

The use of semantic prototypes in semantic role annotation

Eckhard Bick

University of Southern Denmark
VISL Project, ISK

- motivation and methods: Background
- semantic classification: 1. semantic prototypes and atomic features vs. 2. semantic roles (form vs. function)
- inducing (2) from (1) using syntactic context:
a semantic role annotation grammar
- advantages and problems:
Dependency trees and verb frames
- an applicative example from the realm of lexicography:
The DeepDict collocation dictionary

Higher-level Corpus Annotation

- Why annotate?
 - structurally and semantically annotated data provide a better basis for linguistic research than raw text data
 - lexicography: grammatical patterns rather than simple collocations (verb complementation, modifiers etc.)
 - training/learning of automatic NLP systems
- What to annotate?
 - Layered annotation: PoS (word class) & morphology -> syntactic function & dependency -> semantic relations
- How to annotate?
 - In our approach. with rule-based CG parsers, i.e. all information encoded as tags on tokens or as treebanks
 - progressively, deducing syntactic structures from morphology & PoS, and semantic structures from lexical types

a modular parsing paradigm

Constraint Grammar (CG)

- Descriptive modularity:
successive levels of linguistic annotation:
 - "classical": morphological, syntactic function, chunking (Karlsson et al. 1995), implement for many languages
 - tree structure: (dependency, constituent)
 - as part of the formalism, (a) FDG/Connexor, (b) VISL-CG3
 - using added PSG or dependency grammars (Bick 2003, 2005)
- Methodological modularity:
adding, removing and substituting information
 - lexical information from either dictionaries, morpheme-analysis or grammar-internal word sets
 - rule sets as consecutive batches
 - mapping rules for the addition of module-specific information
 - integration of external structural rules

Semantic tags

- **secondary**: semantic tags employed to aid disambiguation and syntactic annotation (traditional CG):
`<vcog>`, `<speak>`, `<Hprof>`, `<aloc>`, `<jnat>`
- **primary**: semantic tags as the object of disambiguation
- **semantic CG modules**:
 - Named Entity classification
 - Danish: Nomen Nescio (Bick 2003)
 - Portuguese: HAREM shared task (Bick 2006-1)
 - semantic prototype tagging for treebanks
 - Floresta Sintáctica (Bick 2006-2), Arboretum
 - semantic tag-based applications
 - pt-da, da-en and inter-scandinavian Machine translation (GramTrans)
 - QA, library IE, sentiment surveys, referent identification (anaphora)

Beyond prototypes

integrating structure and lexicon

- (a) "lexical perspective": contextual selection of
 - a (noun) sense [WordNet style] or
 - semantic prototype [SIMPLE style]
- (b) "structural perspective":
the semantics of argument structure (verb - nominal)

thematic/semantic roles

- Fillmore 1968
- Jackendoff 1972
- Foley & van Valin 1984
- Dowty 1987

Semantic argument slots

- the semantics of a noun can be seen as a "compromise" between its lexical potential or "form" (e.g. prototypes) and the projection of a syntactic-semantic argument slot by the governing verb ("function")
- e.g. <civ> (country, town) prototype
 - (a) **location, origin, destination** slots
(adverbial argument of movement verbs)
 - (b) **agent** or **patient** slots
(subject of cognitive or agentive verbs)
- Rather than hypothesize different senses or lexical types for these cases, a **role annotation level** can be introduced as a bridge between syntax and semantics

DanGram Semantic prototypes

- SIMPLE and DanNet-compatible, ca. 160 types, thus a possibility for cross-mapping or even integration
- Ontology with umbrella classes and subcategories, e.g.
 - <H> : <Hprof>, <HH>, <Hnat>, <Htitle>, <Hfam> ...
 - <L> : <Ltop>, <Lh>, <Lwater>, <Labs>, <Lsurf>
 - <sem> : <sem-r>, <sem-l>, <sem-c>, <sem-s> ...
- allows composite “ambiguous” tags:
 - <civ> (<HH> + <L>), <media> (<HH> + <sem>)
- prototypes expressed as bundles of atomic semantic features, e.g. <V> (vehicle) = +concrete, -living, -human, +movable, +moving, -location, -time ...
- cross-language: same prototypes used for 7 other languages besides Danish

E = entities (\pm CONCRETE)

V = \pm VERBAL

C = \pm CONTROL

P = \pm PERFECTIVE

I = \pm MOVING

S = \pm MEASURING

J = \pm MOVABLE

D = \pm PARTITIVE

A = \pm ANIMATE (living)

X = \pm HUMAN-EXPRESSION (allowing human modifier-ADJ)

H = \pm HUMAN ENTITY

F = feature (\pm ADJECTIVAL)

