Valency in Other LRs:
PropBank, VerbNet, FrameNet

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK

lopatkova@ufal.mff.cuni.cz

Proposition Bank

Annotations in PropBank

- based on Penn TreeBank
- transfer sentences to propositions
- goal is to annotate *every tree* systematically
  - so statistics in the corpus are meaningful
- generally more data-driven & bottom up
  - no level of abstraction beyond verb senses
  - annotate every verb you see
- Martha Palmer, Mitch Marcus, Joseph Rosenzweig, Paul Kingsbury, Hoa Dang, Karin Kipper, Scott Cotton, Laren Delfs, Christiane Fellbaum

http://verbs.colorado.edu/~mpalmer/projects/ace.html

Proposition Bank: Generalizing from Sentences to Propositions

Powell met Zhu Rongji

Powell and Zhu Rongji met

Powell met with Zhu Rongji

Powell and Zhu Rongji had a meeting

. . .

battle
wrestle
join
debate
consult

Proposition: meet(Powell, Zhu Rongji)

meet(Somebody1, Somebody2)

When Powell met Zhu Rongji on Thursday they discussed the return of the spy plane.

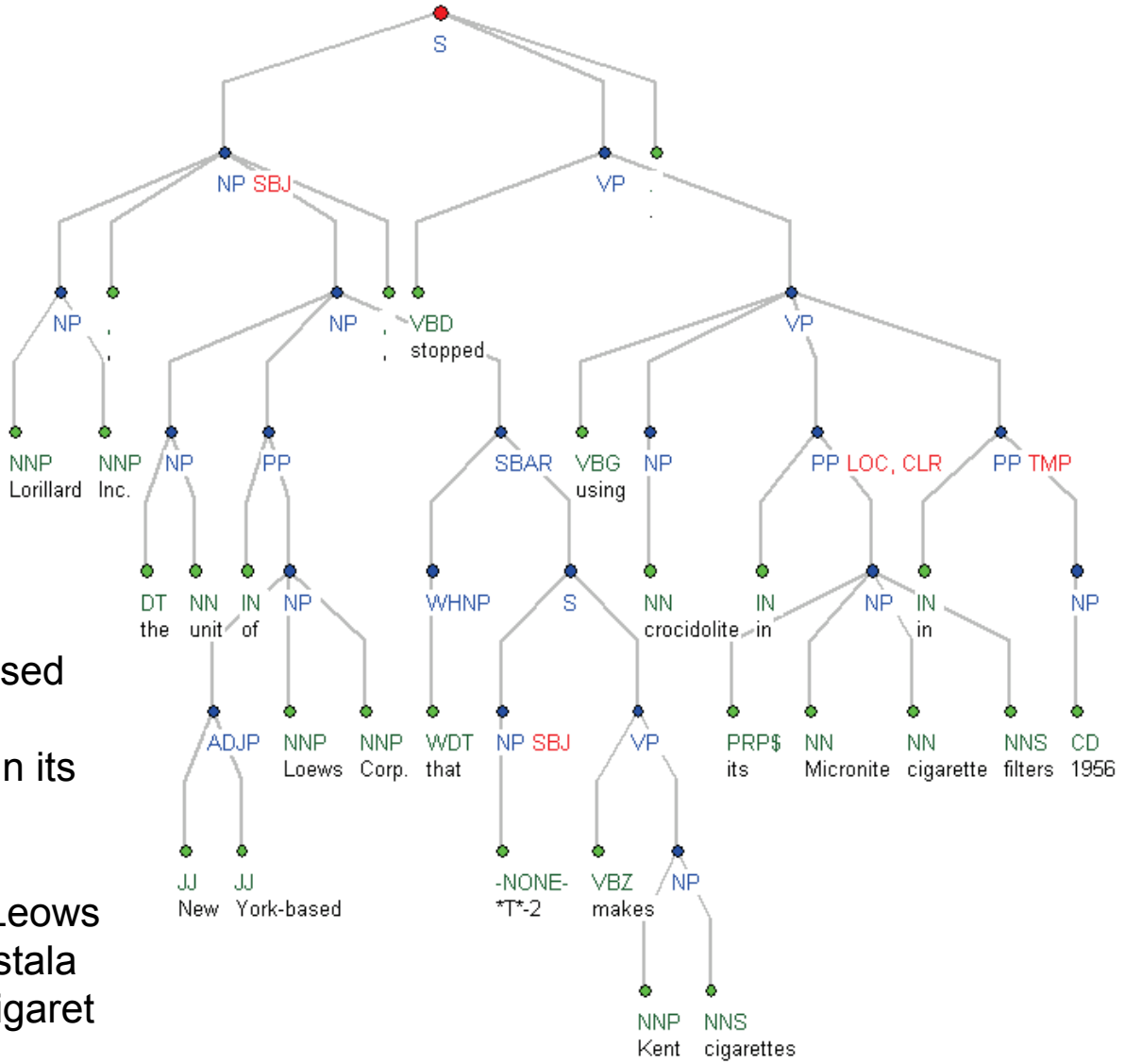
meet(Powell, Zhu) discuss([Powell, Zhu], return(X, plane))

Penn English Treebank

- 1.3 million words
- Wall Street Journal and other sources
- tagged with Part-of-Speech
- syntactically Parsed
- widely used in NLP community
- available from Linguistic Data Consortium

- the same data as in the (eng. part of) PCEDT

Penn English Treebank

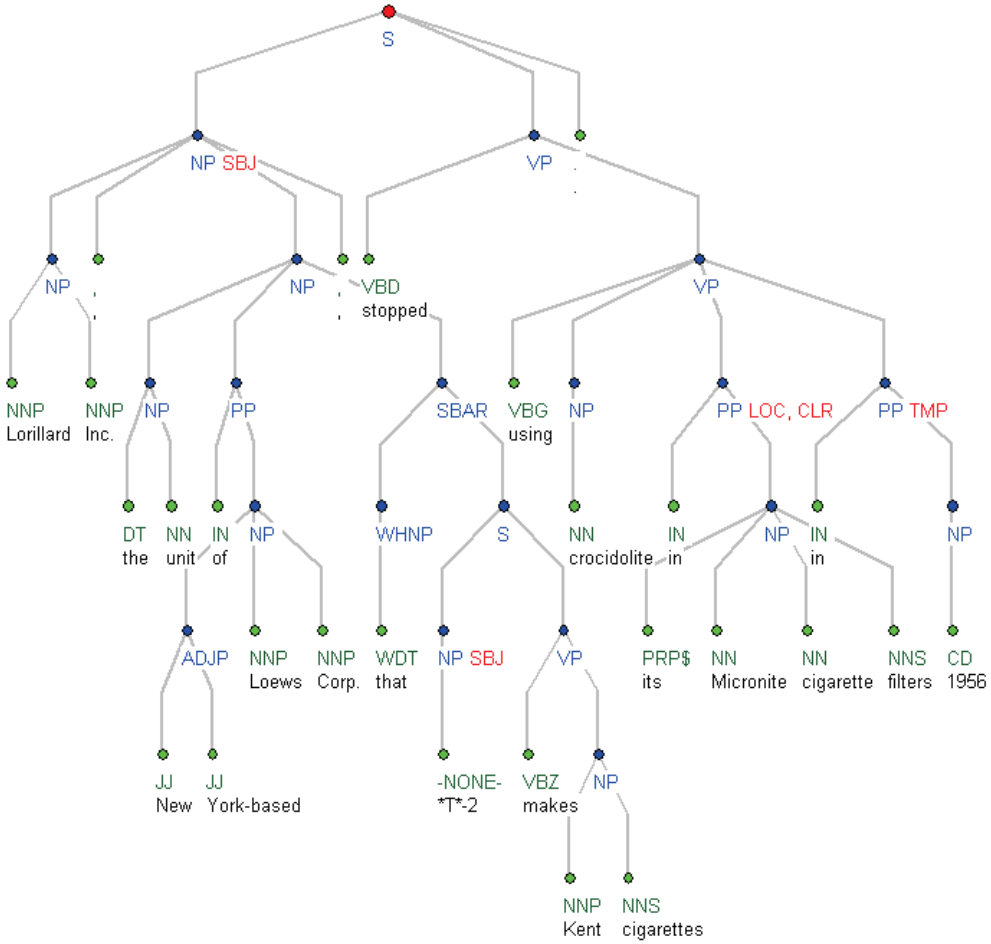


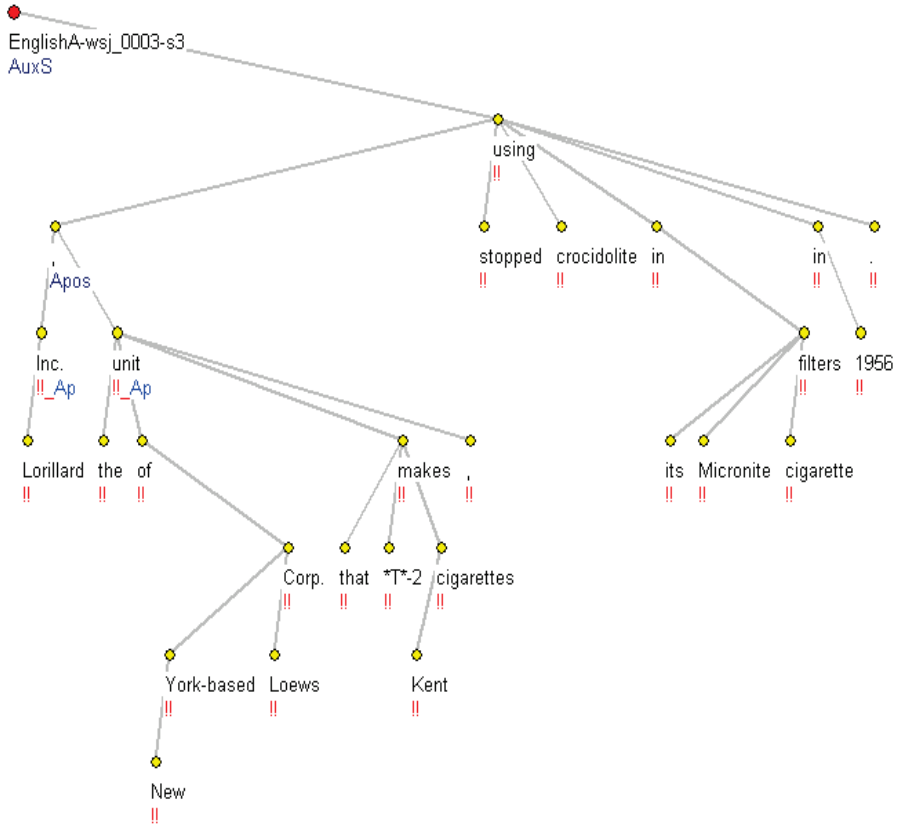
Lorillard Inc., the unit of New York-based Loews Corp. that *T*-2 makes Kent cigarettes, stopped using crocidolite in its Micronite cigarette filters in 1956.

