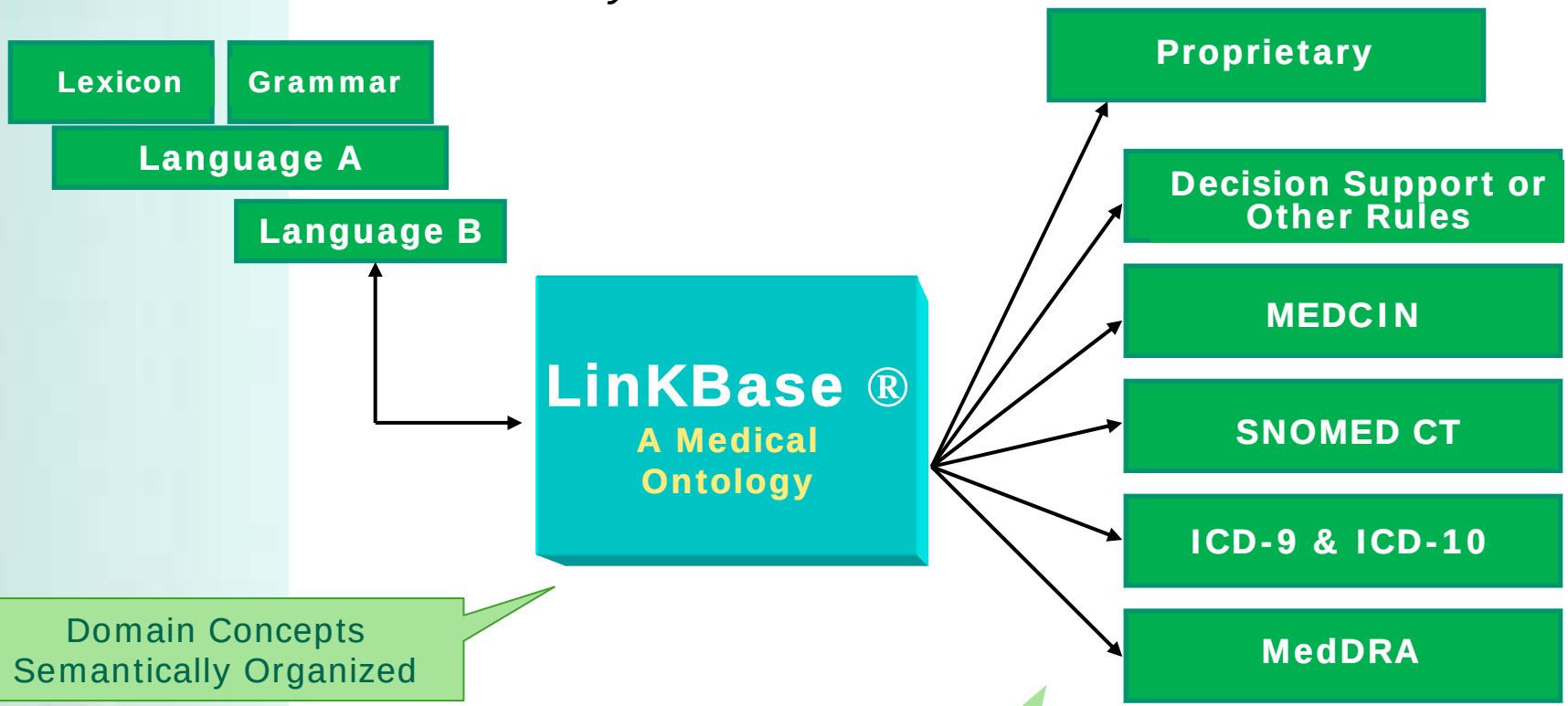
LinKBase® and SNOMED some distinct features and impact on NLP

Maria van Gurp PhD
Marnix Holvoet
Mariana Casella dos Santos MD

KR-MED 2008

The LinKBase® Ontology

Main purpose: serving NLP/NLU
About 100 man years of development
Commercially available



Domain Concepts Semantically Organized

Cross-referenced to 3rd party Content Systems

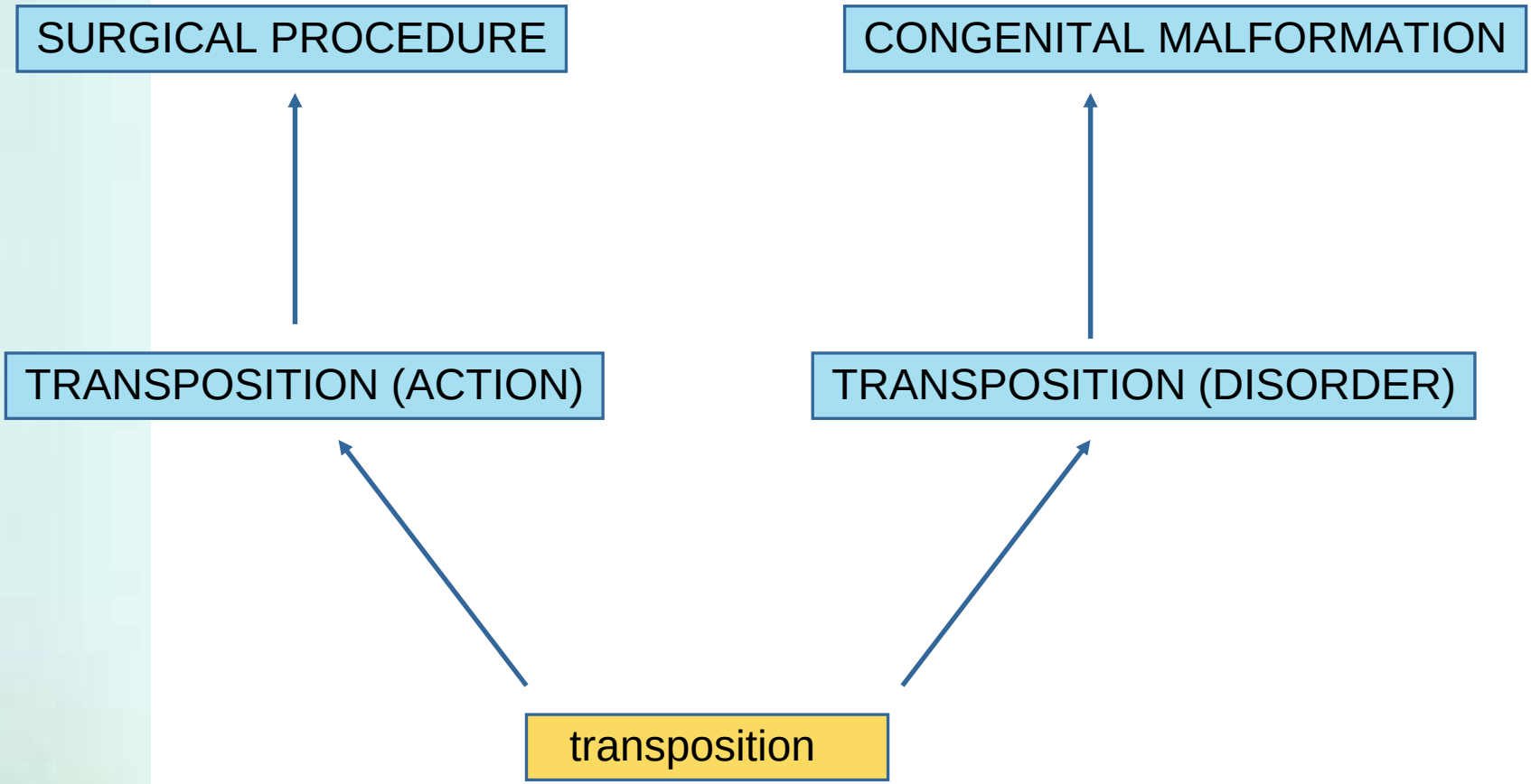
What impacts NLP?