M = \pm MASS

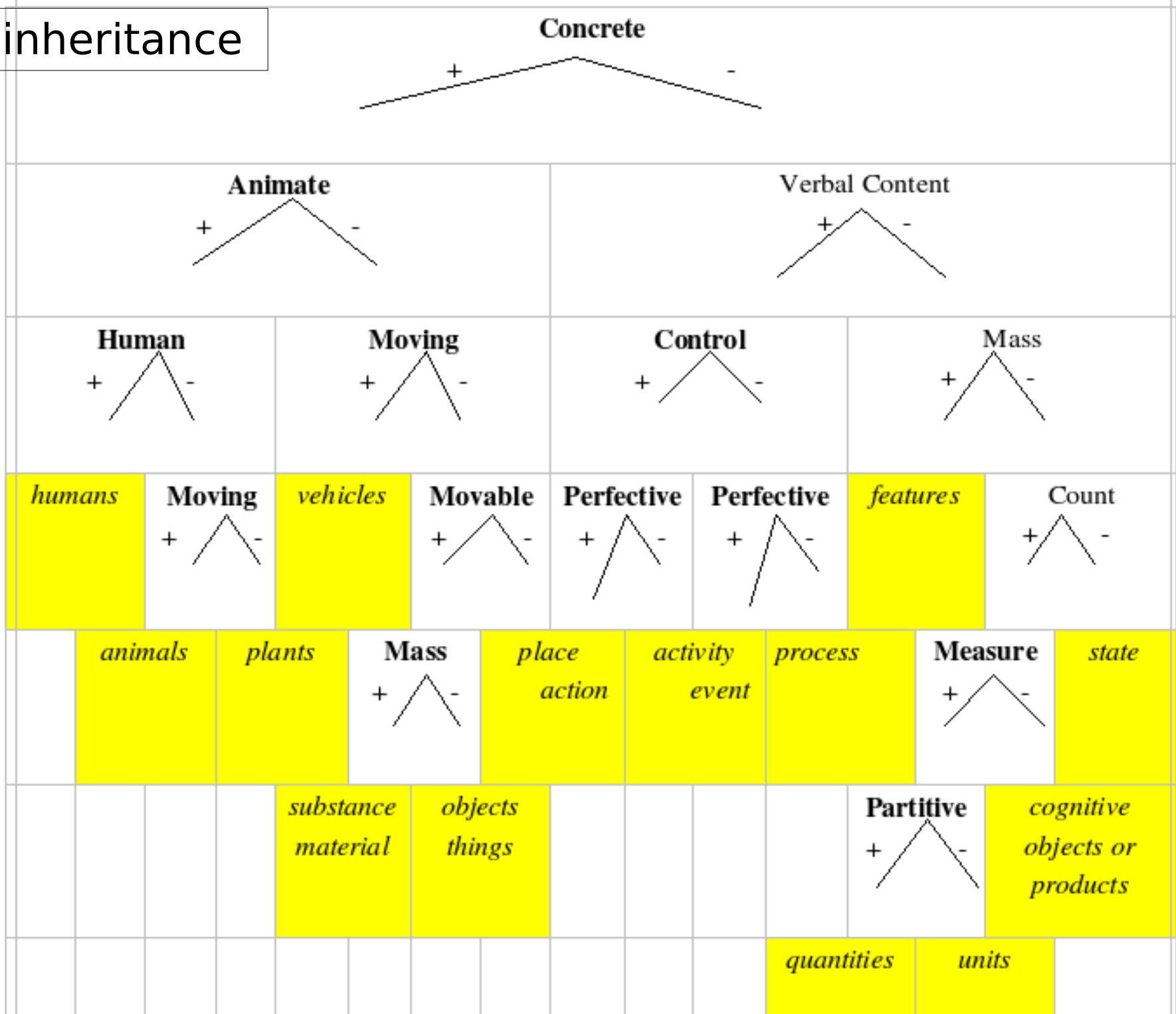
L = \pm LOCATION

N = number (\pm COUNTABLE)

T = \pm TEMPORAL

E	C	I	J	A	H	M	N	V	P	S	D	X	F	L	T	
+		'+	+	'+	'+	.	'+			.		+	.	.	.	H, <u>Hprof</u> , <u>Hnat</u> , <u>Hmyth</u> , <u>Hfam</u> , <u>Htit</u> , i, <u>Hbio</u> , <u>Hsick</u> , <u>Hattr</u> , *hum
+		'+	+	'+	'+	.	.			.		+	.	.	.	HH, <u>Hhparty</u> , *party, *media
+			.	.	'+	.	.			.		+	.	'+	.	inst, *inst
+			.	.	'+	.	'+			.		+	.	'+	.	<u>Lciv</u> , *civ
+		'+	+	'+	.	.	'+			A, <u>Azo</u> , <u>Aorn</u> , <u>Aent</u> , <u>Aich</u> , <u>Amyth</u> , <u>Acell</u> , <u>Adom</u>
+		'+	+	'+	AA (<u>AAzo</u> ..)
+			.	'+	.	.	'+			B, <u>Btree</u> , <u>Bbush</u> , <u>Bflo</u> , <u>Bveg</u>
+			.	'+	'+	.	BB
'+			'+			.		.	.	'+	.	<u>Ltop</u> , <u>Lwater</u> , <u>Lpath</u> , <u>Lh</u> , <u>Lstar</u> , build, part-build, bar, an, <u>anorg</u> , <u>anost</u> , <u>anzo</u> , <u>anorn</u> , <u>anbo</u> , sick-c, dir, *top
.			'+			.		.	.	'+	.	Labs, <u>Ltip</u> , <u>Lopening</u> , <u>Lsurf</u>
+		.	'+	.	.	.	'+			.		.	.	'+	.	furn, <u>Ltrap</u> , <u>Lcover</u> , con, pict
.			'+	.	<u>pos-an</u> , <u>pos-soc</u> , sit, <u>anfeat</u>
.			.	.	.	'+	'+	.	<u>wea</u> , <u>wea-wind</u> , <u>wea-rain</u>
+		.	'+	.	.	.	'+			.		.	.	'+	.	<u>anmov</u>
'+			'+	.	<u>Lsurf</u>
+		'+	+	.	.	.	'+			V, <u>Vground</u> , <u>Vwater</u> , <u>Vair</u>
+		'+	+	VV (ground)