[Lorillard, pobočka newyorské firmy Leows corp., která vyrábí cigarety Kent, přestala používat crocidolit ve filtrech svých cigaret Micronite v roce 1956.]

Penn English Treebank

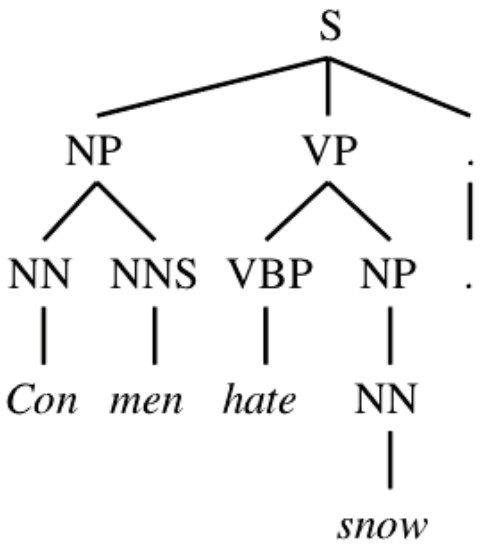
original annotation → a-layer from PCEDT

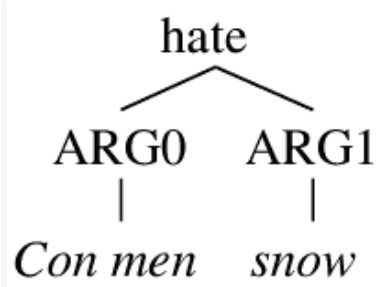




Transfer sentences to propositions

- Penn TreeBank → PropBank
  - Add a semantic layer on Penn TreeBank
  - Define a set of semantic roles for each verb
  - Each verb's roles are numbered



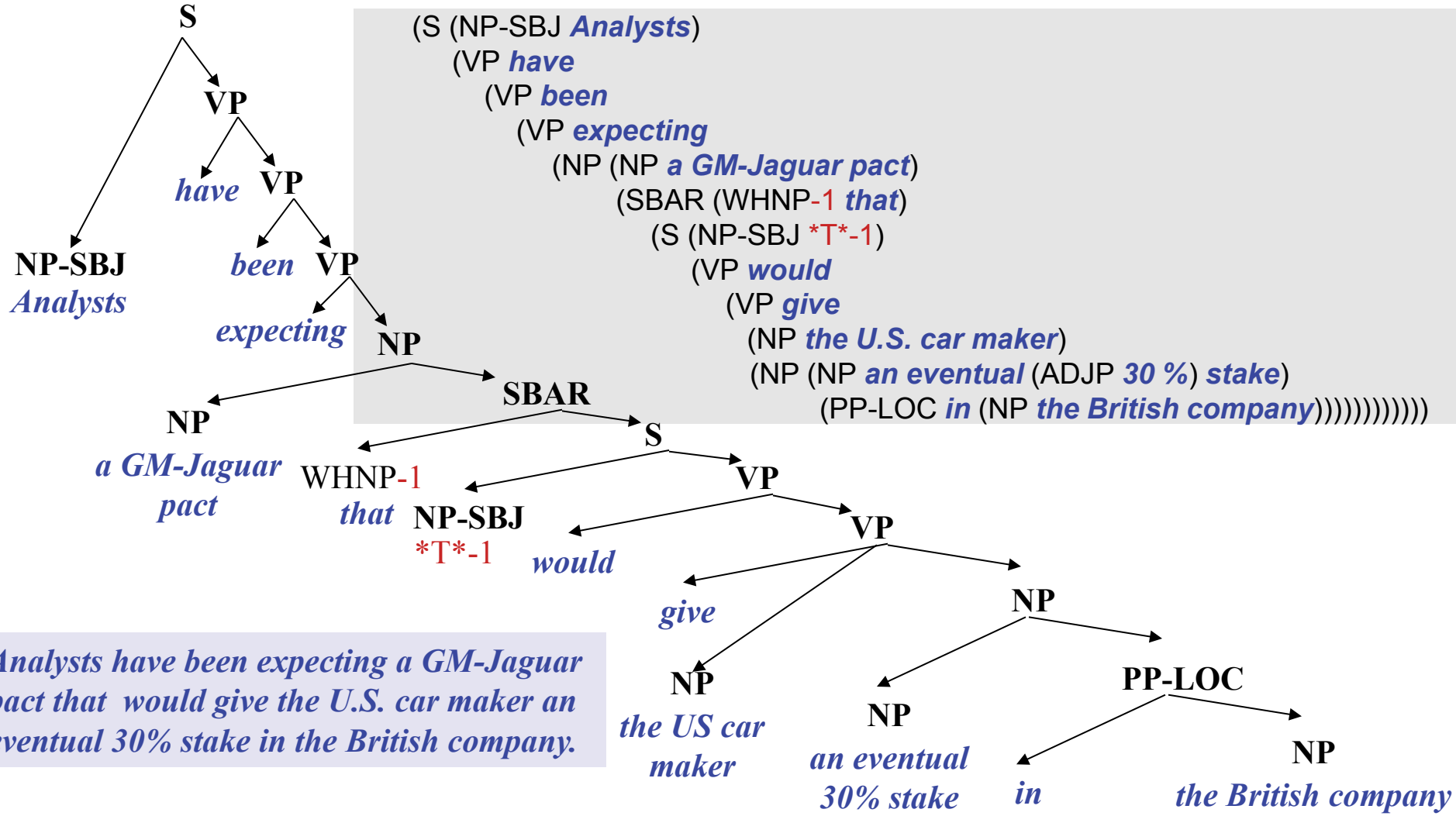


Frameset: *hate.01*

ARG0: *experiencer*

ARG1: *target*

A TreeBanked Sentence

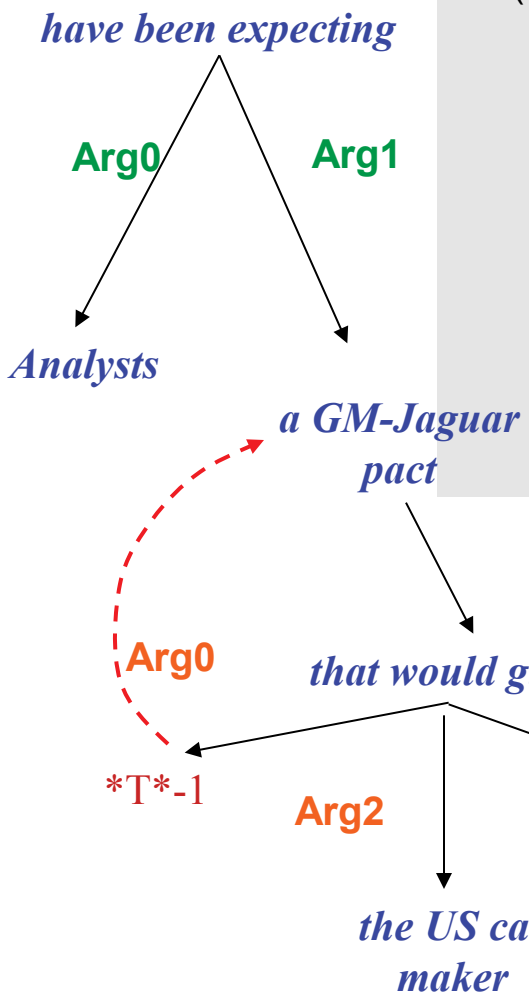



```
(S (NP-SBJ Analysts)
   (VP have
      (VP been
         (VP expecting
            (NP (NP a GM-Jaguar pact)
                (SBAR (WHNP-1 that)
                   (S (NP-SBJ *T*-1)
                      (VP would
                         (VP give
                            (NP the U.S. car maker)
                            (NP (NP an eventual (ADJP 30 %) stake)
                                (PP-LOC in (NP the British company))))))))))))
```

Analysts have been expecting a GM-Jaguar pact that would give the U.S. car maker an eventual 30% stake in the British company.

[Analytici očekávali smlouvu mezi GM a Jaguárem, která by zaručovala výrobci aut případnou 30% investici v britské firmě.]