1. Top structure: presence of upper- and mid layer
2. Relationships and full definitions
3. Synonym assignments and the lexicon

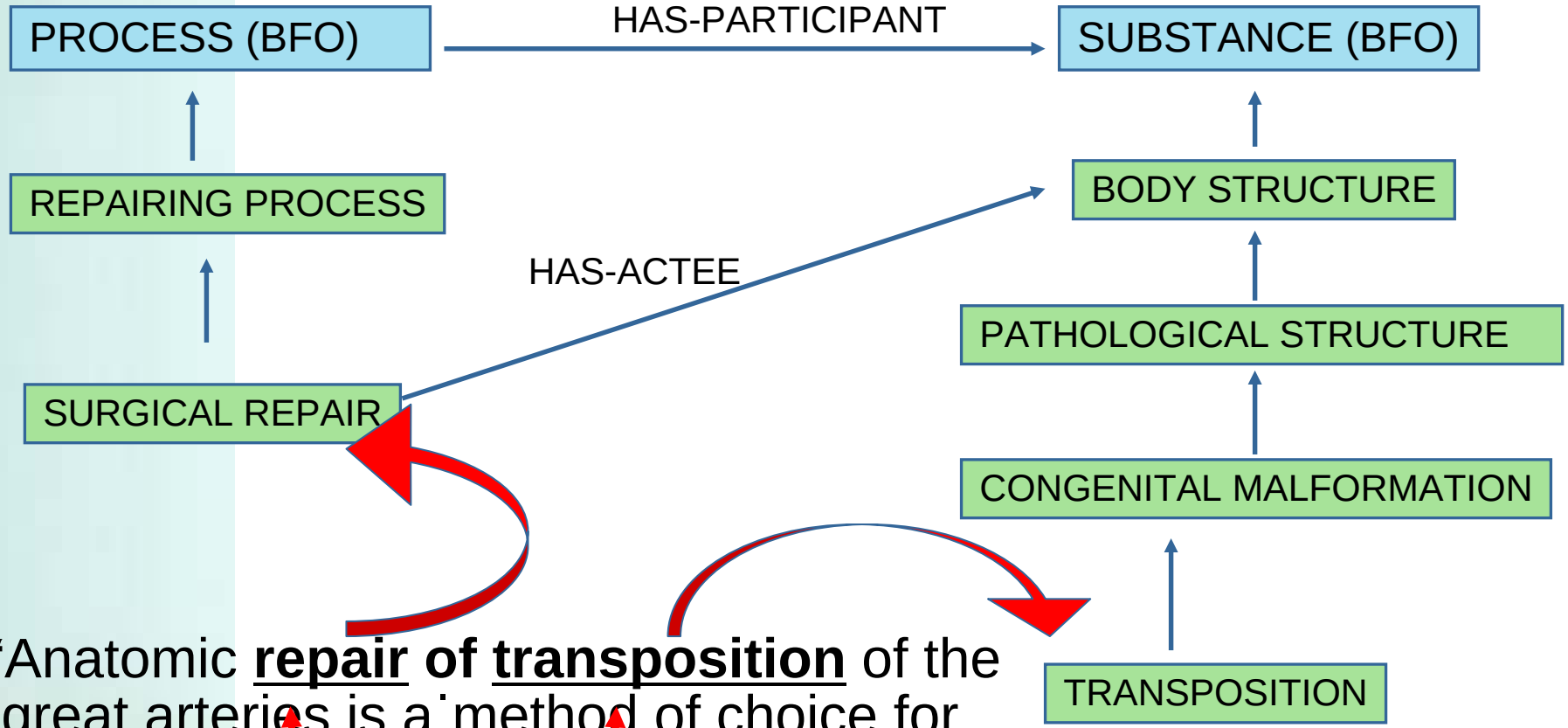
What impacts NLP?

1. Top structure: presence of upper- and mid layer
2. Relationships and full definitions
3. Synonym assignments and the lexicon

Top structure and NLP



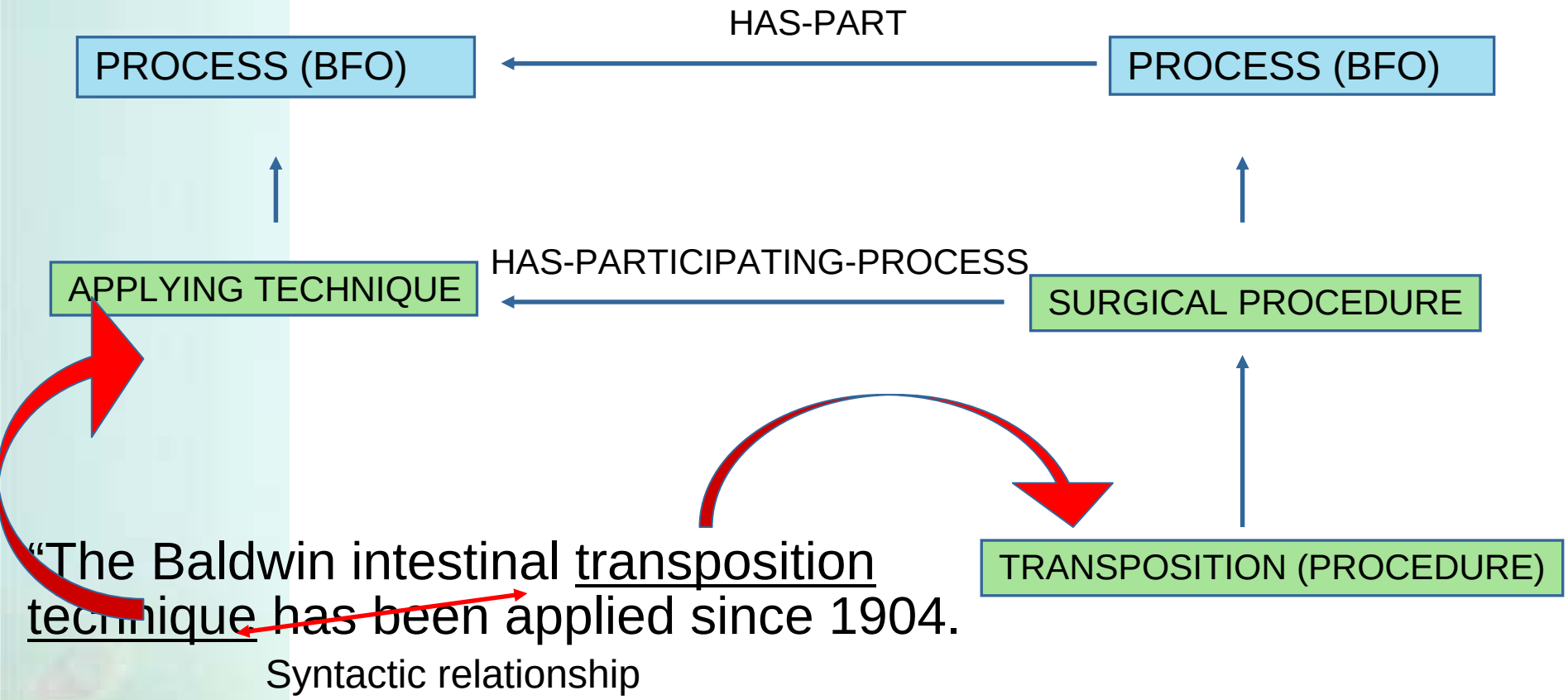
Top structure and NLP



"Anatomic **repair of transposition** of the great arteries is a method of choice for treatment of this congenital heart defect.

Syntactic relationship

Top structure and NLP

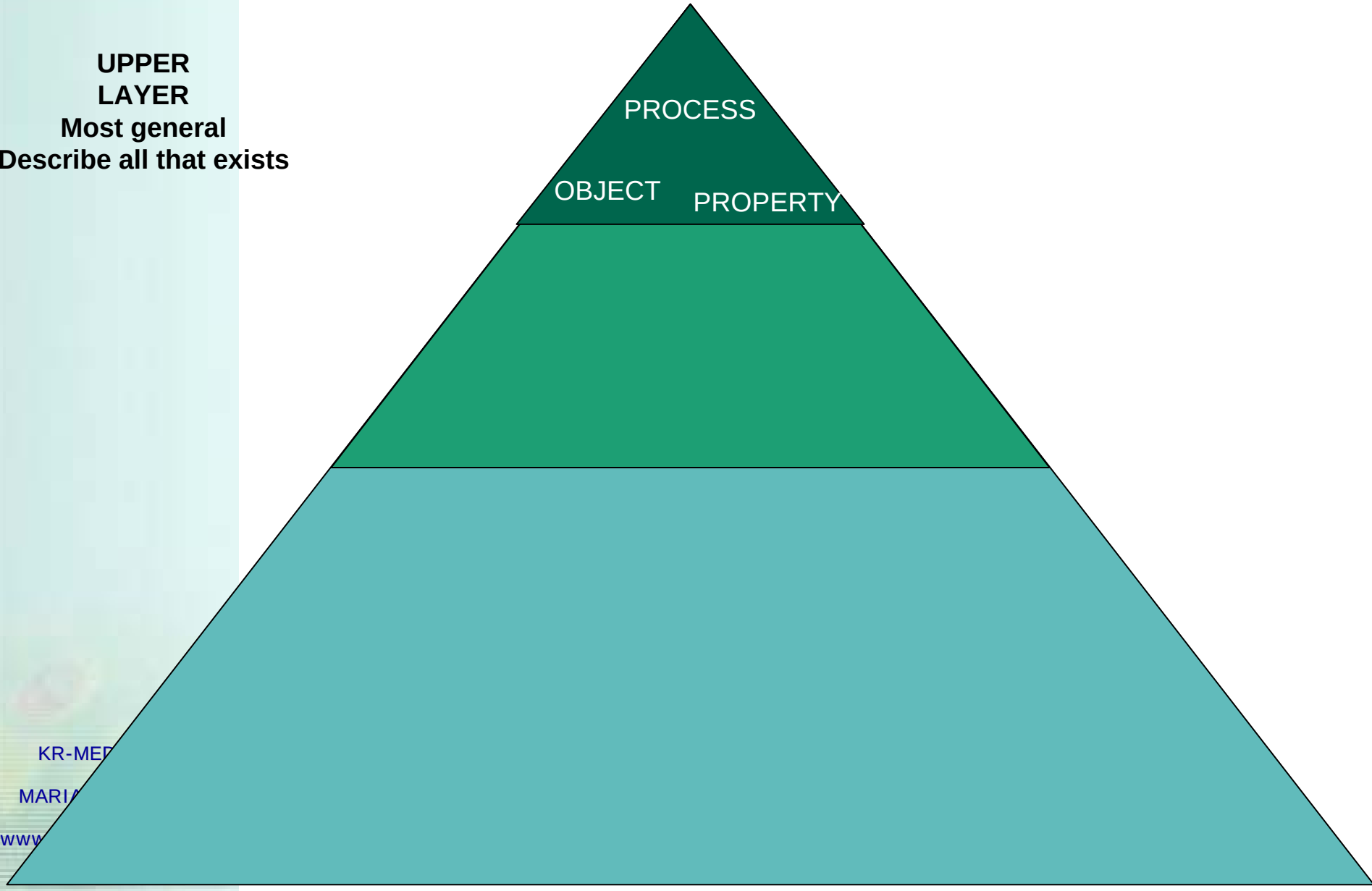


"The Baldwin intestinal transposition technique has been applied since 1904.