feature inheritance



Concrete

Animate

Verbal Content

Human

Moving

Control

Mass

humans

Moving

vehicles

Movable

Perfective

Perfective

features

Count

animals

plants

Mass

place

activity

process

Measure

state



action

event



*substance
material*

*objects
things*

Partitive

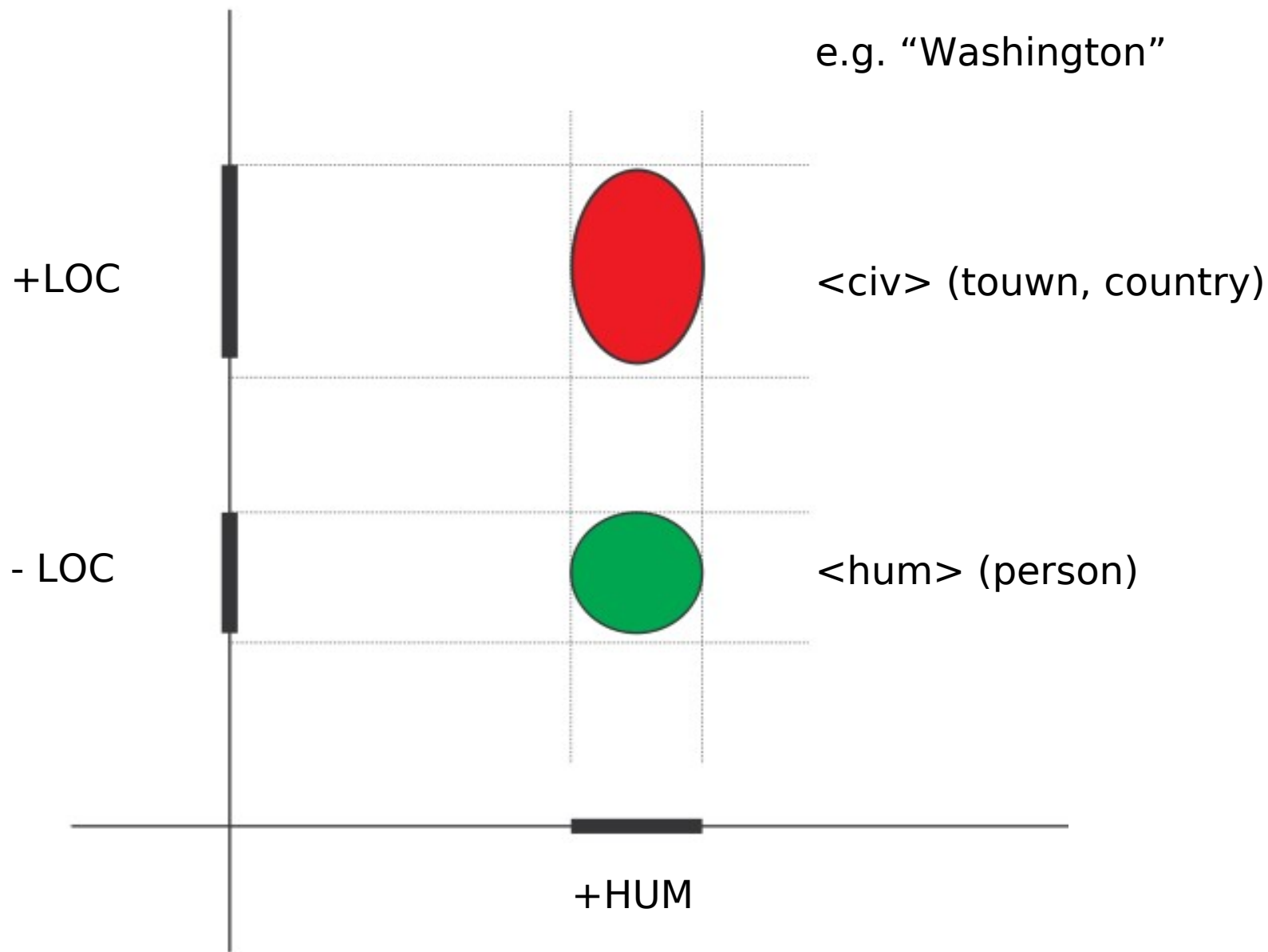
*cognitive
objects or
products*



quantities

units

Disambiguation of semantic prototype bubbles by dimensional downscaling (lower-dimension projections)



Semantic role annotation for Danish

- inspired by the Spanish **3LB-LEX** project (Taulé et al. 2005)
- allows, together with the syntactic function annotation (ARG structure) a mapping onto **PropBank** argument frames (Palmer et al. 2005)
- will allow the extraction of argument frames from treebanks (i.e. the 400kW **Arboretum**)
- **manual vs. automatic**: due to the quality of the syntactic parser and the existence of the prototype lexicon, a boot-strapping is envisioned, where
 - *syntactic* valency is exploited in conjunction with the prototype lexicon (ontology) to create semantic role annotation,
 - which in turn provides "*semantic* valency frames"
 - which then is used to improve the semantic role annotation

The semantic role inventory

<i>"Nominal" roles</i>	<i>definition</i>	<i>example</i>
§AG	agent	X spiser Y
§PAT	patient	Y spiser X, X gik i stykker, X blev +PCP2
§BEN	benefactive	give Y til X
§EXP	experiencer	X frygter Y, overraske X
§TH	theme	iagttage X, X er syg, X ligger der, X har Y
§RES	result	Y bygger X
§ROLE	role	Y arbejder som <i>guide</i>
§COM	co-argument, comitative	Y danser med X
§ATR	static attribute	X er syg, en ring <i>af guld</i>
§ATR-RES	resulting attribute	gøre ngn. <i>nervøs</i>
§POS	possessor	Y tilhører X, <i>Christians</i> bil
§CONT	content	en flaske <i>vin</i>
§PART	part	Y består af X, X danner en helhed
§ID	identity	Byen <i>Itatiaia</i> , det svenske selskab <i>Volvo</i>
§VOC	vocative	rolig, <i>Peter!</i>

<i>"Adverbial"roles</i>	<i>definition</i>	<i>example</i>
§LOC	location	bo i X, her, hjemme
§ORI	origin, source	flygte fra X, kød fra Argentina
§DES	destination	sende til X, en flyforbindelse til X
§PATH	path	hen over huset, igennem hullet, langs floden
§EXT	extension, amount	marchere 7 kilometer, veje 70kg
§LOC-TMP	temporal location	sidste år, i morgen aften, når vi mødes igen
§ORI-TMP	temporal origin	siden januar
§DES-TMP	temporal destination	indtil torsdag
§EXT-TMP	temporal extension	3 uger lang, 4 år
§FREQ	frequency	sommetider, 14 gange
§CAU	cause	pga. X, fordi han ikke kunne komme
§COMP	comparation	bedre end nogensinde
§CONC	concession	på trods af X, selvom vi ikke har hørt noget
§COND	condition	i tilfælde af X, medmindre vi hører andet
§EFF	effect, consequence	med risiko for, der var så mange at ...
§FIN	purpose, intention	arbejde for optagelsen af Tyrkiet
§INS	instrument	vha. X, skære brødet med, betale med
§MNR	manner	på denne måde, i takt med X, hvordan ...
§COM-ADV	accompanier (ArgM)	ud over Anne, med ngt i hånden