The same sentence, PropBanked



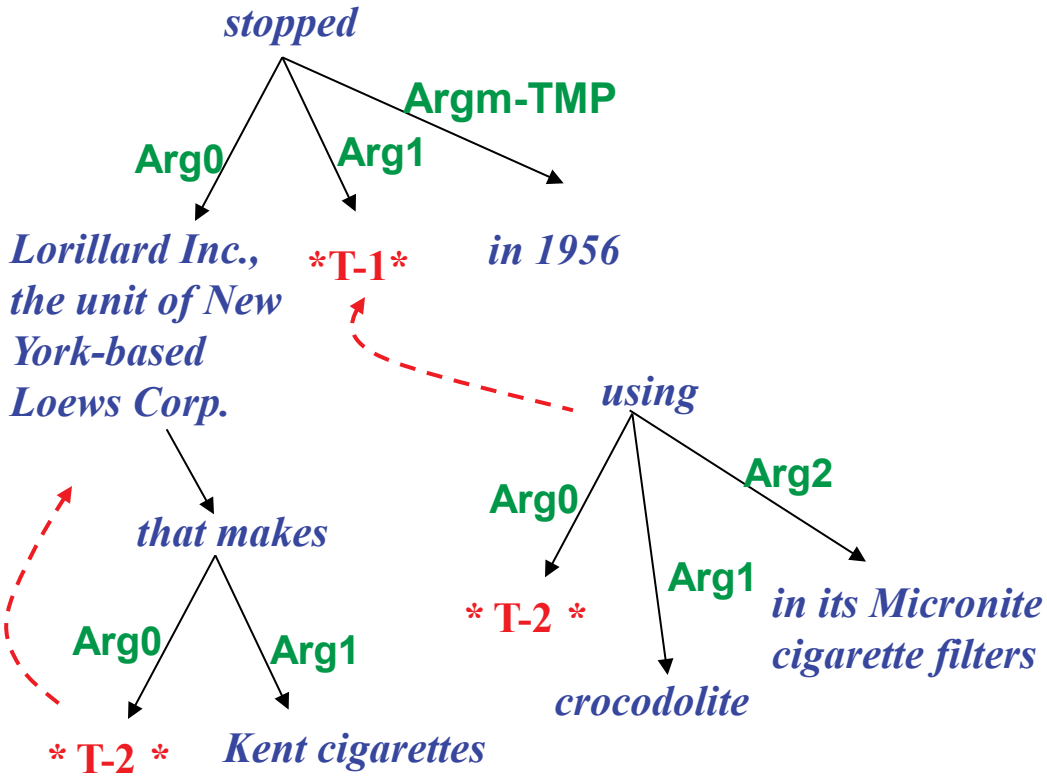

```
(S Arg0 (NP-SBJ Analysts)
    (VP have
      (VP been
        (VP expecting
          Arg1 (NP (NP a GM-Jaguar pact)
                   (SBAR (WHNP-1 that)
                     (S Arg0 (NP-SBJ *T*-1)
                       (VP would
                         (VP give
                           Arg2 (NP the U.S. car maker)
                             Arg1 (NP (NP an eventual (ADJP 30 %) stake)
                                      (PP-LOC in (NP the British company)))))))))))))
```

expect(Analysts, GM-J pact)
give(GM-J pact, US car maker, 30% stake)

[Analytici očekávali smlouvu mezi GM a Jaguárem, která by zaručovala výrobci aut případnou 30% investici v britské firmě.]

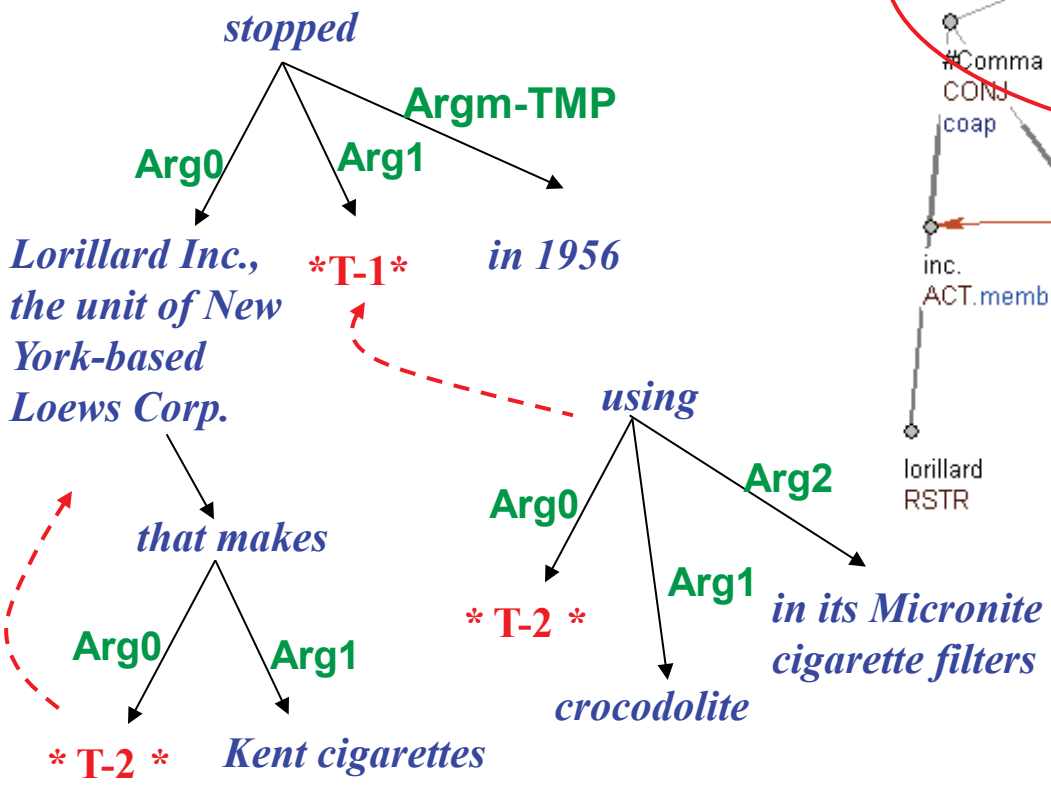
PropBank annotation

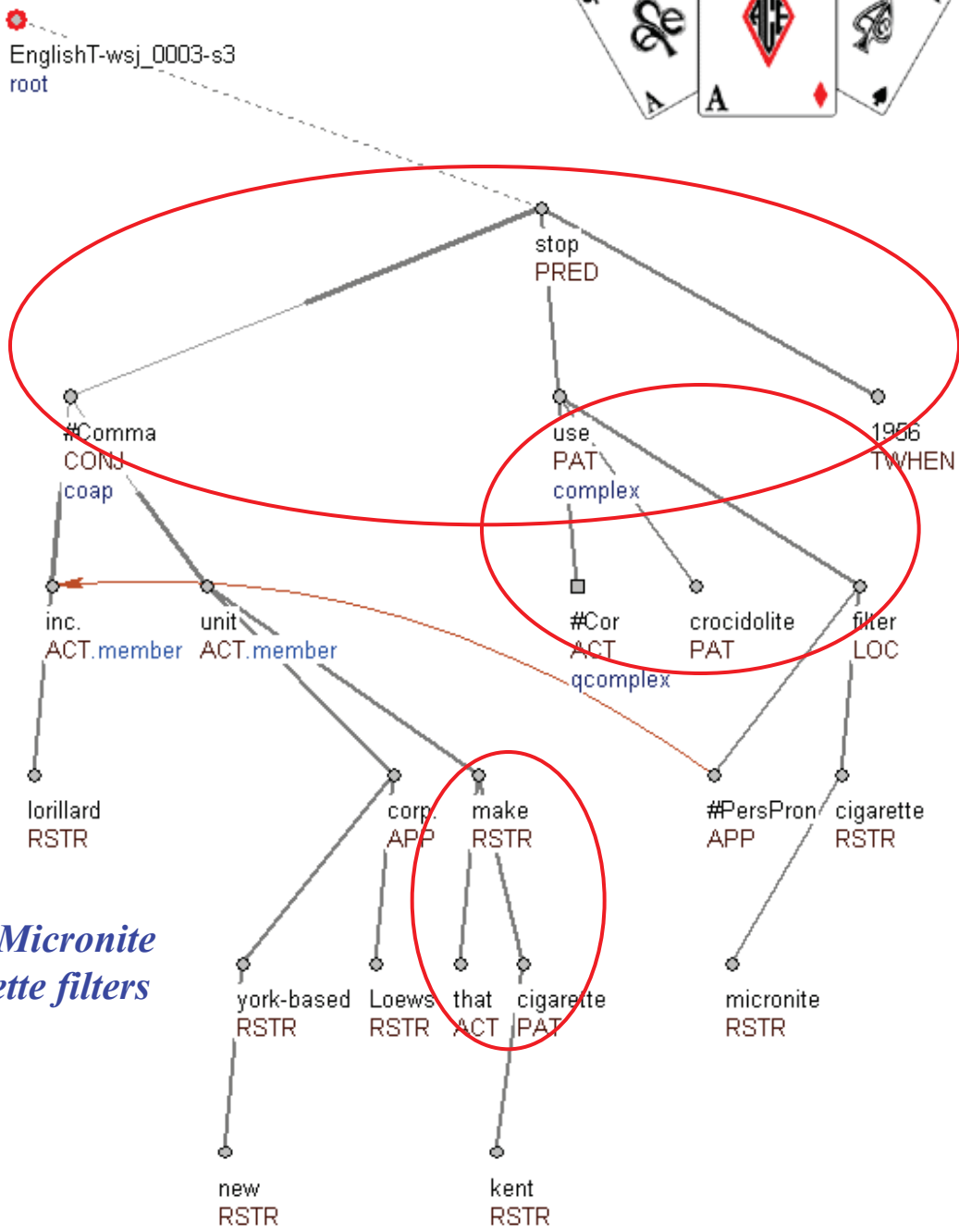
Lorillard Inc., the unit of New York-based Loews Corp. that *T*-2 makes Kent cigarettes, stopped using crocidolite in its Micronite cigarette filters in 1956.



→ PDT t-layer

Lorillard Inc., the unit of New York-based Loews Corp. that *T*-2 makes Kent cigarettes, stopped using crocidolite in its Micronite cigarette filters in 1956.