Syntactic relationship

LinKBase® Layers

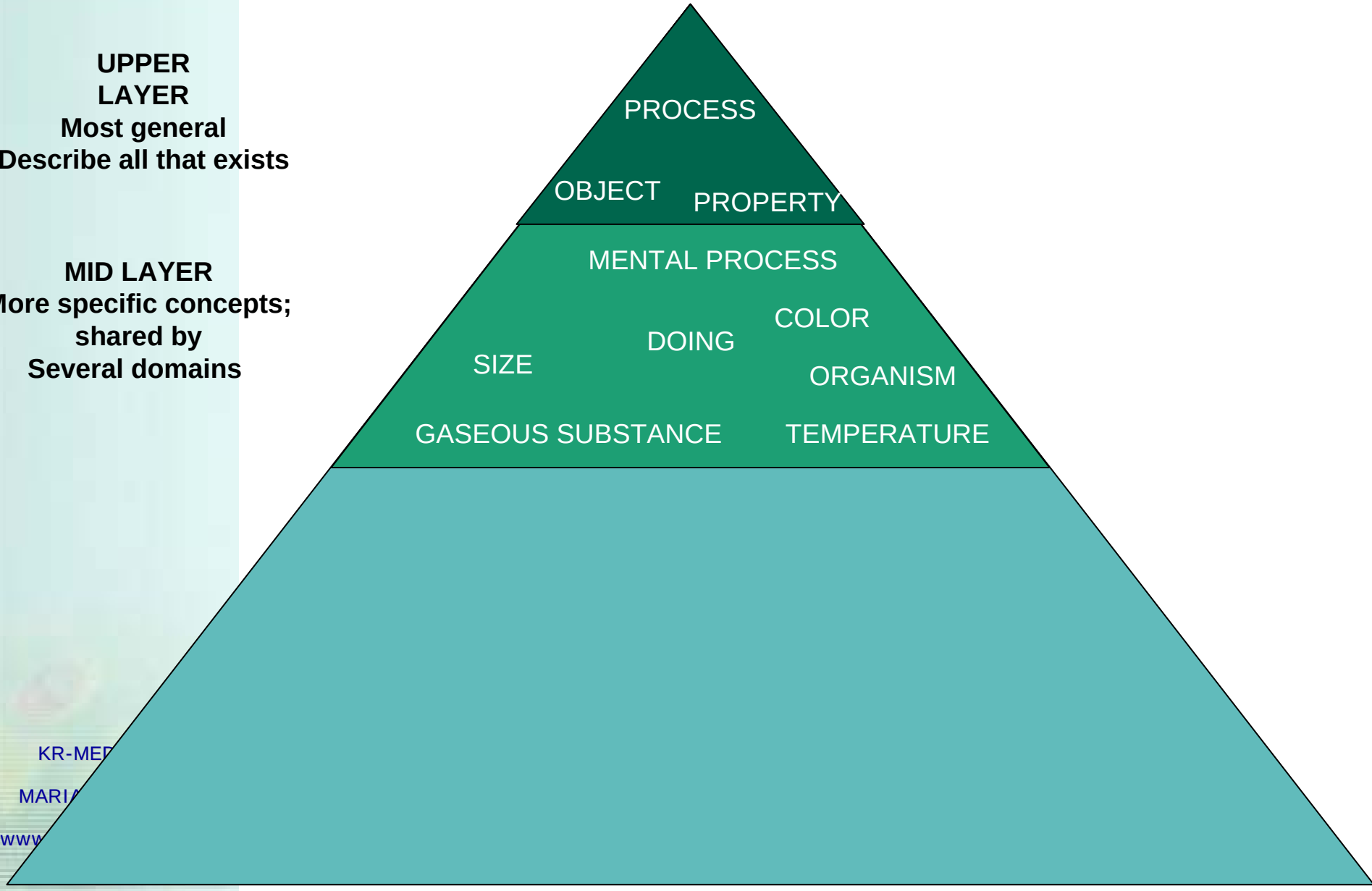
UPPER LAYER
Most general
Describe all that exists