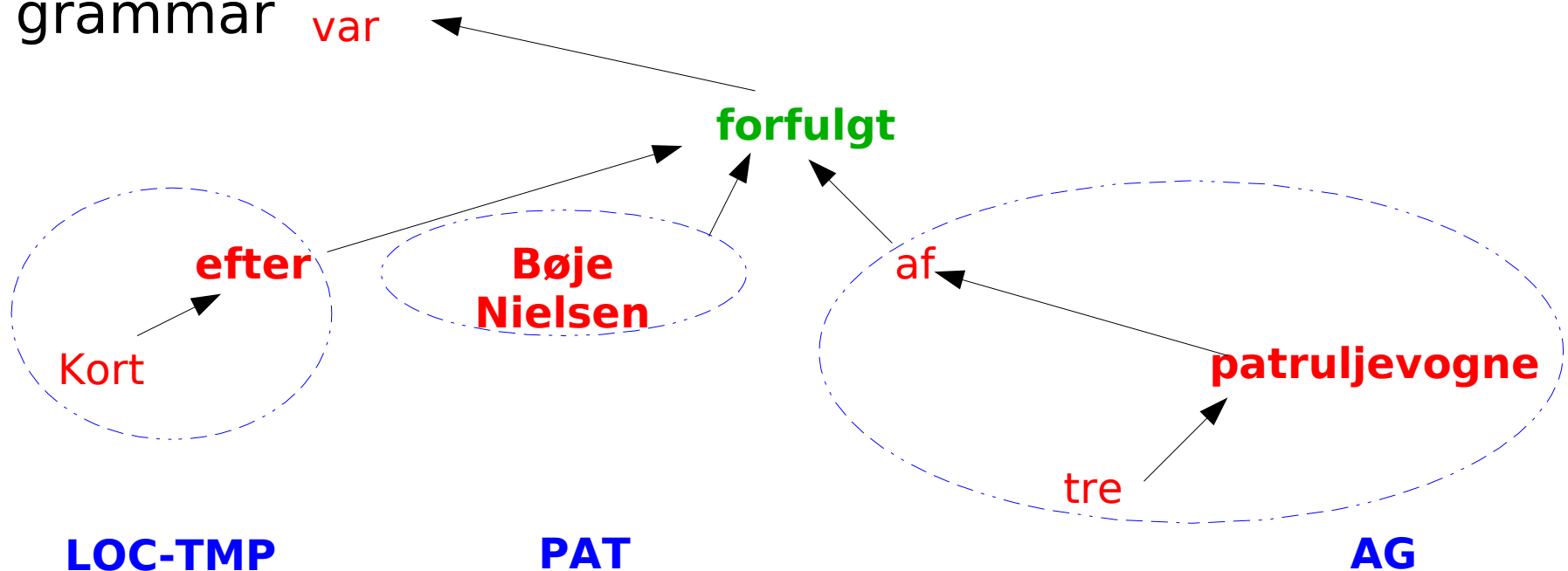
<i>"Syntactic roles"</i>	<i>definition</i>	<i>example</i>
§META	meta adverbial	<i>ifølge X, måske, åbenbart</i>
§FOC	focalizer	<i>only, også, endda</i>
§ADV	dummy adverbial	if no other adverbial categories apply
§EV	event, act, process	<i>tillade/påbegynde X, ... X starter/slutter</i>
§PRED	(top) predicator	main verb in main clause
§DENOM	denomination	lists, headlines
§INC	verb-incorporated	<i>finde sted</i> (not fully implemented)

A new CG formalism

- new cg design: CG3 (<http://beta.visl.sdu.dk/cg3.html>) [Didriksen & Bick]
 - regular expressions, probabilistic/"mathematical" tags, multi-window-references, dynamic meta-variables, templates, flexible rule section administration
- to link thematic roles to dependency slots:
 - dependency "positions" in rule and link contexts, supplementing the ordinary relative and global position indicators
 - p - parent (mother nodes)
 - c - child (daughter nodes)
 - s - sibling (sister nodes)
- Needs dependency trees as input, created independently or with preceding CG modules

Dependency trees

- **CG-attachment grammar:** rules for close/long attachment, coordination, clause boundaries and ellipsis
- **Dependency grammar:** function attachment rules with built-in uniqueness and non-circularity conditions (ca. 100 rules), *could* be CG3, but *is* a Perl-compiled grammar



Source format

Kort (<i>soon</i>)	[kort] ADV @>A #1->2
efter (<i>after</i>)	[efter] <atemp> ADV @ADVL> #2->6 §LOC-TMP
var (<i>was</i>)	[være] <aux> V IMPF AKT @FS-STA #3->0
Bøje=Nielsen	[Bøje=Nielsen] <hum> PROP NOM @<SUBJ #4->3 §PAT
således (<i>thus</i>)	[således] ADV @<ADVL #5->6
forfulgt (<i>chased</i>)	[forfølge] <mv> V PCP2 PAS @ICL-AUX< #6->3
af (<i>by</i>)	[af] PRP <pass> @<ADVL #7->6
tre (<i>three</i>)	[tre] <card> NUM P @>N #8->9
patruljevogne (<i>police cars</i>)	[patruljevogn] <Vground> N UTR P IDF NOM @P< #9->7 §AG
\$.	[\$.] PU @PU #10->0

(Example from the Arboretum Treebank part of Korpus90/2000)

Inferring **semantic roles** from **semantic prototype sets** using syntactic function (@) and dependency (p, c and s)

implicit inference of semantics:

syntactic function (e.g. @SUBJ) and valency potential (e.g. ditransitive <vdt>) are *not* semantic by themselves, but help restrict the range of possible argument roles (e.g. §BEN for @DAT)

- *subjects of ergatives*
MAP (§PAT) TARGET @SUBJ (p <ve> LINK NOT c @ACC) ;
- *the give sb-DAT s.th.-ACC frame*
MAP (§TH) TARGET @ACC (s @DAT) ;

Inferring **semantic roles** from **semantic prototype sets** using syntactic function (@) and dependency (p, c and s)

explicit use of lexical semantics: semantic prototypes:

- <Hprof> (human professional),
- <Hideo> (ideology-follower),
- <Hnat> (nationality) ...