PropBank: Predicate Annotation

- a general set of semantic roles for all types of predicates
- semantic roles for each verb and sense in the frame files
- the (core) arguments are labeled by numbers
- adjunct-like arguments – universal to all verbs

PropBank: Predicate Annotation

- **argument types** defined on a "per-verb" basis
- consistent across uses of a single verb (sense)
- but the same tags are used (Arg0, Arg1, Arg2, …)

Arg0 ≈proto-typical agent (Dowty)

Agent (85%), Experiencer (7%), Theme (2%), …

Arg1 ≈proto-typical patient

Theme (47%),Topic (23%), Patient (11%), …

Arg2= Recipient (22%), Extent (15%), Predicate (14%), …

Arg3= Asset (33%), Theme2 (14%), Recipient (13%), …

Arg4= Location (89%), Beneficiary (5%), …

Arg5= Location (94%), Destination (6%)

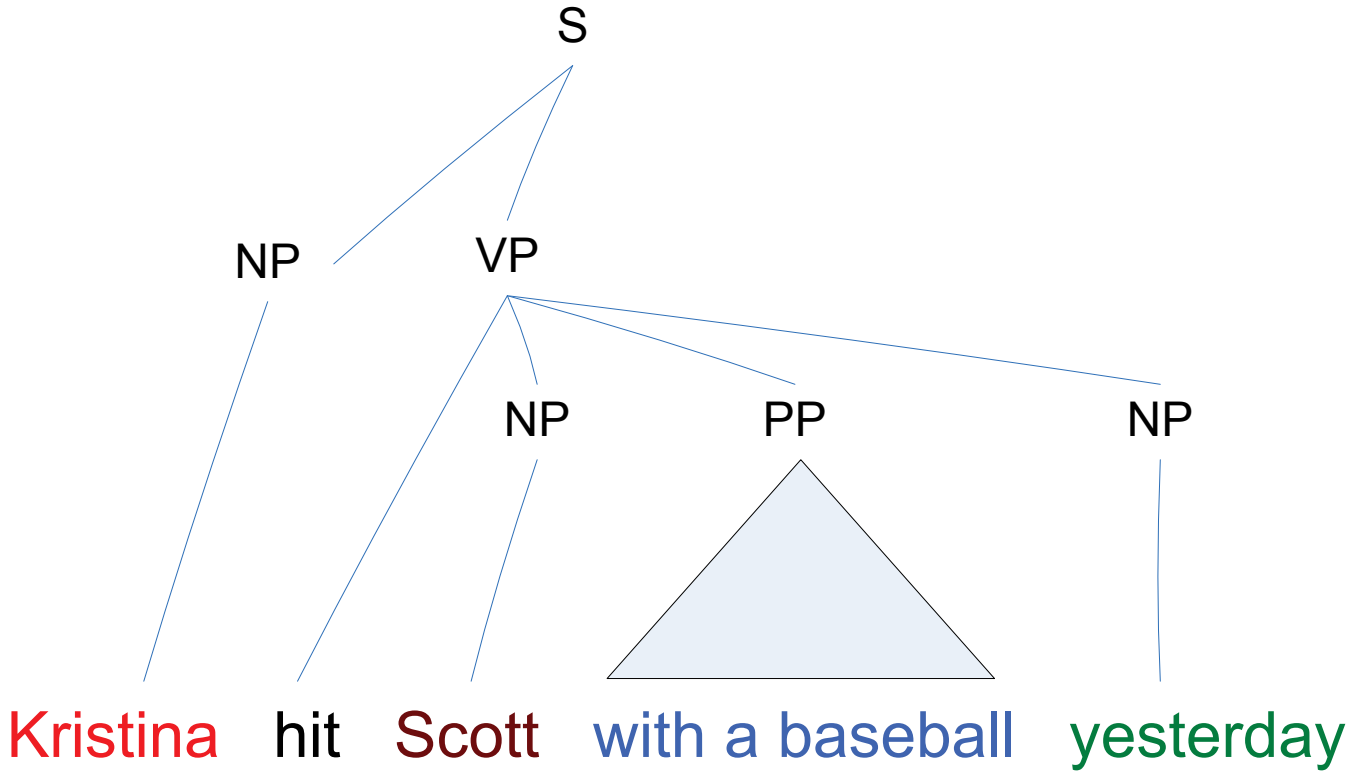
PropBank: Predicate Annotation

- **adjunct-like arguments** – universal to all verbs:
  - LOC (where at), TMP (when), DIR (where to), MNR (how), ADV (others), PRP (why)
  - REC (himself, themselves, each other), PRD (this argument refers to or modifies another)
  - NEG, MOD (modals and negation marked as ArgMs)
  - EXT (amount), PNC, DIS

Frame Examples

- Transfer sentences to propositions

  Kristina hit Scott → hit(Kristina,Scott,with a baseball)



Frame Examples

- Transfer sentences to propositions

  Kristina hit Scott → hit(Kristina,Scott,with a baseball)

- hit.01 “strike”

  Arg0: agent, hitter;

  Arg1: thing hit;

  Arg2: instrument, thing hit by or with

  [$_{A0}$ *Kristina*] ***hit*** [$_{A1}$ *Scott*] [$_{A2}$ *with a baseball*] *yesterday*.

PropBank: Predicate Annotation

- look.02 “seeming”
  - Arg0: seemer;
  - Arg1: seemed like;
  - Arg2: seemed to
  - [A0 *It*] ***looked***$_{REL}$ [A2 *to her] like* [A1 *he deserved this*].
- deserve.01 “deserve”
  - Arg0: deserving entity;
  - Arg1: thing deserved;
  - Arg2: in-exchange-for
  - *It looked to her like* [A0 *he*] ***deserved***$_{REL}$ [A1 *this*].
- expect.0? “anticipate”
  - Arg0: expecter;
  - Arg1: thing expected
  - [A0 *Portfolio managers*] ***expect***$_{REL}$ [A1 *further declines in interest rates*].
- give.01 “transfer”
  - Arg0: giver;
  - Arg1: thing given;
  - Arg2: entity given to
  - [A0 *The executives*] ***gave***$_{REL}$ [A2 *the chefs*] [A1 *a standing ovation*].

PropBank: Argument Numbering

- task: not just undoing passives

  [A0 *The earthquake*] *shook* [A1 *the building*].
  [A1*The walls*] *shook*; [A1 *the building*] *rocked*.

- how are arguments numbered?
  - Examination of example sentences
  - Determination of required / highly preferred elements
  - Sequential numbering, Arg0 is typical first argument

PropBank: Argument Numbering

- task: not just undoing passives

  [A0 *The earthquake*] *shook* [A1 *the building*].
  [A1 *The walls*] *shook*; [A1 *the building*] *rocked*.

- how are arguments numbered?
  - Examination of example sentences
  - Determination of required / highly preferred elements
  - Sequential numbering, Arg0 is typical first argument

BUT:
- ergative/unaccusative verbs (shake example)

rise.01 “go up”
Arg1: Logical subject, patient, thing rising;
Arg2: EXT, amount risen;
Arg3* = start point
Arg4 = end point

Example: [A1 *Sales*] *rose*$_{REL}$ [A2 *4%*] [A4 *to $3.28 billion*] [A3 *from $3.16 billion*].

- arguments mapped for "synonymous" verbs
  (consistent between different verbs that share similar argument structures)

VerbNet

VerbNet – main characteristics

- the largest on-line verb lexicon for English
- Kipper-Schuler 2006
- main characteristics:
  - hierarchical
  - domain-independent
  - broad-coverage
- mappings to other lexical resources
  - PropBank … 90.86% coverage of PropBank tokens
  - WordNet (Miller, 1990; Fellbaum, 1998)
  - FrameNet (Baker et al., 1998)
  - Lexicalized Tree Adjoining Grammar (XTAG Research Group, 2001)

Organization of VerbNet

- organized into verb classes extending Levin (1993) classes
- syntactic and semantic coherence among members of a class
- verb class in VN is described by
  - thematic roles
  - selectional restrictions on the arguments
  - frames = syntactic description + semantic predicates

VerbNet entry: Hit-18.1 class (simplified)

Class Hit-18.1			
Roles and Restrictions: Agent[+int_control] Patient[+concrete] Instrument[+concrete]			
Members: bang, bash, hit, kick, ...			
Frames:			
Name	Example	Syntax	Semantics
Basic Transitive	Paula hit the ball	Agent V Patient	• cause(Agent, E) • manner (during(E), directedmotion, Agent) • !contact (during(E), Agent, Patient) • manner(end(E),forceful, Agent) • contact(end(E), Agent, Patient)