LinKBase® Layers

UPPER LAYER
Most general
Describe all that exists

MID LAYER
More specific concepts; shared by Several domains

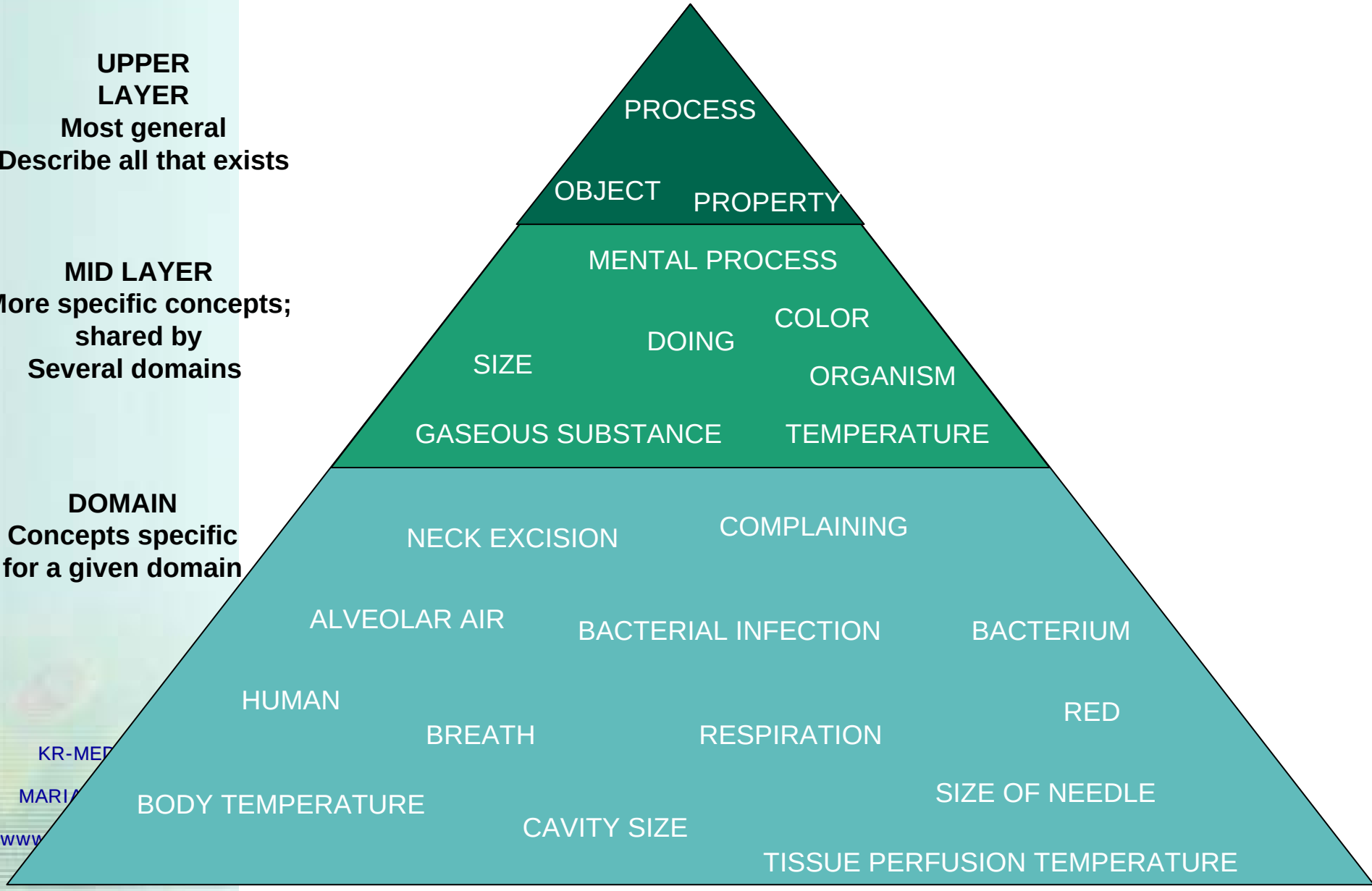


LinKBase® Layers

UPPER LAYER
Most general
Describe all that exists

MID LAYER
More specific concepts; shared by Several domains

DOMAIN
Concepts specific for a given domain



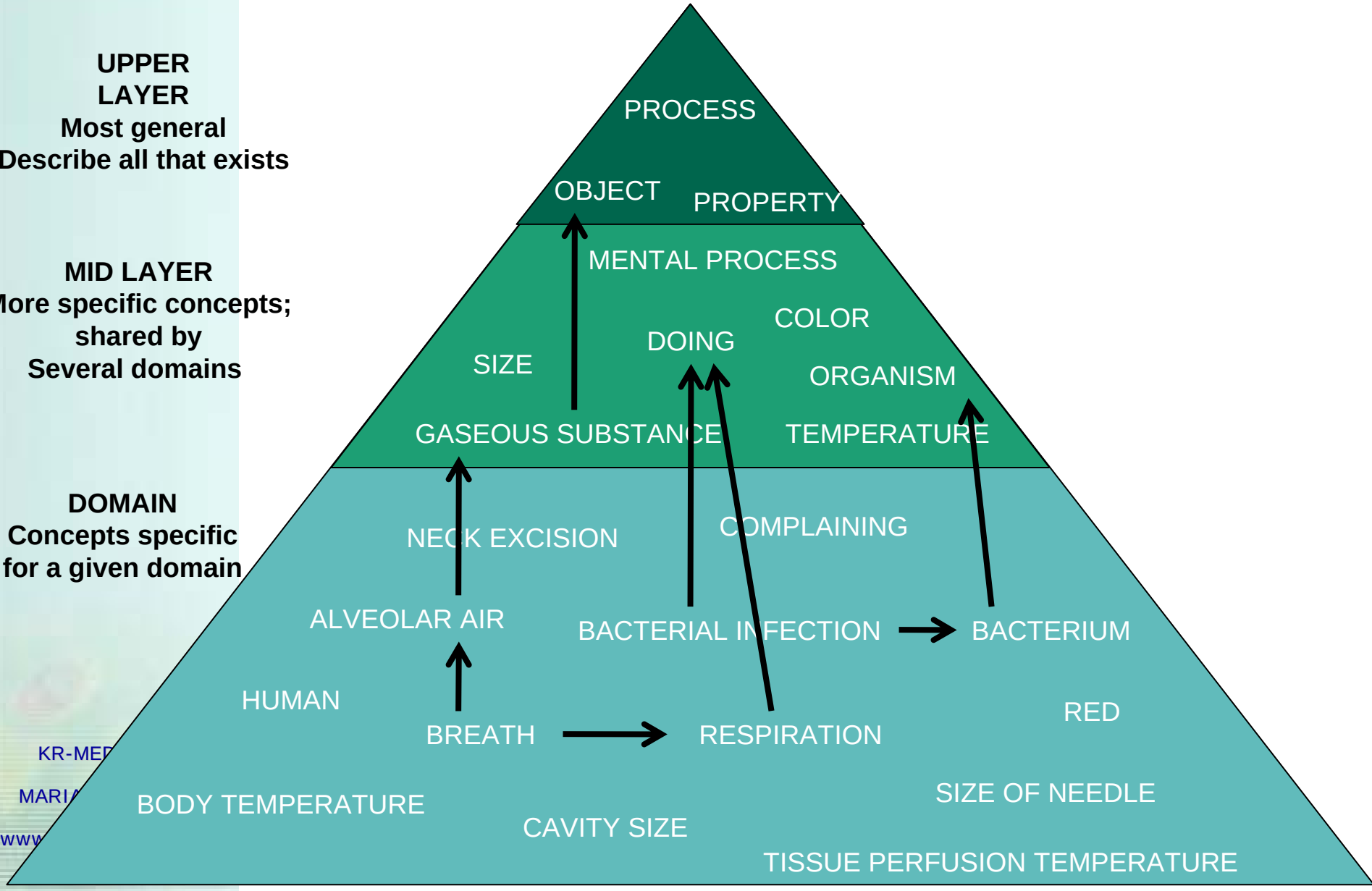
KR-MED
MARIA
www

LinKBase® Layers

UPPER LAYER
Most general
Describe all that exists

MID LAYER
More specific concepts; shared by Several domains

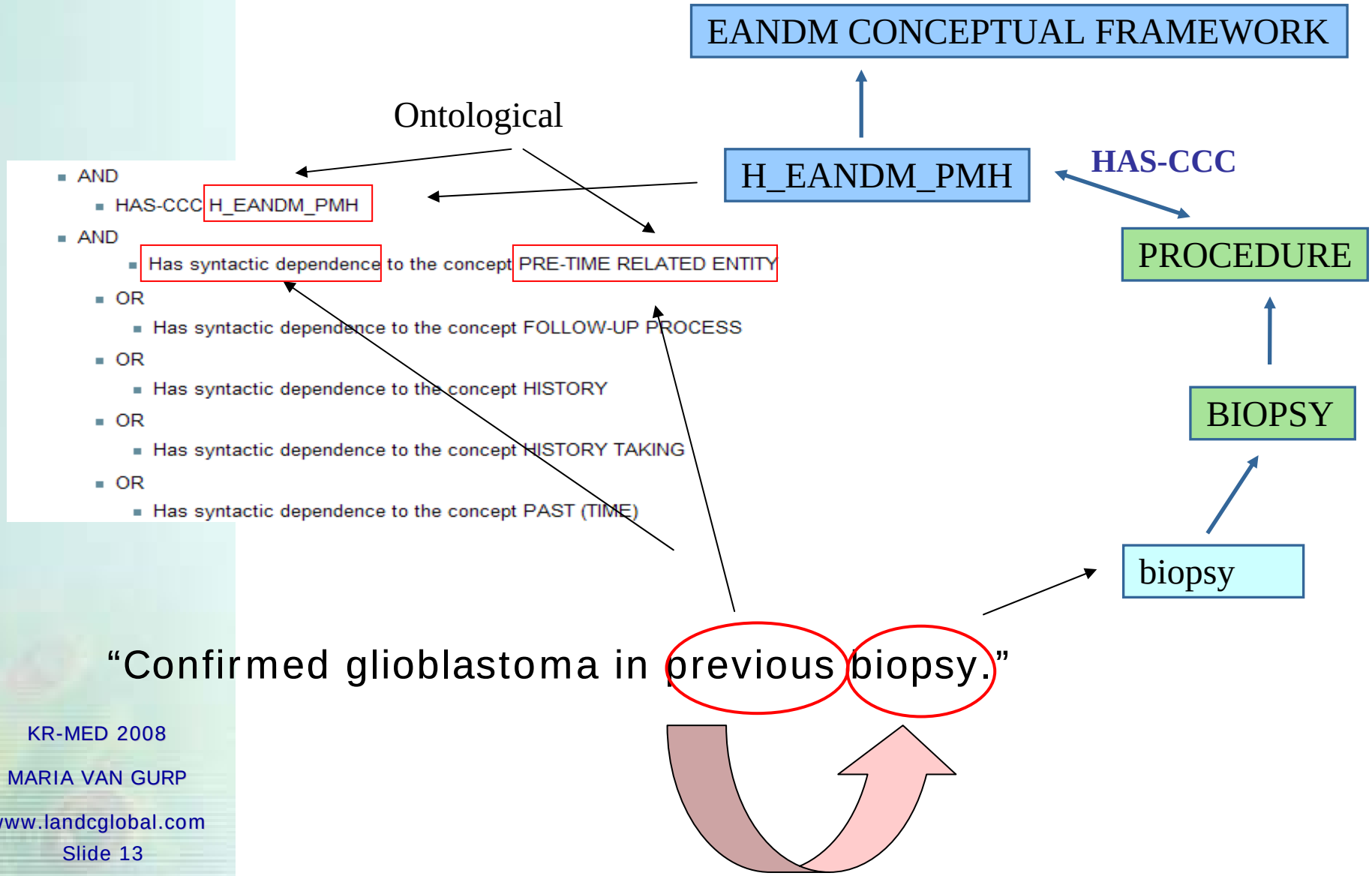
DOMAIN
Concepts specific for a given domain



SNOMED top concepts

: 105590001 : SUBSTANCE (SUBSTANCE)
: 106237007 : LINKAGE CONCEPT (LINKAGE CONCEPT)
: 123037004 : BODY STRUCTURE (BODY STRUCTURE)
: 123038009 : SPECIMEN (SPECIMEN)
: 243796009 : SITUATION WITH EXPLICIT CONTEXT (SITUATION)
: 254291000 : STAGING AND SCALES (STAGING SCALE)
: 260787004 : PHYSICAL OBJECT (PHYSICAL OBJECT)
: 272379006 : EVENT (EVENT)
: 308916002 : ENVIRONMENT OR GEOGRAPHICAL LOCATION (ENVIRONMENT / LOCATION)
: 362981000 : QUALIFIER VALUE (QUALIFIER VALUE)
: 363787002 : OBSERVABLE ENTITY (OBSERVABLE ENTITY)
: 370115009 : SPECIAL CONCEPT (SPECIAL CONCEPT)
: 373873005 : PHARMACEUTICAL / BIOLOGIC PRODUCT (PRODUCT)
: 404684003 : CLINICAL FINDING (FINDING)
410607006 : ORGANISM (ORGANISM) LinKBase® mid layer
: 419891008 : RECORD ARTIFACT (RECORD ARTIFACT)
: 48176007 : SOCIAL CONTEXT (SOCIAL CONCEPT)
71388002 : PROCEDURE (PROCEDURE) LinKBase® mid layer
: 78621006 : PHYSICAL FORCE (PHYSICAL FORCE)