restrict the role range by themselves, but are ultimately still dependent on verb argument frames

(a) "*Genitivus objectivus/subjectivus*"

- MAP (§PAT) TARGET @P< (p PRP-AF + @N< LINK p N-VERBAL) ;
ødelæggelsen af byen
- MAP (§AG) TARGET GEN @>N (p N-ACT) ;
Springers publicering af .
- MAP (§PAT) TARGET GEN @>N (p N-HAPPEN) ;
Økonomiens sammenbrud

- **Agent:** han blev forfulgt af tre patruljevogne (he was chased by three police cars)
MAP (§AG) TARGET @P< (p ("af" @ADVL) LINK p PAS) (0 N-HUM OR N-VEHICLE) ;
- **Possessor:** “malerens pensel” (the painter's brush)
MAP (§POS) TARGET @P< (0 N-HUM + GEN LINK 0 @>N) (p N-OBJECT) ;
- **Instrumental:** “destroy the piano with a hammer”
MAP (§INS) TARGET @P< (0 N-TOOL) (p PRP-MED + @ADVL) ;
- **Content:** “en flaske vin” (a bottle of wine)
MAP (§CONT) TARGET @N< (0 N-MASS OR (N P)) (p <con>) ;
- **Attribute:** “en statue af guld”
MAP (§ATR) TARGET @P< + N-MAT (p PRP-AF + @N<) ;
- **Location:** “live in a big house”
MAP (§LOC) TARGET @P< + N-LOC (p PRP-LOC LINK 0 @ADVL OR @N<);
- **Origin:** “send greetings from Athens”, “drive all the way from the border”
MAP (§ORI) TARGET @P< (0 N-LOC) (p PRP-FRA LINK 0 @<ADVL OR @<SA OR @<OA LINK p V-MOVE/TR) ;
- **Temporal extension:** “The session lasted 4 hours”
MAP (§EXT-TMP) TARGET @SA (0 N-DUR) ;

Integrating the lexical resources in the grammar: Set definitions for rule use

- N-HUM = <H> <Hprof> <Hideo> <Hnat> ...
- N-LOC = <L> <Ltop> <Lh> <Lwater> <Lsurf> ...
- N-TOOL = <tool> <tool-cut> <tool-shoot> ..
- N-OBJECT = <cc> <cc-stone> ... <food> <food-h> ...
- verb semantics, largely defined as lists in the grammar itself:
 - V-SPEAK, V-MOVE, V-MOVE-TR ...
 - with valency: VT-ACC-HUM, VT-ACC-NON-HUM, VT-ACC-ORG, VT-ACC-TITLE, V-SUBJ-HUM, VT-SUBJ-HUM
- adjective semantics (in part matching noun prototypes):
 - <jh>, <jnat>, <jcol> ...

Grammar profile

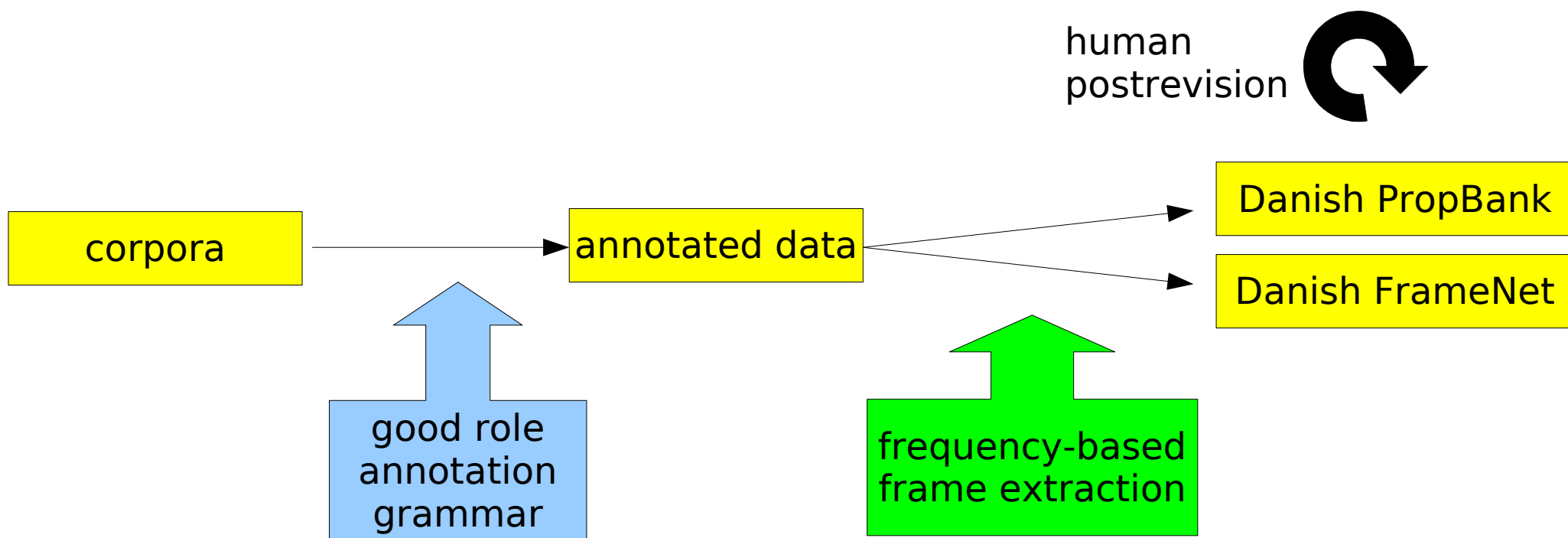
- current grammar size:
 - a few hundred general rules
 - a growing body of token-specific rules for irregular cases
- a backbone of unambiguous mapping rules, that constitute a clear majority of rules
- enhanced rule economy by using semantic tags and sets rather than individual lexemes - for both targets and contexts
 - nouns: ~160 semantic prototypes
 - verbs: \pm HUM information for subject arguments and some object cases, <move>, <speak>, <cog> ...
 - adjectives: domain classes