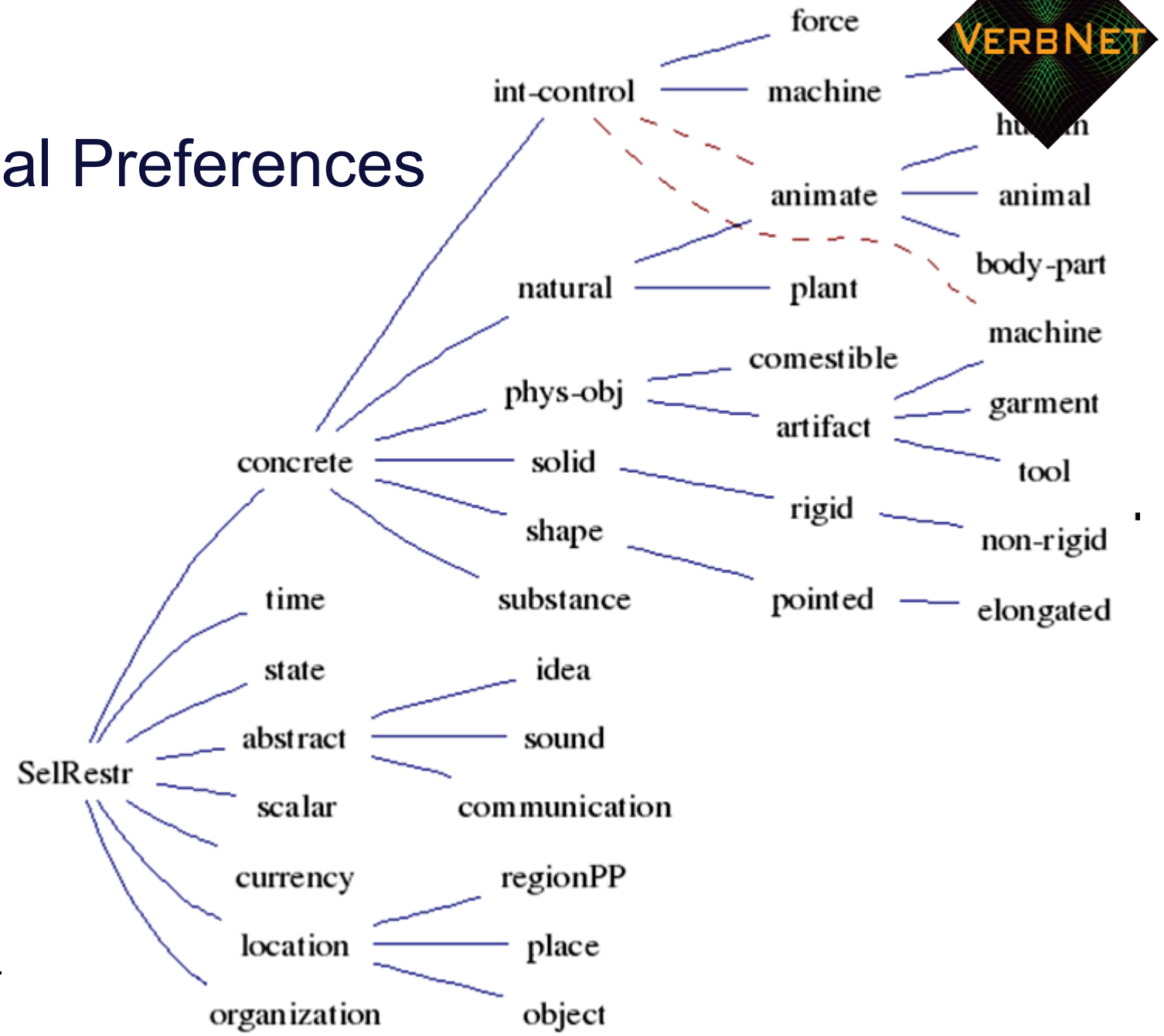
VerbNet – basic statistics

	VerbNet
First-level classes	274
Thematic roles	23
Semantic predicates	94
Syntactic restrictions (on sentential compl)	55
Number of verb senses	5257
Number of lemmas	3769

VerbNet – Thematic Roles

- Actor
- Agent
- Asset
- Attribute
- Beneficiary
- Cause
- Location, Destination, Source
- Destination
- Source
- Location
- Experiencer
- Extent
- Instrument
- Material and Product
- Material
- Product
- Patient
- Predicate
- Proposition
- Recipient
- Stimulus
- Theme
- Time
- Topic

VerbNet
Selectional Preferences



VerbNet – Unified Verb Index

- complete list of the thematic roles, selectional and syntactic restrictions, predicates, and frame types:

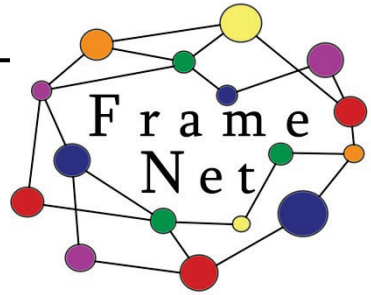
  http://verbs.colorado.edu/verb-index/reference

- mappings:
  - VerbNet senses to FrameNet senses

    http://verbs.colorado.edu/verb-index/vn-fn.xml
  - VerbNet Role to FrameNet Frame Elements

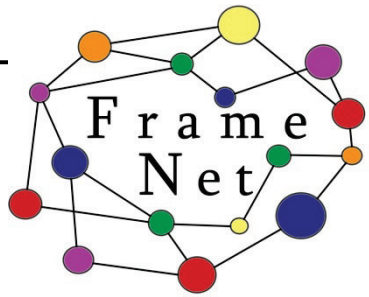
    http://verbs.colorado.edu/verb-index/vn-fn-roles.xml
  - SemLink: PropBank, FrameNet, VerbNet

    http://verbs.colorado.edu/semlink/



FrameNet

FrameNet Background: Case Theory



- Fillmore: Case Theory (1968)
  - deep structure
  - surface-structure cases are derived
  - universal set of roles

A(gentive)
I(nstrumental)
D(ative)
F(actitive)
L(ocative)
O(bjective)

(i) *The door opened.* [__O]
(ii) *John opened the door.* [__O+A]
(iii) *The wind opened the door.* [__O+I]
(iv) *John opened the door with a chisel.* [__O+A+I]

→ **[__O (A) (I)]**

- benefits
  - fewer tokens: fewer verb senses
  - verbs with the same semantics, but with different subject selection preferences

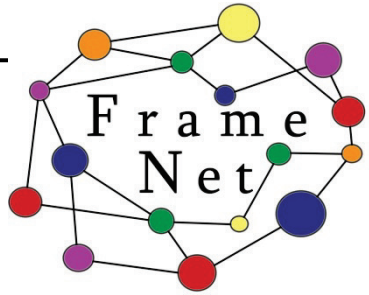
like, *please* …[__O+D]

  - only noun phrases of the same case may be conjoined

**John and a hammer broke the window. *The car broke the window with a fender*

- drawbacks
  - limited set of cases
  - mapping rules syntax → semantics

From Case Theory to FrameNet



University of Berkeley
http://framenet.icsi.berkeley.edu/

- on-line lexical resource for English
- verbs, nouns, adjectives, prepositions
- the aim:

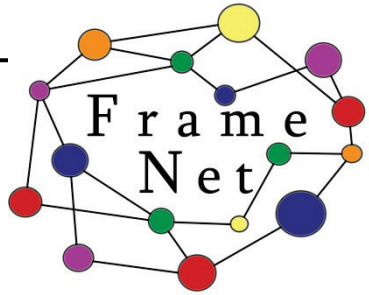
  to document the range of semantic and syntactic combinatory possibilities- valences -of each word in each of its senses
- FrameNet data:

  more than 13,000 lexical units,

  more than 1,200 hierarchically-related semantic frames

  exemplified in more than 200,000 annotated sentences

From Case Theory to FrameNet



- Frame semantics
  semantic frames and its frame elements
- Semantic frames are evoked by lexical units

- Semantic frame:
  a conceptual structure that describes a particular type of situation, object, or event
- Frame elements:
  components of semantic frames, semantic roles
- each word's valence possibilities,
  the ways in which information about the frames is provided in the linguistic structures

FrameNet Data

FrameNet data is available online as browsable reports, a clickable visualization, and a searchable database.
You can also download the data in XML format (OWL format is also available in the current data release).

Note that the data shown here is **newer** than the data in the data release, but has not been as carefully checked for completeness and consistency.

Sample output from FrameGrapher

Full Text Annotation

In addition to our ongoing lexicographic work, FrameNet has begun to annotate some continuous texts, as a demonstration of how frame semantics can contribute to text understanding. This style of annotation typically involves marking frame elements of frames evoked by multiple predicators in each sentence or even in each clause.

Frame Index

Frame definitions, semantic roles/frame elements (FEs), and other frame information.

Lexical Unit Index

Word senses (Lexical Units) with annotation and related syntactic patterns report.

Communication

definition of the semantic frame

Definition:

A Communicator conveys a Message to an Addressee; the Topic and Medium of the communication also may be expressed. This frame includes no specification of the method of communication (speech, writing, gesture, etc.). The frames that inherit the general Communication frame can add elaboration to the Medium in a variety of ways (*in French, on the radio program, in a letter*) or to the Manner of communication (*babble, rant, shout, whisper*). There are also frames that either do not inherit all of the FEs of this frame or do not inherit them in a straightforward manner (such as Conversation, in which Communicator and Addressee alternate roles, and are often expressed by a single, plural NP).

Let's hope it did n't have time to COMMUNICATE any of its findings to its overlords .

Communication

definition of the frame elements

FEs:

Core:

Communicator [Com]
Semantic Type Sentient
The sentient entity that uses language in the written or spoken modality to convey a Message to another person.
He finds it hard to COMMUNICATE with people, not least his separated parents .

Medium [Medium]
The physical or abstract setting in which the Message is conveyed.
Opinions are usually COMMUNICATED over the telephone and are often given within 24 hours.

Message [Msg]
Semantic Type Message
Message A proposition or set of propositions that the Communicator wants the Addressee to believe or take for granted.
How do you COMMUNICATE to them that you really like them ?

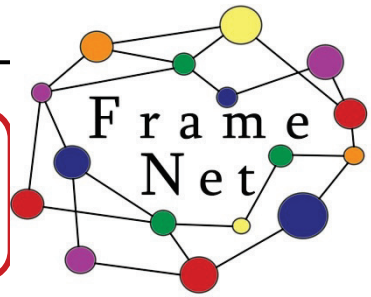
Topic [Top]
Topic is the entity that the proposition or propositions conveyed relate to, that they are about.
Had someone COMMUNICATED to the capital about the flagrant disregard of the religious law ?