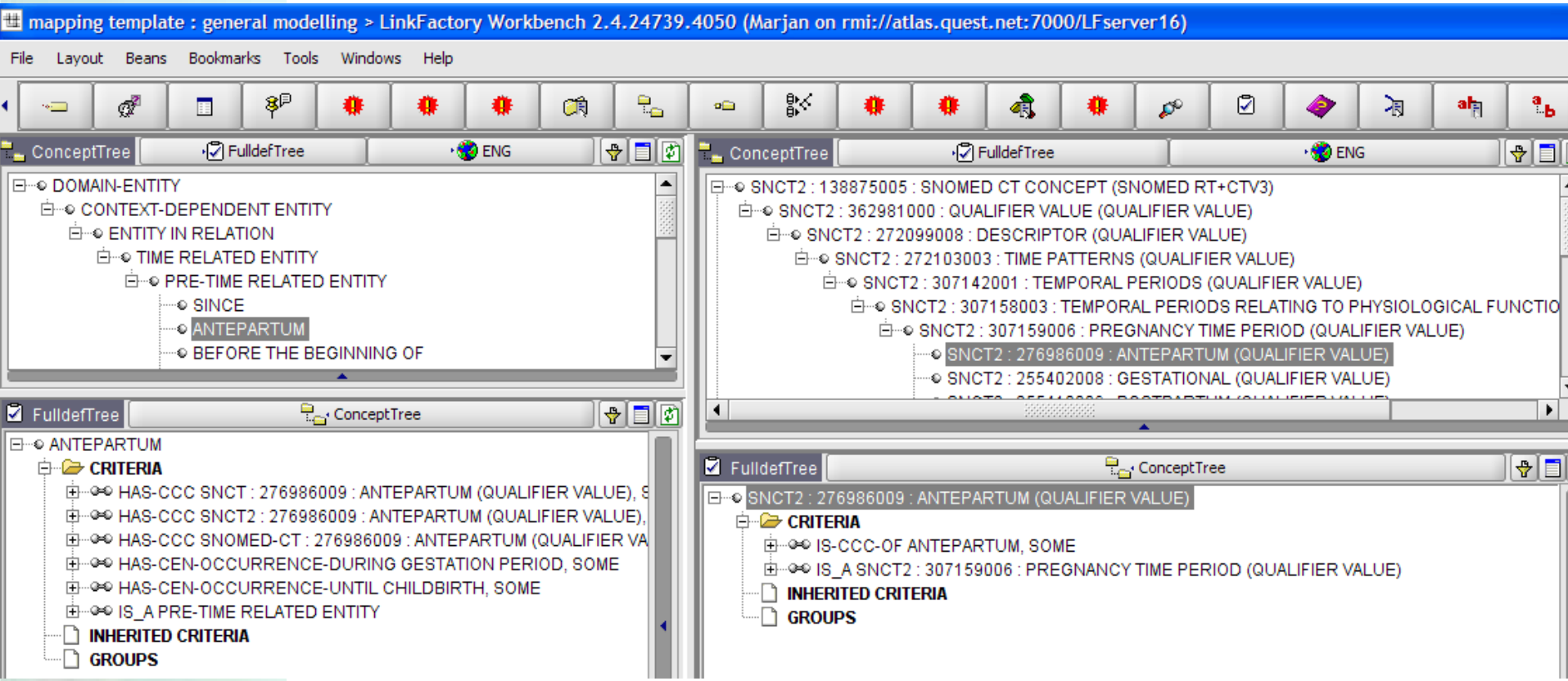
Upper layer and NLP



Upper layer and NLP

LinKBase®

SNOMED

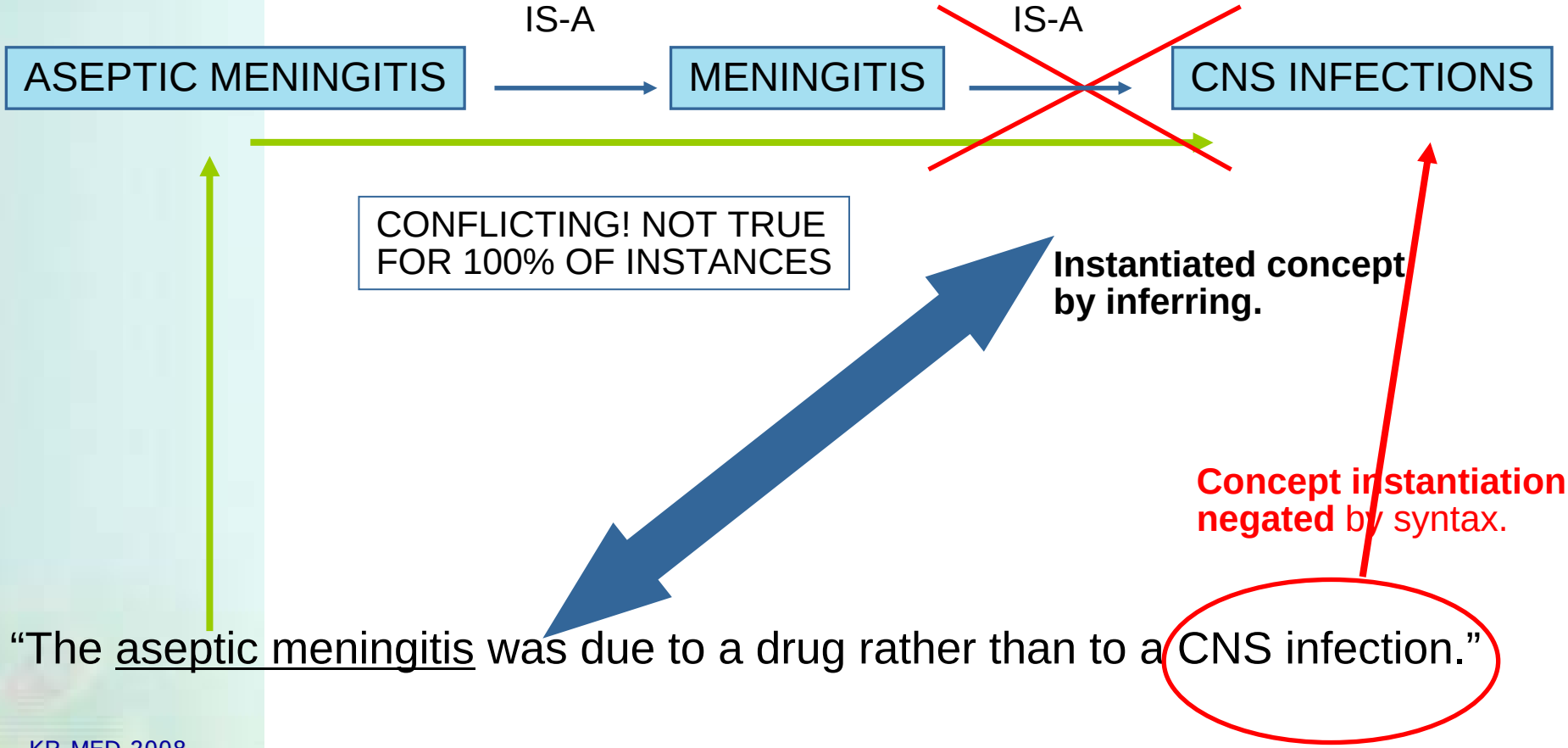


What impacts NLP?