Advantages of the dependency-based method in semantic annotation

- in a traditional CG (without dependency), about 3 times as many rules would be needed:
 - left-pointing arguments vs. right-pointing arguments
 - special rules for embedded clauses (e.g. relative clauses after subjects)
- greater context precision of dependency-based rules
 - safer mapping rules
 - less need for ambiguous mapping followed by REMOVE/SELECT-disambiguation
- dependency rules are simpler and less error prone
 - no need for BARRIER contexts and "negative" sets to handle/delimit constituent boundaries

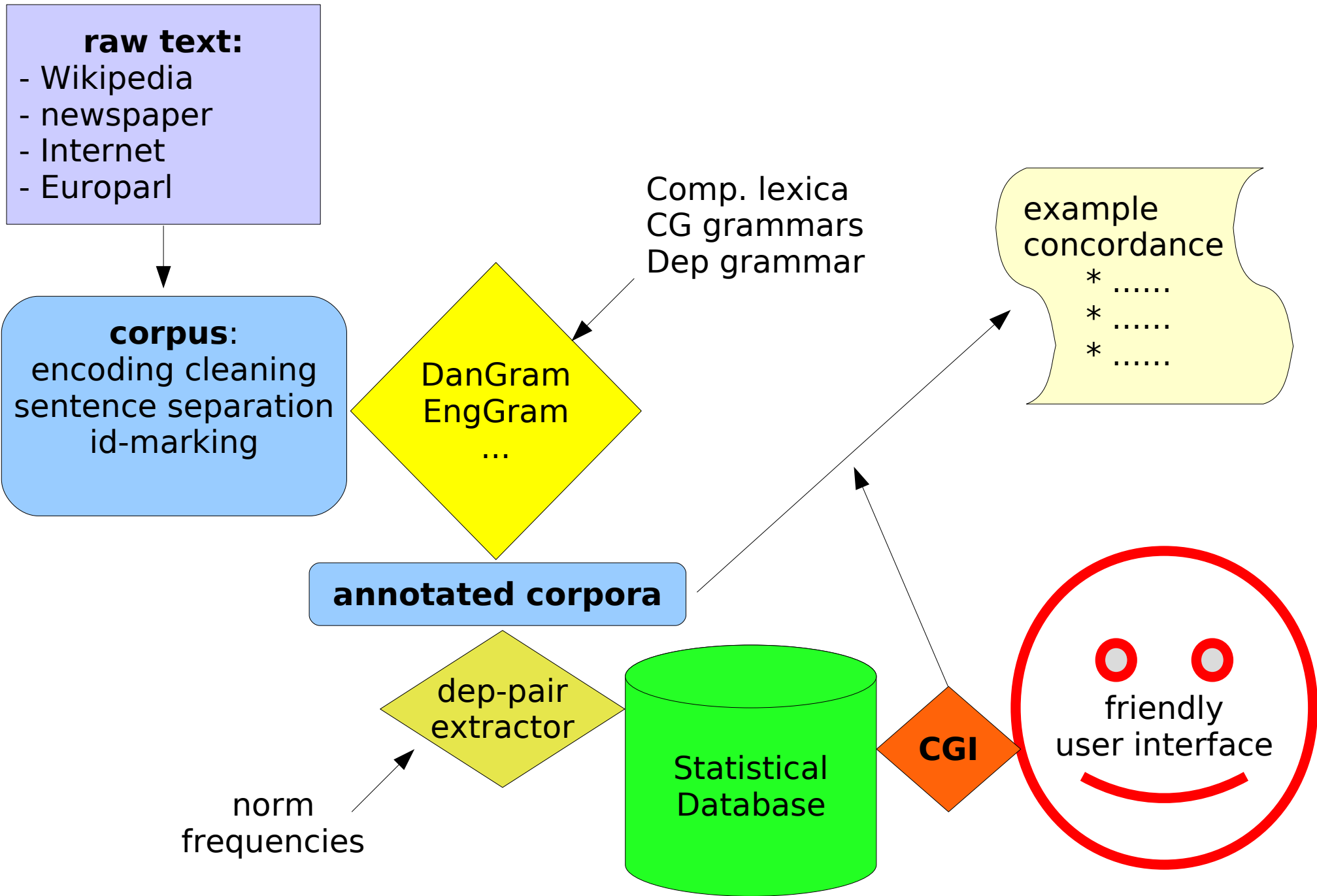
Problems: The verb frame bottle neck

- few semantic verb classes in the DanGram lexicon at present
- need to compensate with CG lexeme sets to support role allocation, e.g. experiencer ('see') or resultative verbs ('bygge'), limited by the fact that
 - sets focus at one argument at a time (not the whole frame!)
 - and do not cover the whole lexicon
- Solution: Boot-strapping idea for building a frame lexicon



A usage example: Using “deeply” annotated corpora for “live” lexicography: **DeepDict**

- 1) **annotate** a corpus with dependency, syntactic function and semantic classes
- 2) **extract dependency pairs** of words and their complements, e.g. V + @ACC, @>N + N
- 3) identify **statistically significant relations** (not n-grams but “dep-grams!”)
- 4) **normalisation** (lemma) and classification
- 5) build a graphical **interface**
 - * semantic prototypes will allow generalisations and cross-language comparisons
 - * uses: advanced learner's dictionaries, production dictionaries, language and linguistics teaching ...



DeepDict user interface

Quick Reference

An entry consists of the following elements, taking "8.37:6 comic" as example:

- **8.37**: The co-occurrence strength between the lookup word and a given relation (relative frequency).
- **6**: Dual logarithmic value of absolute frequency. Scale is from 1 to 9.
- **comic**: The co-occurring word. In bold face if the absolute frequency value is 4 or higher.

Red numbers can be clicked to show examples in concordance form, if available.

For more information, please see the [DeepDict Reference page](#).