Non-Core:

Addressee [Add]
Semantic Type Sentient
The Addressee receives a Message from the Communicator.
The company must be able to COMMUNICATE to potential customers the way in which its product would satisfy their needs, and provide competitive value.

and others …

Communication



relations between the Communication SF and other SFs

Frame-frame Relations:

Inherits From:
Is Inherited By: Communication_manner, Communication_noise, Communication_response, Gesture, Reassuring, Statement, Summarizing
Subframe of:
Has Subframes:
Precedes:
Is Preceded by:
Uses: Information, Topic
Is Used By: Attempt_suasion, Bail_setting, Candidness, Claim_ownership, Commitment, Communication_means, Contacting, Convey_importance, Deny_permission, Discussion, Encoding, Entering_of_plea, Expressing_publicly, Grant_permission, Hear, Justifying, Name_conferral, Notification_of_charges, Prevarication, Questioning, Reasoning, Reporting, Request, Ruling_legally, Sentencing, Speak_on_topic, Suasion, Successfully_communicate_message, Text_creation, Verdict, Volubility
Perspective on:
Is perspectivized in: Mention
Is Causative of:
See Also:

the list of lexical units that evoke the particular SF

Lexical Units

codeword.n, communicate.v, communication_((act)).n, communication_((entity)).n, contact.n, convey.v, indicate.v, password.n, signal.v, speech.n

lexical unit (LU) … **a pairing of a word with a sense**;
roughly, a single word in a single meaning (Cruse 1986),

From Case Theory to FrameNet

example: convey … 3 lexical units

Search

FrameNet Data Search for convey

Frame search results: Closest match is convey

Convey_importance

Lexical unit search results: Closest match is convey

Lexical Unit	Frame	LU Status	Lexical Entry Report	Annotation Report
convey.v	Bringing	Finished_Initial	LE	Anno
convey.v	Successfully_communicate_message	Finished_Initial	LE	Anno
convey.v	Communication	Created	LE	

communicate.v

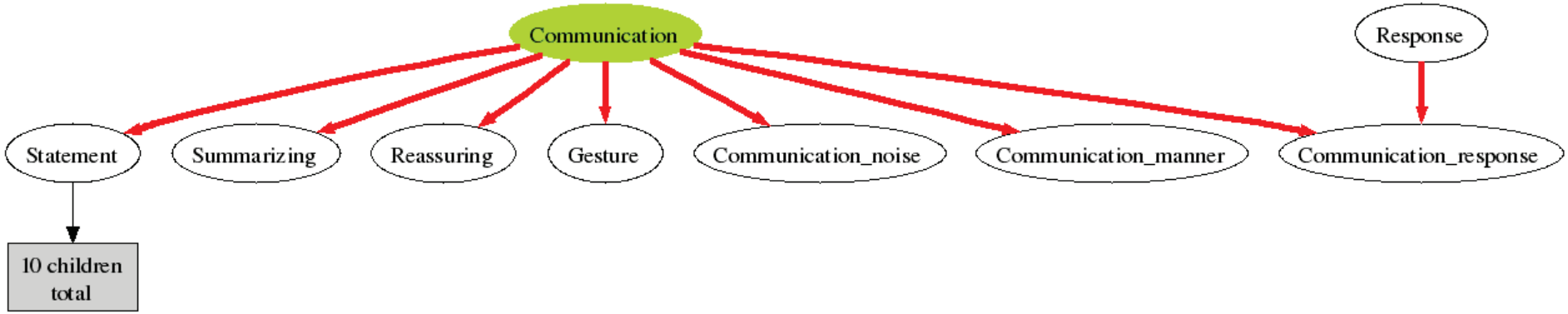
Frame Elements	Core Type
Addressee	Core
Amount_of_information	Peripheral
Communicator	Core
Depictive	Extra-Thematic
Duration	Extra-Thematic
Manner	Peripheral
Means	Peripheral
Medium	Peripheral
Message	Core
Time	Peripheral
Topic	Core

annotated examples from BNC

- T-Wwith-(1)
  1. I have been COMMUNICATING with the Minister since 1988 on that problem .
  2. Nerves that COMMUNICATE with the use of acetylcholine are said to be cholinergic and are found in the peripheral and central nervous systems .INI INI
  3. With the facility of computers it is possible , as shown by other unions as well as forward-looking branches of our own , to COMMUNICATE with members on a personal basis .CNI INI
  4. Let him COMMUNICATE with her by letter , if he must : let solicitors arrange money matters .INI
  5. By 1928 when the vote was granted to women over 21 , the two sisters had ceased to COMMUNICATE with each other -- their ideas and lifestyles were now poles apart .INI
  6. EDI has become extremely important in international freight enabling shippers , forwarders , carriers and HM Customs and Excise to COMMUNICATE with each other using a set of standard messages .INI
- T-NPinformation,message,idea-(1)
  1. Teletext uses a broadcast signal to COMMUNICATE information to specially adapted television sets , while Prestel uses a telephone link to allow access to information .
  2. Central to the overall strategy is the ability to COMMUNICATE information about individual patients and their care throughout the NHS .CNI about individual patients and their care

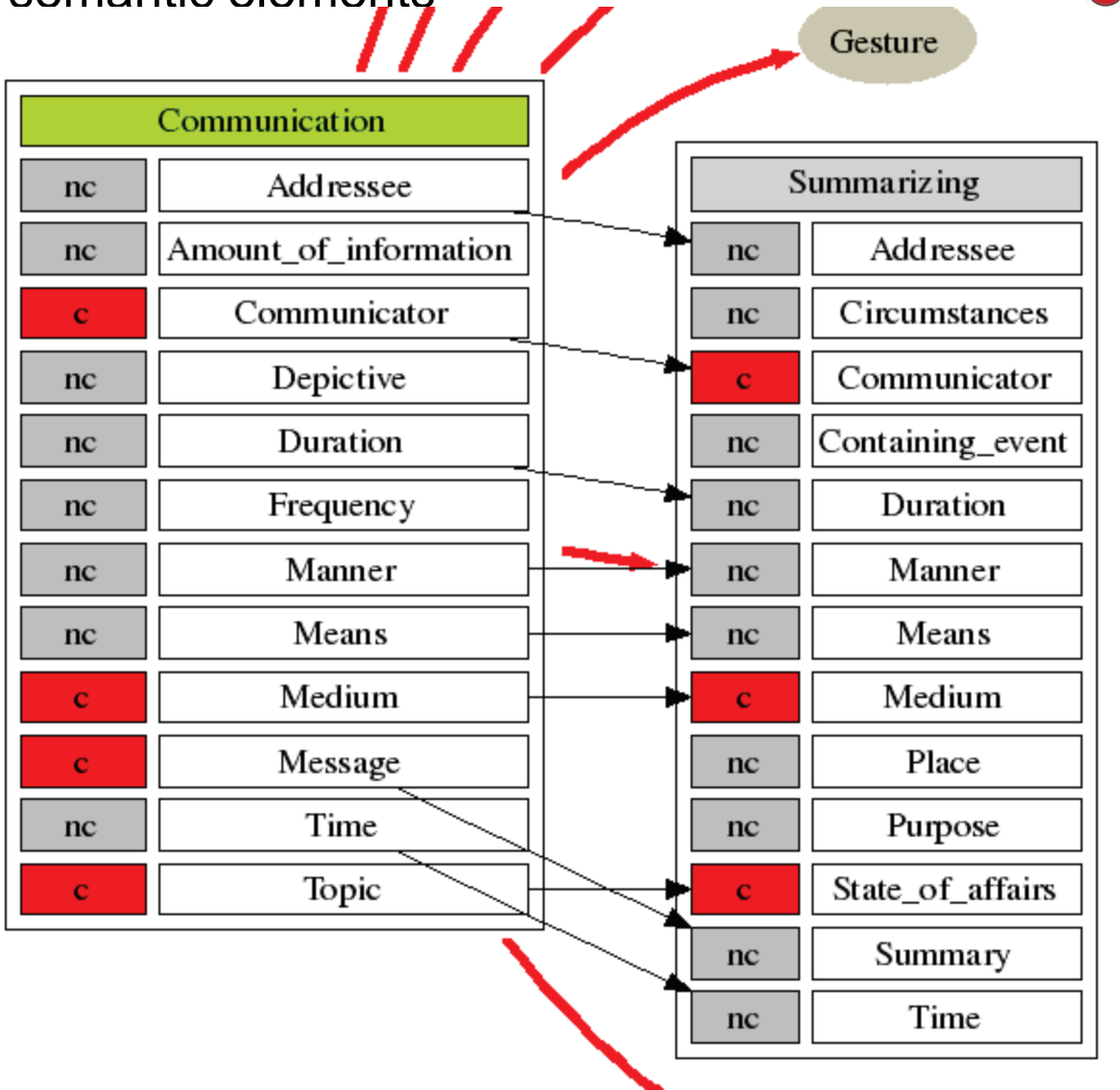
Communication

relation of inheritance: semantic frames

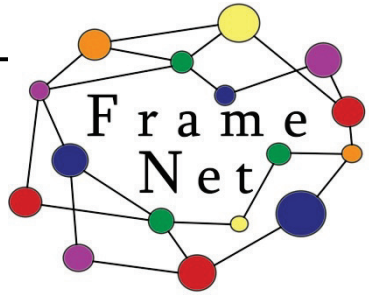


Communication

relation of inheritance: semantic elements



From Case Theory to FrameNet



- Principals: **Charles J. Fillmore, Collin F. Baker**
- Senior Linguist: Michael Ellsworth
- System Analyst: Jisup Hong
- FrameSQL Designer: Hiroaki Sato (Senshu University)
- Lexicographic Consultant: Sue Atkins
- Lexicographers, Annotators and Programmers:
  John Keesling, Albert Kong, Russell Lee-Goldman
- number of other participants and affiliates

(Fillmore, Baker et al., 1998; Ruppenhofer et al., 2010)

PropBank vs. FrameNet

FRAMENET ANNOTATION:

[Buyer Chuck] *bought* [Goods a car] [Seller from Jerry] [Payment for $1000].