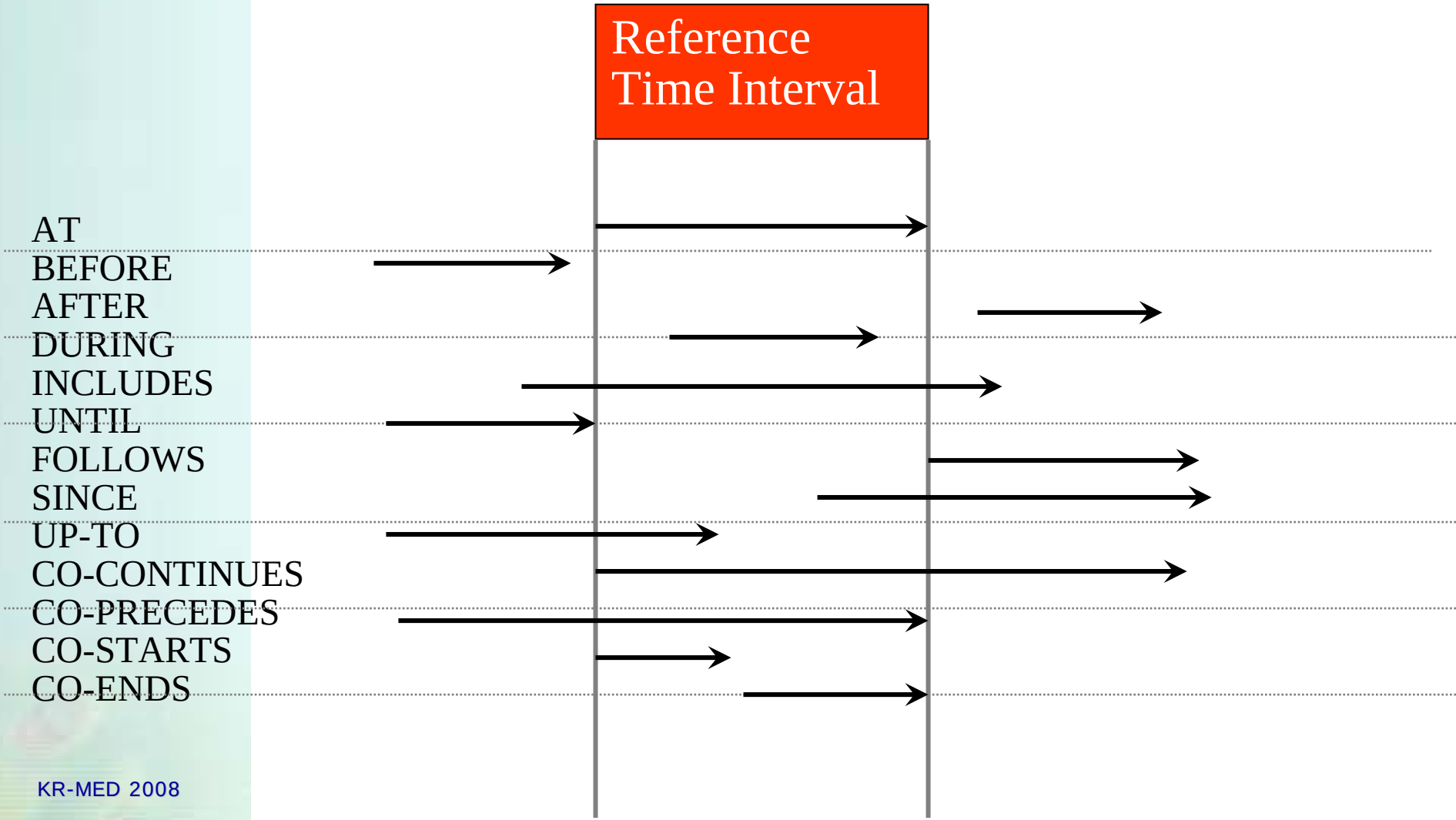
1. Top structure: presence of upper- and mid layer
2. Relationships and full definitions
3. Synonym assignments and the lexicon

Principle of 100 % true relationships

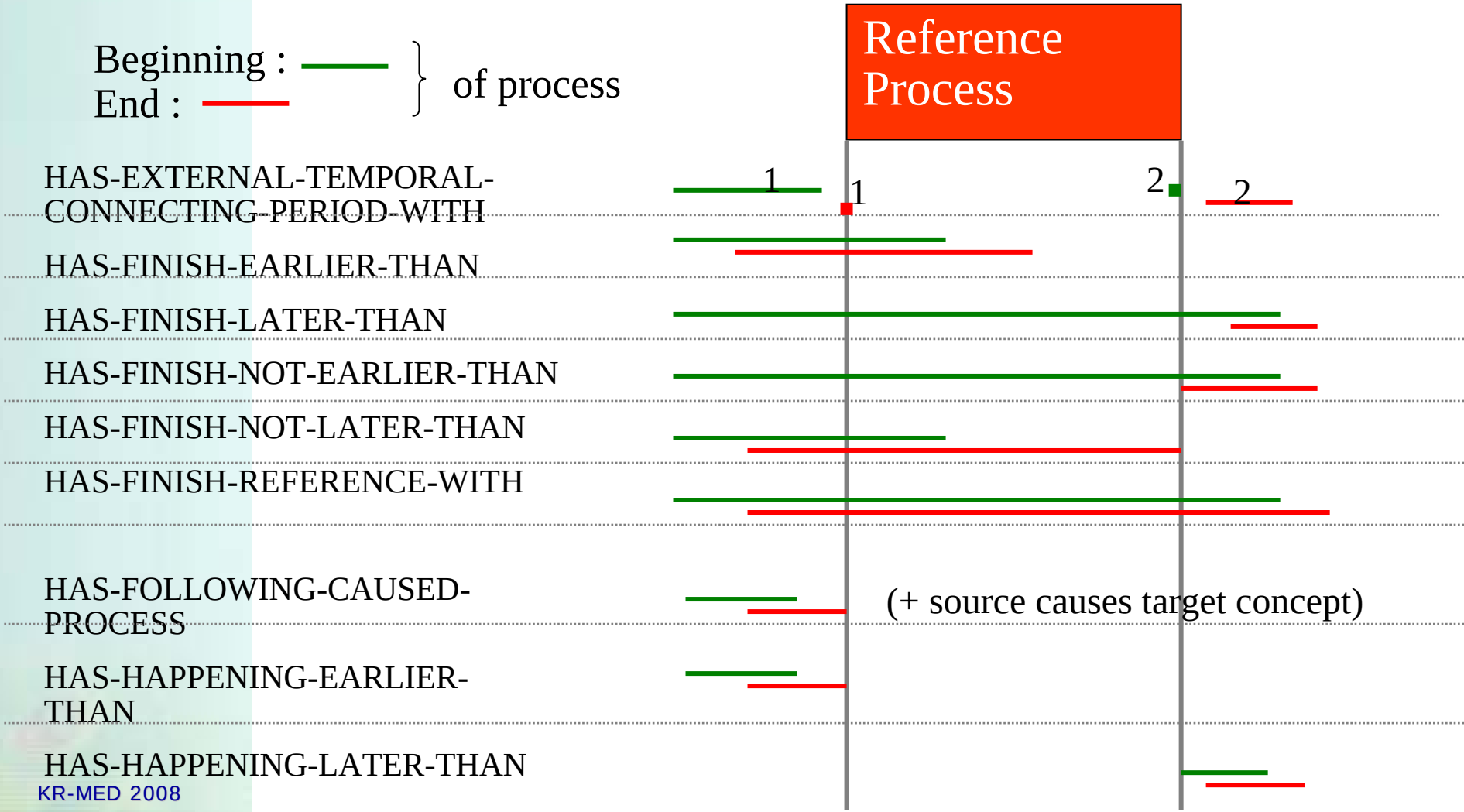
relationships should be true for all subsumed by that entity



Temporal relations diagram



Non-CEN-temporal relations



Why so many links?

= the only way to fully define a concept

Postoperative complication
- Is_A Complication of Surgical procedure
- Has-Temporal-Relation Surgical Doing

-**Has-Happening-Later-Than** Surgical Doing *(LinKBase®)*

-**After** 387713003 : Surgical procedure (procedure) *(SNOMED)*

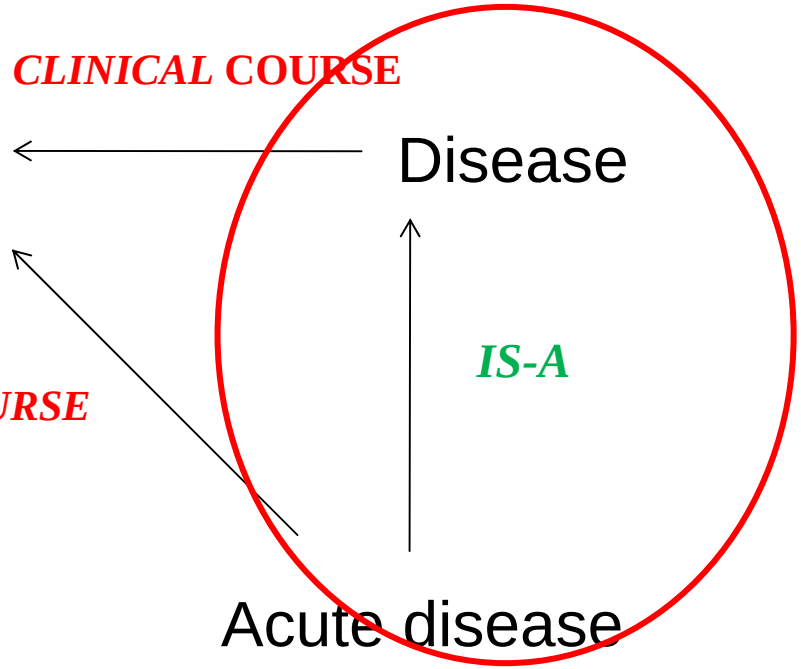
Intraoperative complication
- Is_A Complication of Surgical procedure
- Has-Temporal-Relation Surgical Doing