Change Lookup Parameters

Word to look up: <input type="text" value="vote"/>	Lookup language: <input type="radio"/> Danish <input checked="" type="radio"/> English <input type="radio"/> French <input type="radio"/> German <input type="radio"/> Portuguese <input type="radio"/> Spanish	Lexical frequency threshold: <input type="radio"/> High <input checked="" type="radio"/> Medium <input type="radio"/> Low <input type="radio"/> None Minimum occurrence: <input type="text" value="2"/> Minimum relative frequency: <input type="text" value="0.0"/> Show top: <input type="text" value="25"/>
Word class: <input type="radio"/> Noun <input checked="" type="radio"/> Verb <input type="radio"/> Adverb <input type="radio"/> Adjective <input type="button" value="Look up via DeepDict"/>		

DeepDict: implicit semantics from lexical relations

voice (noun)

countable

Premodifiers: 6.73:7 loud · 6.57:7 NUM · 6.41:6 distinctive · 5.05:7 deep · 6.64:5 soprano · 7.46:4 gravelly · 7.44:4 husky · 4.34:7 single · 5.21:6 inner · 4.2:7 own · 6.59:4 baritone · 5.58:5 passive · 6.52:4 hoarse · 4.46:6 soft · 5.46:5 authoritative · 4.32:6 quiet · 4.28:6 human · 6.28:4 squeaky · 6.16:4 narrative · 5.98:4 gruff	PP postmodifiers: 8.72:8 rel-INDP 1.35:3 interr-INDP 2.58:5 of character 2.43:5 of reason 2.25:5 of god 3.08:4 from behind 1.75:5 of america 2.7:4 of conscience 2.43:3 of dissent	Modifier of: 6.04:7 actor · 5.94:4 telephony · 3.58:5 actress · 4.91:3 coil · 2.04:5 communication · 2.88:4 talent · 2.88:4 recorder · 3.52:3 choir · 2.19:4 transmission · 1.02:5 vote · 1.76:4 channel · 2.7:3 characterization · 3.59:2 synthesizer · 3.47:2 inflection · 3.41:2 synthesis · 0.31:5 system · 1.27:4 message · 1.6:3 directive · 0.51:4 call · 1.43:3 lesson
---	---	--

one can ...	14.93:2 modulate · 12.4:2 murmur · 8.62:5 recognise · 11.36:1 hush · 10.96:1 shriek · 3.44:8 hear · 9.28:2 amplify · 8.34:2 imitate · 3.72:6 lower · 6.7:3 obey · 7.2:2 mimic · 4.9:4 lend · 5.02:3 possess · 4.68:3 dub · 0.53:7 raise · 5.52:2 heed · 4.36:3 drown · 6.14:1 clip · 5.08:2 equal · 6.06:1 sharpen 8.54:4 creep into · 10.09:2 exclaim in · 9.09:2 mutter in · 2.63:8 speak with · 4.18:5 sing in · 8.07:1 retort in · 6.42:2 whisper in · 3.65:4 reply in · 4.24:3 cry in · 6.06:1 recite in · 5.78:1 startle by · 4.77:2 inject into · 2.77:4 listen to · 0.54:6 speak in · 3.39:3 detect in · 3.34:3 consist of · 2.91:3 shout in · 2.04:3 sing with · 0.78:3 sound like	a voice
a voice can ...	14.14:3 muffle · 12.44:4 tremble · 9.54:4 whisper · 11.44:2 crackle · 11.32:2 growl · 6.13:7 sound · 10.61:1 wobble · 9.49:2 drip · 9.49:2 thicken · 10.29:1 squeak · 7.85:3 falter · 5.82:5 echo · 8.69:2 waver · 7.43:3 harden · 8.24:2 reverberate · 8.7:1 exclaim · 5.38:4 fade · 5.25:4 shout · 6.17:3 deepen · 7.17:2 startle	
a voice can be	13.33:3 muffle · 12.39:2 hush · 8.1:3 clip · 9.75:1 tinge · 8.7:1 amplify · 1.58:7 hear · 4.68:3 dub · 5.12:2 choke · 4.09:2 drown · 3.62:2 strain	... 'ed

DeepDict prepositions: more syntax than semantics

vote ...	6.18:8 therefore · 4.87:8 today · 5.85:7 tomorrow · 4.24:6 unanimously · 3.62:5 against · 3.61:5 overwhelmingly · 1.49:7 justly · 2.31:5 yesterday · 4.22:3 firstly · 3.21:4 accordingly · 0.76:6 now · 3.47:3 tactically · 2.33:4 differently · 1.6:4 separately · 0.34:5 however · 0.16:5 so · 1.57:3 namely · 2.51:2 wholeheartedly · 1.37:3 actually · 0.26:4 simple · 0.93:3 freely · 0.89:3 naturally · 0.74:3 except · 1.53:2 intelligently · 0.43:3 narrowly
vote in ...	12.04:9 favour · 3.36:8 election · 3.66:6 referendum · 4.52:4 entirety · 1:6 committee · 0.61:5 vote · 0.07:4 chamber · 0.25:3 may · 0.06:1 plebiscite
vote against ...	5.36:9 report · 3.7:7 resolution · 3.66:7 amendment · 2.59:7 proposal · 3.58:6 motion · 3.03:4 paragraph · 2.33:4 directive · 1.6:4 text · 0.84:4 recommendation · 0.49:3 PROP-hum · 0.17:3 appointment · 1.16:2 accession · 0.53:2 discharge · 0.03:2 ratification
vote *for ...	5.06:9 report · 4.35:7 resolution · 3.67:7 amendment · 4.08:6 motion · 2.56:6 reason · 3.51:5 directive · 1.66:6 candidate · 3.22:4 censure · 1.12:5 proposal · 1.6:4 text · 2.24:3 accession · 0.54:4 bush · 0.03:4 regulation · 1.84:2 incorporation · 0.83:3 paragraph · 0.49:3 PROP-hum · 0.85:2 abolition · 0.51:2 postponement · 0.22:2 discharge · 0.01:2 tomorrow · 0.99:1 E · 0.82:1 deletion · 0.7:1 continuance · 0.7:1 guillotine · 0.5:1 assent
vote in ...	6.54:7 favour · 0.35:5 poll
vote *on ...	3.21:6 amendment · 1.97:7 report · 1.92:6 resolution · 2.38:5 text · 2.23:4 directive · 1.02:5 proposal · 1.01:5 matter · 1.34:4 motion · 1.72:3 paragraph · 0.53:4 basis · 1.42:3 enlargement · 0.29:3 statute · 0.25:3 may · 0.85:2 accession · 0.48:2 recital · 0.99:1 website · 0.01:1 scoreboard