[Seller Jerry] *sold* [Goods a car] [Buyer to Chuck] [Payment for $1000].

PROPBANK ANNOTATION:

[Arg0 Chuck] *bought* [Arg1 a car] [Arg2 from Jerry] [Arg3 for $1000].

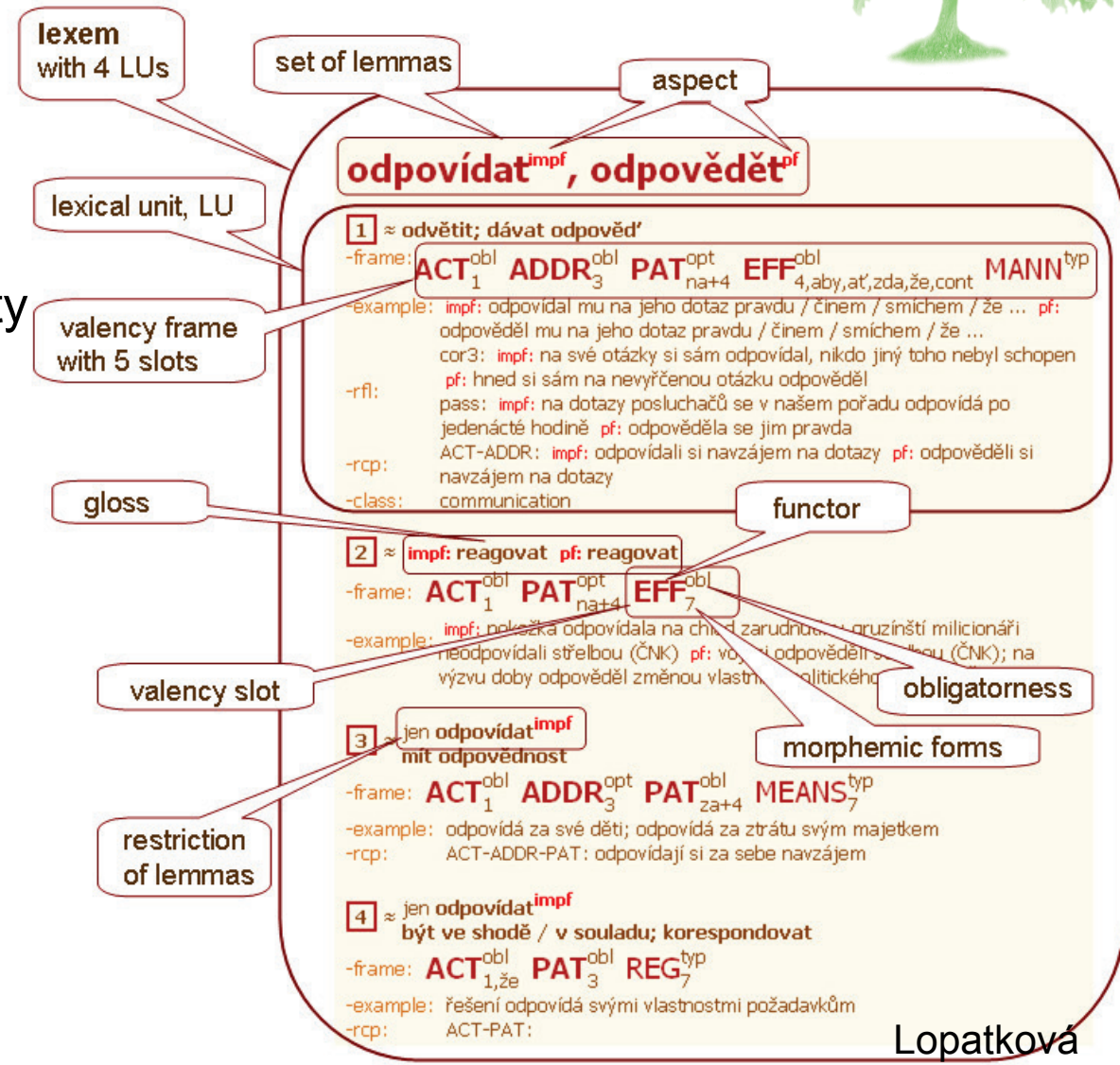
[Arg0 Jerry] *sold* [Arg1 a car] [Arg2 to Chuck] [Arg3 for $1000].

References

- Fillmore, Baker et al. (1998) The Berkeley FrameNet project. In *Proceedings of the COLING-ACL*, Montreal, Canada.
- Ruppenhofer et al. (2010) *FrameNet II: Extended Theory and Practice.* http://framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=126
- Cruse (1986) *Lexical Semantics*. Cambridge University Press, Cambridge.
- Palmer et al. (2005) The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics* 31, 1, p. 71-106.
- Kipper-Schuler, K. (2005) *VerbNet: A broad-coverage, comprehensive verb lexicon*. PhD thesis, Computer and Information Science Dept., University of Pennsylvania, Philadelphia
- Dorr, B. J. et al. (2001) *LCS Verb Database, Online Software Database of Lexical Conceptual Structures and Documentation.* Technical report, University of Maryland, 2001.
- Hanks, P., Pustejovsky, J. (2005) A Pattern Dictionary for Natural Language Processing, In *Revue française de linguistique appliquée* 10 (2)
- Šindlerová, J., Toman, J., Cinková, S., Semecký, J. (2007) *EngVallex 1.0.* UFAL MFF UK
- Lopatková et al. (2008) *Valenční slovník českých sloves*. Praha, Karolinum
- Hlaváčková, D., Horák, A. (2006) VerbaLex - New Comprehensive Lexicon of Verb Valencies for Czech. In *Computer Treatment of Slavic and East European Languages*. SNK, Bratislava p. 107-115
- Skoumalová, H. (2001) *Czech syntactic lexicon*. Ph.D. dissertation, FF UK, Prague
- Svozilová et al. (1997) *Slovesa pro praxi*. Praha, Academia
- Svozilová et al. (2005) *Slovník slovesných, substantivních a adjektivních vazeb a spojení*

Valency in Other Lexical Resources: Czech

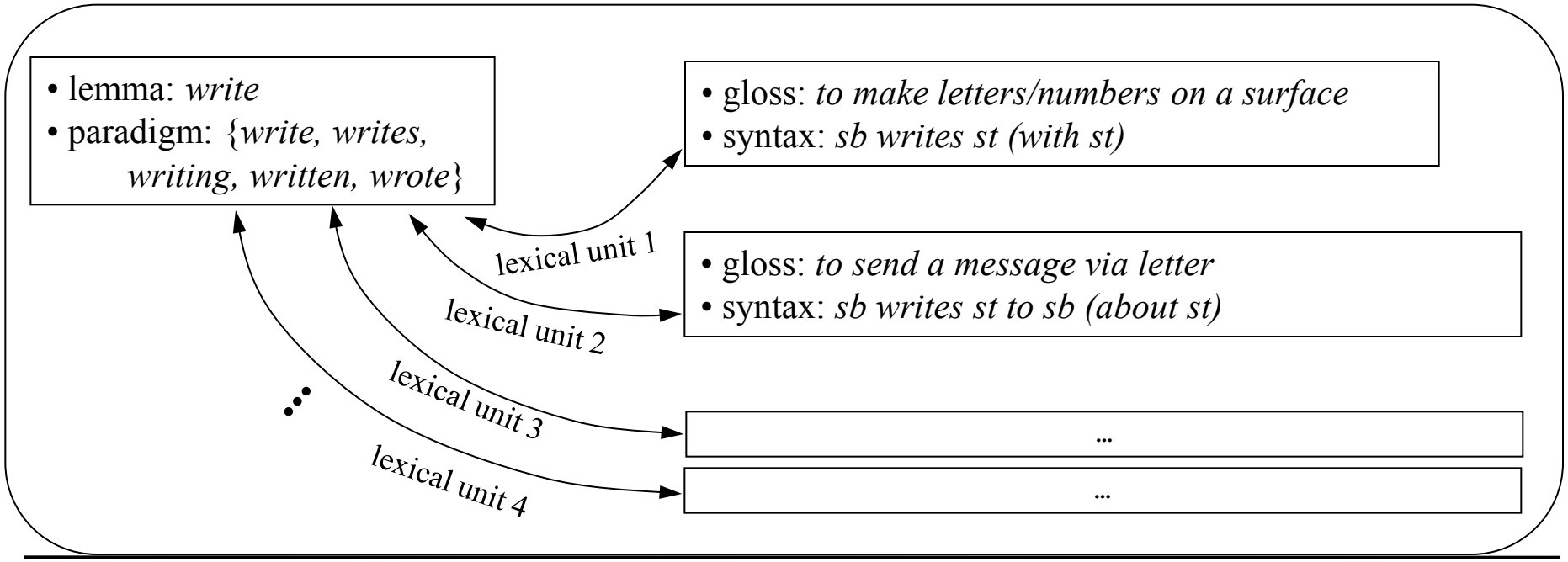
- ***VALLEX***
- most frequent Czech verbs
- valency frames ~ meaning
  - aspectual counterparts
  - diatheses
  - reflexivity and reciprocity
  - semantic classes