- Has-Cen-Occurrence-During Surgical Doing

Relationships in SNOMED

- DEFINING: logically define concepts
- REFINING: allow concepts to be connected to "qualifiers" in specific instances

Sudden onset and/or short duration (qualifier value)



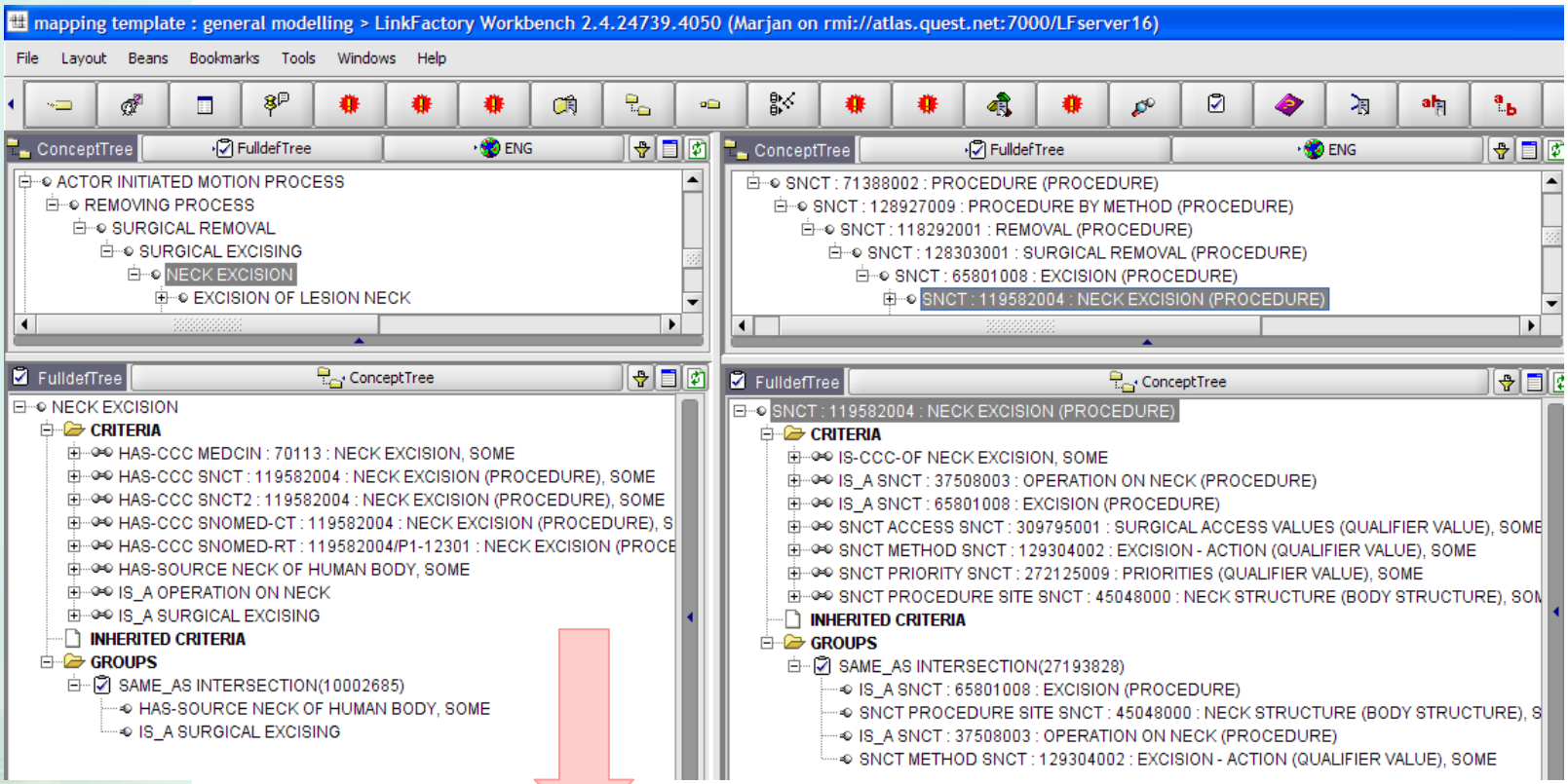
Full definitions

LinKBase®

necessary and sufficient

SNOMED

necessary



Reduction of computing time when reasoning over the ontology
Enhancement of data retrieval due to limited set of conditions to fulfill

What impacts NLP?

1. Top structure: presence of upper- and mid layer
2. Relationships and full definitions
3. Synonym assignments and the lexicon

What impacts NLP?

1. Top structure: presence of upper- and mid layer

2. Relationships and full definitions

3. Synonym assignments and the lexicon

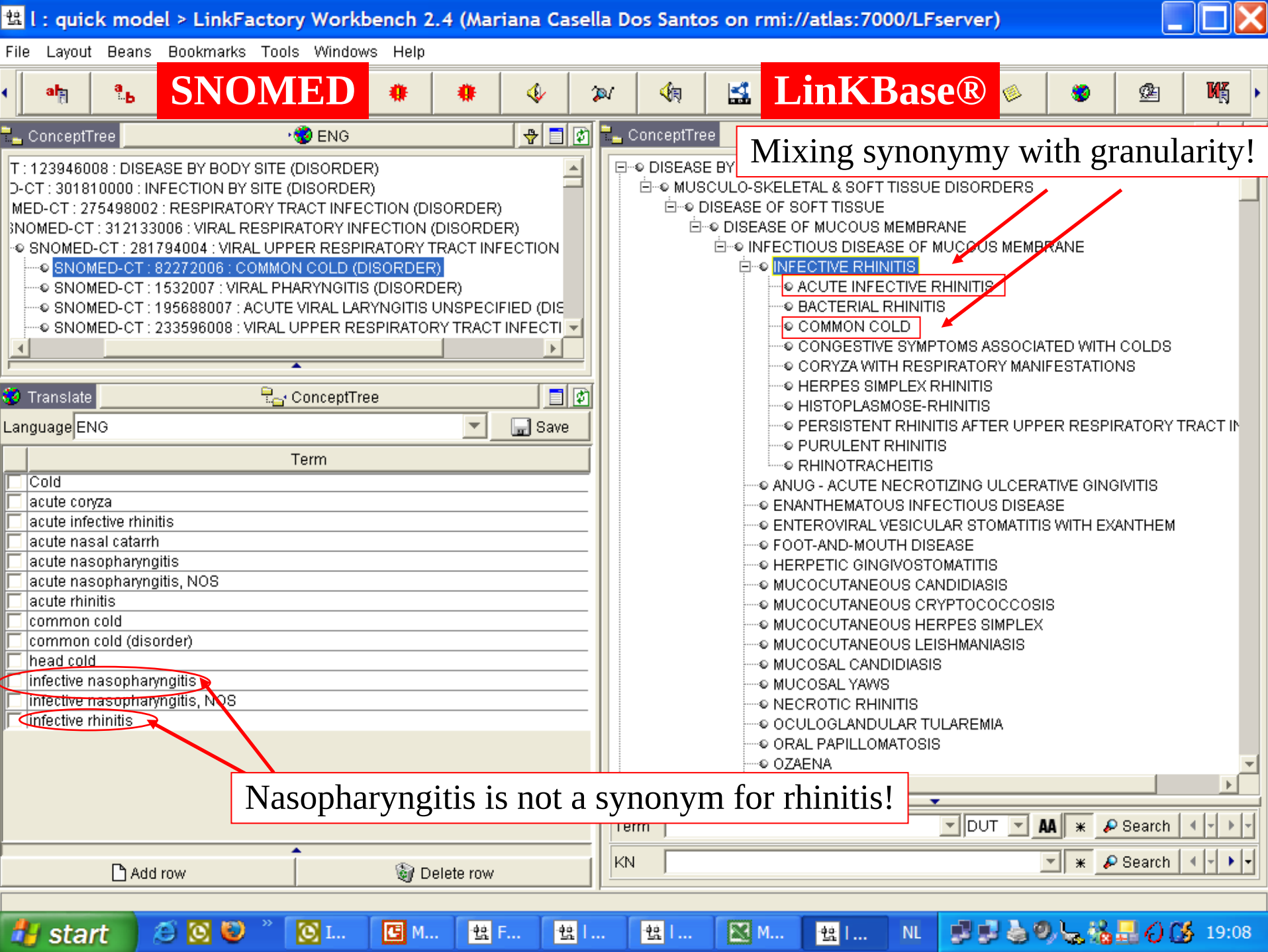
   The connection between an ontology and its lexicon is given by the term assignment to the concepts in the ontology

synonym assignment

- LinKBase®: terms are true synonyms of the ontological concepts

- SNOMED: terms can be either true synonyms or related elements either from a

  - taxonomic point of view (i.e. more generic terms assigned to more specific concepts) or
  - from a clinical point of view (e.g. symptoms assigned to a disorder)