DeepDict: explicit semantics (semantic prototypes)

spise (verb)

total of 28482 relations

Subjects:

PERS: jeg, vi, de, du, hun, han, man, dere, som, den, det, ingen, en

9.59:7 PROP-hum · 1.38:8 <H> ·
1.45:5 <Adom> · 2.74:3 katt · 1.46:4 folk ·
2.21:3 larve · 1.85:3 gjest · 2.72:2 mamma
· 2.53:2 ku · 0.38:4 barn · 3.14:1 fanden ·
3.02:1 bikkje · 3.02:1 kattunge ·
2.89:1 passasjerbåt · 2.77:1 lefse ·
1.54:2 elg · 0.35:3 dyr ·
2.33:1 spekkhogger · 1.26:2 dame ·
0.22:3 hund · 0.05:3 mor · 1.97:1 jaguar ·
1.88:1 pappa · 0.61:2 fange ·
0.58:2 pasient

Subclauses:

2.56:2 interr

Accusative objects:

PERS: hva

9.72:8 middag · 10.2:7 lunsj · 8.1:9 <food-h> ·
9.55:7 frokost · 7.23:8 <food-m-h> ·
6.52:8 <food-c-h> · 6.15:8 <food-m> · 6.93:7 mat ·
5.84:8 <food> · 7.92:5 brødskiye · 6.67:6 kjøtt ·
8.46:4 kveldsmat · 4.62:7 <Aich> · 6.59:5 PROP-hum ·
3.34:8 <temp> · 5.08:6 fisk · 5.39:5 brød · 6.32:4 kake
· 4.12:6 <food-c> · 7.07:3 matpakke · 6:4 pizza ·
5.99:4 kjeks · 2.99:7 <occ> · 4.62:5 frukt ·
6.59:3 potetgull · 2.26:6 <Azo> · 1.53:6 <amount> ·
2.38:5 <Aent> · 1.83:5 · 1.73:5 <Aorn> ·
1.35:5 <Adom> · 1.3:5 <A> · 1.29:5 <cm> ·
1.22:5 <drink-h> · 1.27:4 <drink-c-h> · 0.36:1 PROP

Verbal particles:

4.67:7 opp · 4.34:7 sammen · 1.69:2 stille














Parsers: beta.visl.sdu.dk

Corpora: corp.hum.sdu.dk

DeepDict: gramtrans.com

eckhard.bick@mail.dk

DeepDict: sources and sizes

	Parser	Lexicon	Grammar		Corpora
	DanGram	100.000 lexemes, 40.000 names	8.400 rules		ca. 67 mill. words (mixed) [+83 mill. news]
	PALAVRAS	70.000 lexemes, 15.000 names	7.500 rules		ca. 210 mill. words (news) [+170 mill. wiki a.o.]
	HISPAL	73.000 lexemes	4.900 rules		ca. 50 mill. words (Wiki, Europarl) [+36 mill. Internet]
	EngGram	81.000 val/sem	4.500 rules		ca. 210 mill. words (mixed) [+106 mill. email & chat]
	SweGram	65.000 val/sem	8.400 rules		ca. 60 mill. words (news, Europarl) [+ Wiki]
	NorGram	OBT / via DanGram	OBT / via DanGram		ca. 30 mill. words (Wikipedia) [+ internet]
	FrAG	57.000 lexemes	1.400 rules		- [+67 mill Wiki, Europarl]
	GerGram	25.000 val/sem	LS+1.300 rules		ca. 44 mill. words (Wiki, Europarl) [+ internet]
	EspGram	30.000 lexemes	2.600 rules		ca. 18 mill. words (mixed) [+ internet, wiki]
	ItaGram	30.600 lexemes	1.600 rules		- [+ 46 mill. Wiki, Europarl]

References

- Bick, Eckhard (2003-4), "A CG & PSG Hybrid Approach to Automatic Corpus Annotation". In: Kiril Simow & Petya Osenova (eds.), *Proceedings of SProLaC2003* (at Corpus Linguistics 2003, Lancaster), pp. 1-12
- Bick, Eckhard (2005-9), "Turning Constraint Grammar Data into Running Dependency Treebanks". In: Civit, Montserrat & Kübler, Sandra & Martí, Ma. Antònia (red.), *Proceedings of TLT 2005*, pp.19-27
- Bick, Eckhard (2006-1), "Functional Aspects in Portuguese NER", in: Renata Vieira et al. (eds.) *Computational Processing of the Portuguese Language* (Proceedings of PROPOR 2006, Itatiaia, May 15th-17th, 2006), pp.80-89. Springer
- Bick, Eckhard (2006-2), "Noun Sense Tagging: Semantic Prototype Annotation of a Portuguese Treebank". In: Hajic, Jan & Nivre, Joakim (red.), *Proceedings of the Fifth Workshop on Treebanks and Linguistic Theories (December 1-2, 2006, Prague, Czech Republic)*, pp.127-138
- Dowty, D. (1987). Thematic proto roles and argument selection, *Language* **67**: 547-619.
- Fillmore, C. (1968). The case for case, in E. Bach & R. Harms (eds), *Universals in linguistic theory*, Holt, Rinehart and Winston, New York.
- Foley, W. & van Valin, R. (1984). *Functional syntax and Universal Grammar*, CUP, Cambridge.
- Jackendoff, R. (1972). *Semantic interpretation in generative grammar*, The MIT Press, Cambridge, Ma.
- Karlsson, F., Voutilainen, A., Heikkilä, J. & Antilla, A. (1995). *Constraint Grammar: A language-independent system for parsing unrestricted text*, Mouton de Gruyter, Berlin.
- Palmer M, Kingsbury P, Gildea D. (2005) "The Proposition Bank: An Annotated Corpus of Semantic Roles". *Computational Linguistics* **31** (1): 71-106
- Taulé, M. et al., "Mapping Syntactic Functions into Semantic Roles". In: *Proceedings of TLT2005*: 185-194