http://ufal.mff.cuni.cz/vallex/

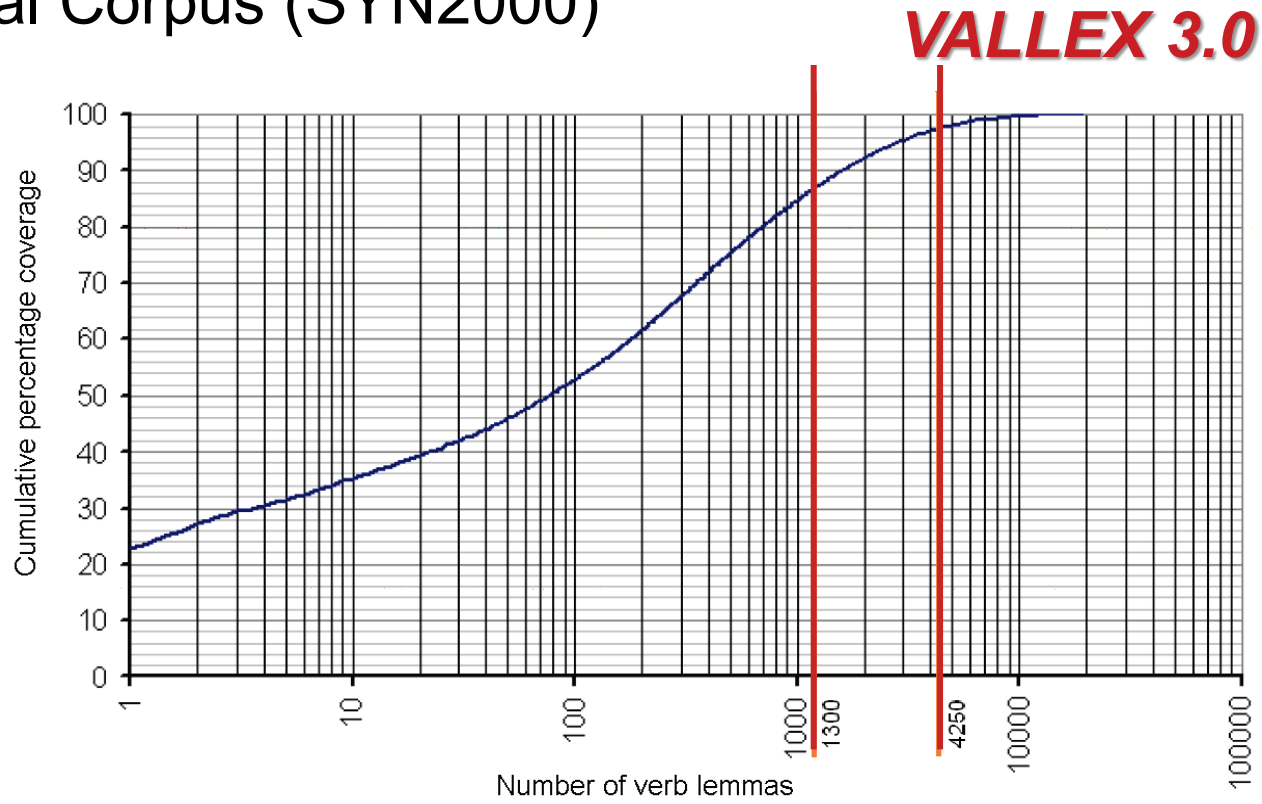
Valency in Other Lexical Resources: Czech

- ***VALLEX 3.0***
  - the lexicon structure
  - lexeme as an abstract unit consisting of
    - a 'formal' part … a list of word forms represented by (one or more) lemmas
    - a 'functional' part … a list of lexical units
  - a single lexeme for aspectual counterparts



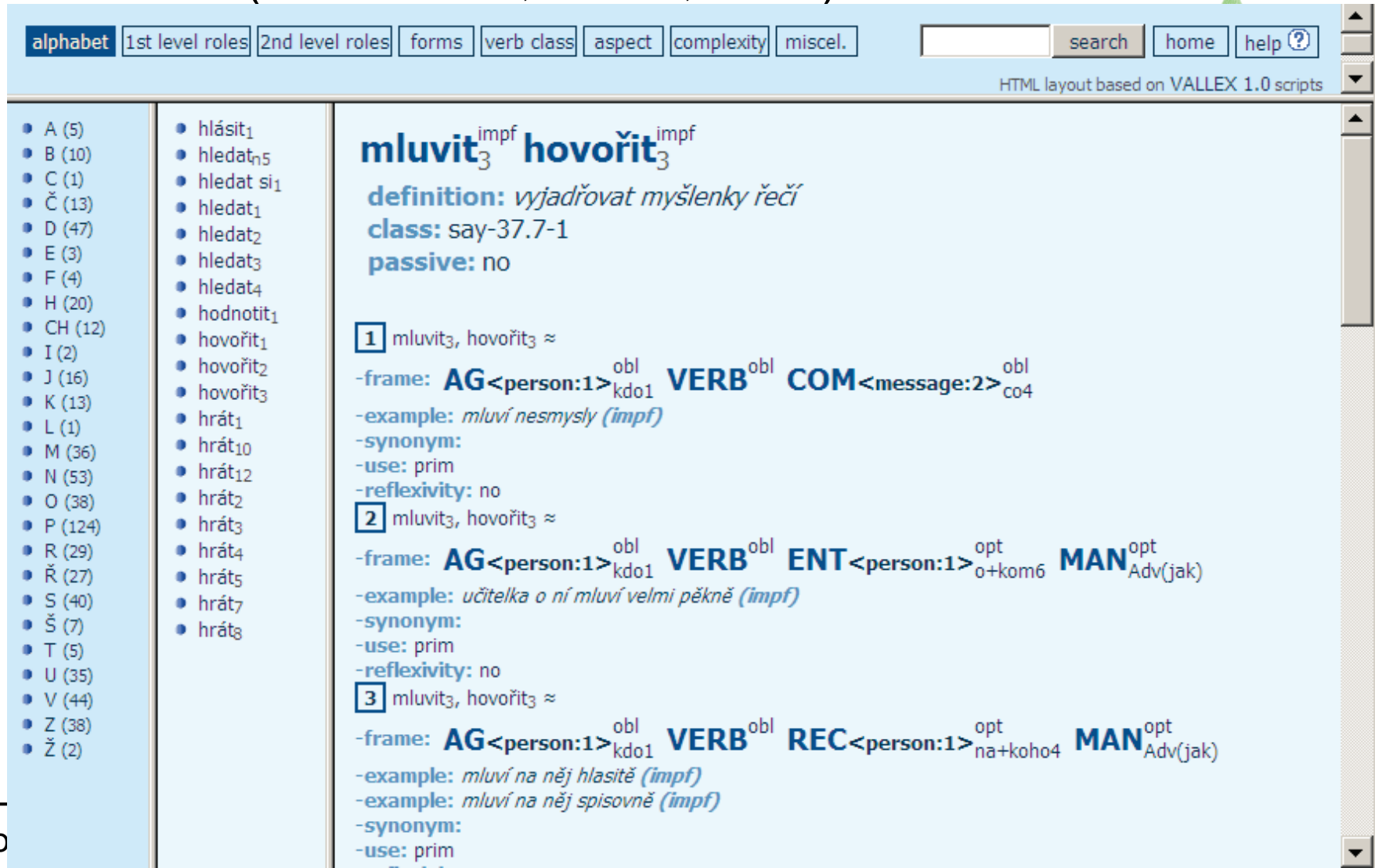
Valency in Other Lexical Resources: Czech

- ***VALLEX 3.0***
  - 2 722 lexemes, 6 711 lexical units
  - 4 586 verbs, 10 821 meanings
- coverage 98% of verbal occurrences in the Czech National Corpus (SYN2000)



Valency in Other Lexical Resources: Czech

- ***VerbaLex*** … (Hlaváčková, Horák, 2006)



PDT: o

Valency in Other Lexical Resourses

- Czech Syntactic Lexicon (Skoumalová, 2001)
- Slovesa pro praxi (Svozilová et al., 1997)
- Slovník slovesných, substantivních a adjektivních vazeb a spojení (Svozilová et al., 2005)

- Pattern Dictionary of English Verbs (PDEV) (Hanks, Pustejovsky)
  - based on Corpus Pattern Analysis (CPE)
  - prototypes and patterns
  - exploitation
  - ***semantic types*** and ***semantic roles***

cognitive concepts such as Human, Institution, Animal, Event, etc

intrinsic property of nouns normally found in that argument slot,
e.g. judge, criminal, punishment for to sentence

References

- Fillmore, Baker et al. (1998) The Berkeley FrameNet project. In *Proceedings of the COLING-ACL*, Montreal, Canada.
- Ruppenhofer et al. (2010) *FrameNet II: Extended Theory and Practice.* http://framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=126
- Cruse (1986) *Lexical Semantics*. Cambridge University Press, Cambridge.
- Palmer et al. (2005) The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics* 31, 1, p. 71-106.
- Kipper-Schuler, K. (2005) *VerbNet: A broad-coverage, comprehensive verb lexicon*. PhD thesis, Computer and Information Science Dept., University of Pennsylvania, Philadelphia
- Dorr, B. J. et al. (2001) *LCS Verb Database, Online Software Database of Lexical Conceptual Structures and Documentation.* Technical report, University of Maryland, 2001.
- Hanks, P., Pustejovsky, J. (2005) A Pattern Dictionary for Natural Language Processing, In *Revue française de linguistique appliquée* 10 (2)
- Šindlerová, J., Toman, J., Cinková, S., Semecký, J. (2007) *EngVallex 1.0.* UFAL MFF UK
- Lopatková et al. (2008) *Valenční slovník českých sloves*. Praha, Karolinum
- Hlaváčková, D., Horák, A. (2006) VerbaLex - New Comprehensive Lexicon of Verb Valencies for Czech. In *Computer Treatment of Slavic and East European Languages*. SNK, Bratislava p. 107-115
- Skoumalová, H. (2001) *Czech syntactic lexicon*. Ph.D. dissertation, FF UK, Prague
- Svozilová et al. (1997) *Slovesa pro praxi*. Praha, Academia
- Svozilová et al. (2005) *Slovník slovesných, substantivních a adjektivních vazeb a spojení*