l : quick model > LinkFactory Workbench 2.4 (Mariana Casella Dos Santos on rmi://atlas:7000/LFserver)
File Layout Beans Bookmarks Tools Windows Help
SNOMED
LinKBase®
ConceptTree
ENG
T : 123946008 : DISEASE BY BODY SITE (DISORDER)
D-CT : 301810000 : INFECTION BY SITE (DISORDER)
MED-CT : 275498002 : RESPIRATORY TRACT INFECTION (DISORDER)
NOMED-CT : 312133006 : VIRAL RESPIRATORY INFECTION (DISORDER)
SNOMED-CT : 281794004 : VIRAL UPPER RESPIRATORY TRACT INFECTION
SNOMED-CT : 82272006 : COMMON COLD (DISORDER)
SNOMED-CT : 1532007 : VIRAL PHARYNGITIS (DISORDER)
SNOMED-CT : 195688007 : ACUTE VIRAL LARYNGITIS UNSPECIFIED (DIS
SNOMED-CT : 233596008 : VIRAL UPPER RESPIRATORY TRACT INFECTI
Translate
ConceptTree
Language ENG
Save
Term
Cold
acute coryza
acute infective rhinitis
acute nasal catarrh
acute nasopharyngitis
acute nasopharyngitis, NOS
acute rhinitis
common cold
common cold (disorder)
head cold
infective nasopharyngitis
infective nasopharyngitis, NOS
infective rhinitis
Add row
Delete row
ConceptTree
Mixing synonymy with granularity!
DISEASE BY
MUSCULO-SKELETAL & SOFT TISSUE DISORDERS
DISEASE OF SOFT TISSUE
DISEASE OF MUCOUS MEMBRANE
INFECTIOUS DISEASE OF MUCOUS MEMBRANE
INFECTIVE RHINITIS
ACUTE INFECTIVE RHINITIS
BACTERIAL RHINITIS
COMMON COLD
CONGESTIVE SYMPTOMS ASSOCIATED WITH COLDS
CORYZA WITH RESPIRATORY MANIFESTATIONS
HERPES SIMPLEX RHINITIS
HISTOPLASMOSE-RHINITIS
PERSISTENT RHINITIS AFTER UPPER RESPIRATORY TRACT IN
PURULENT RHINITIS
RHINOTRACHEITIS
ANUG - ACUTE NECROTIZING ULCERATIVE GINGIVITIS
ENANTHEMATOUS INFECTIOUS DISEASE
ENTEROVIRAL VESICULAR STOMATITIS WITH EXANTHEM
FOOT-AND-MOUTH DISEASE
HERPETIC GINGIVOSTOMATITIS
MUCOCUTANEOUS CANDIDIASIS
MUCOCUTANEOUS CRYPTOCOCCOSIS
MUCOCUTANEOUS HERPES SIMPLEX
MUCOCUTANEOUS LEISHMANIASIS
MUCOSAL CANDIDIASIS
MUCOSAL YAWS
NECROTIC RHINITIS
OCULOGLANDULAR TULAREMIA
ORAL PAPILLOMATOSIS
OZAENA
Term
DUT
AA
Search
KN
Search
Nasopharyngitis is not a synonym for rhinitis!
start
NL
19:08

Synonym assignment: effect on NLP

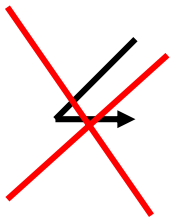
Broader term assignment

- enhances search and retrieval of related concepts

But negative effect on NLP due to

- introduction of more ambiguity
- introduction of an indirect hierarchical inconsistency (when the ontology is viewed from the lexical point of view)
- clinical associations may lead to incorrect conclusions

The acute coryza was caused by an allergic reaction



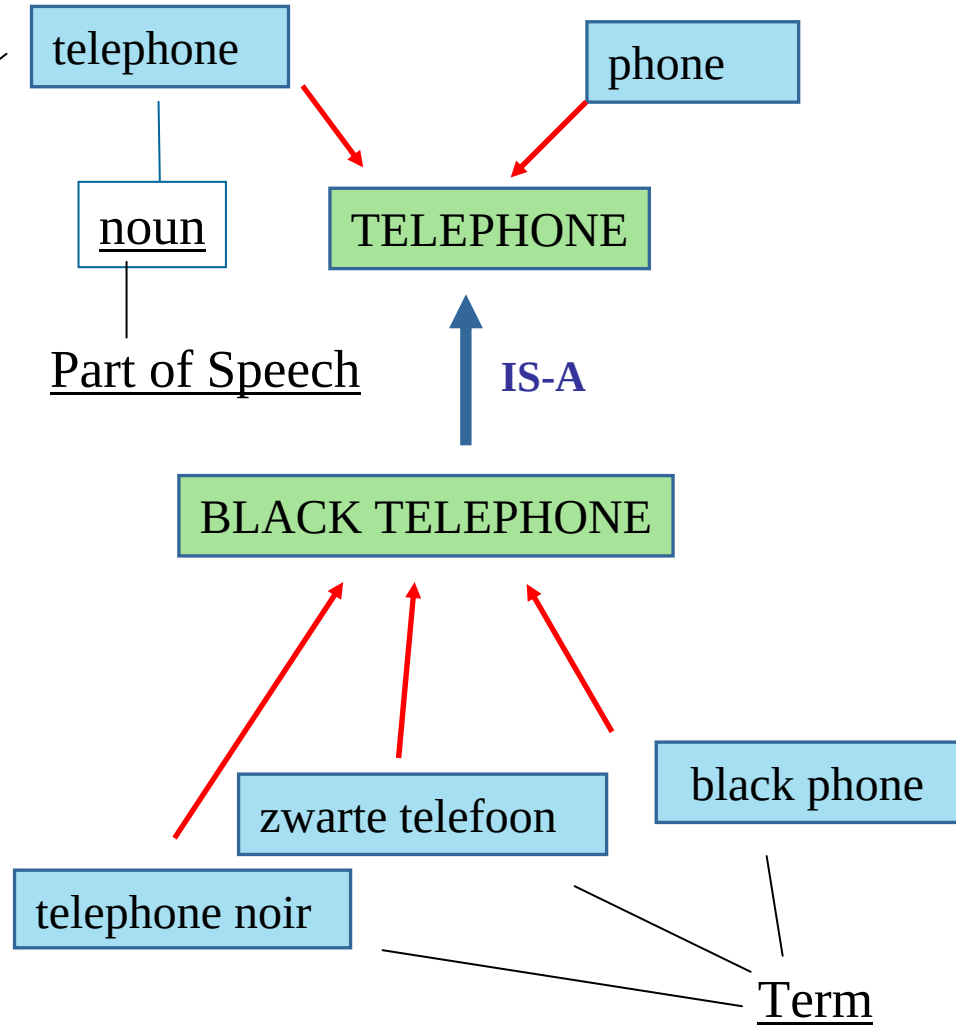
common cold

Lexicon content

LinKBase®

ROOT: telephone
POS: noun
WORD_CLASS: common
SE: plural-common telephones
SE: singular-common telephone
FEATURES: countable-noun normal-noun

ROOT: telephone
POS: verb
WORD_CLASS: lexverb
SE: gerund telephoning
SE: infinitive telephone
SE: participle telephoned
SE: past-verb telephoned
SE: thirdpersonsingular telephones
FEATURES: monotransitive nonphrasal



Conclusion

- the SNOMED structure and synonym assignment makes it very suitable for specific (clinical) applications
- LinKBase® is designed for advanced NLP, information extraction applications

SNOMED and NLP

- Adaptations to integrate SNOMED with NLP:
  - create a general top level structure that provides meaning to the lower levels, aids in resolving disambiguity and allows reasoning over the ontology
  - Extend and adapt the set of relationships to allow to define the sometimes subtle differences between concepts and the generation of full definitions
  - Only add terms which are true synonyms of the concepts and thus create a correct (lexical) hierarchy
  - Extend the lexicon; additional grammatical info will enhance use of NLP technology

SNOMED and NLP

- Adaptations to integrate SNOMED with NLP:
  - create a general top level structure that provides meaning to the lower levels, aids in resolving disambiguity and allows reasoning over the ontology
  - Extend and adapt the set of relationships to allow to define the sometimes subtle differences between concepts and the generation of full definitions
  - Only add terms which are true synonyms of the concepts and thus create a correct (lexical) hierarchy
  - Extend the lexicon; additional grammatical info will enhance use of NLP technology

SNOMED and NLP

- Adaptations to integrate SNOMED with NLP:
  - create a general top level structure that provides meaning to the lower levels, aids in resolving disambiguity and allows reasoning over the ontology
  - Extend and adapt the set of relationships to allow to define the sometimes subtle differences between concepts and the generation of full definitions
  - Only add terms which are true synonyms of the concepts and thus create a correct (lexical) hierarchy
  - Extend the lexicon; additional grammatical info will enhance use of NLP technology

SNOMED and NLP

- Adaptations to integrate SNOMED with NLP:
  - create a general top level structure that provides meaning to the lower levels, aids in resolving disambiguity and allows reasoning over the ontology
  - Extend and adapt the set of relationships to allow to define the sometimes subtle differences between concepts and the generation of full definitions
  - Only add terms which are true synonyms of the concepts and thus create a correct (lexical) hierarchy
  - Extend the lexicon; additional grammatical info will enhance use of NLP technology

Questions?