English Resource Grammar + WordNet

A progress report

Dan Flickinger

DELPH-IN Summit 2025

Amsterdam

7 July 2025

Goals

- Link each (contentful) ERG lexical entry to WordNet synsets
- Update Redwoods treebank to include these sense distinctions
- Try to improve disambiguation using lexical semantics
- Produce MRSs enriched with WordNet synset information

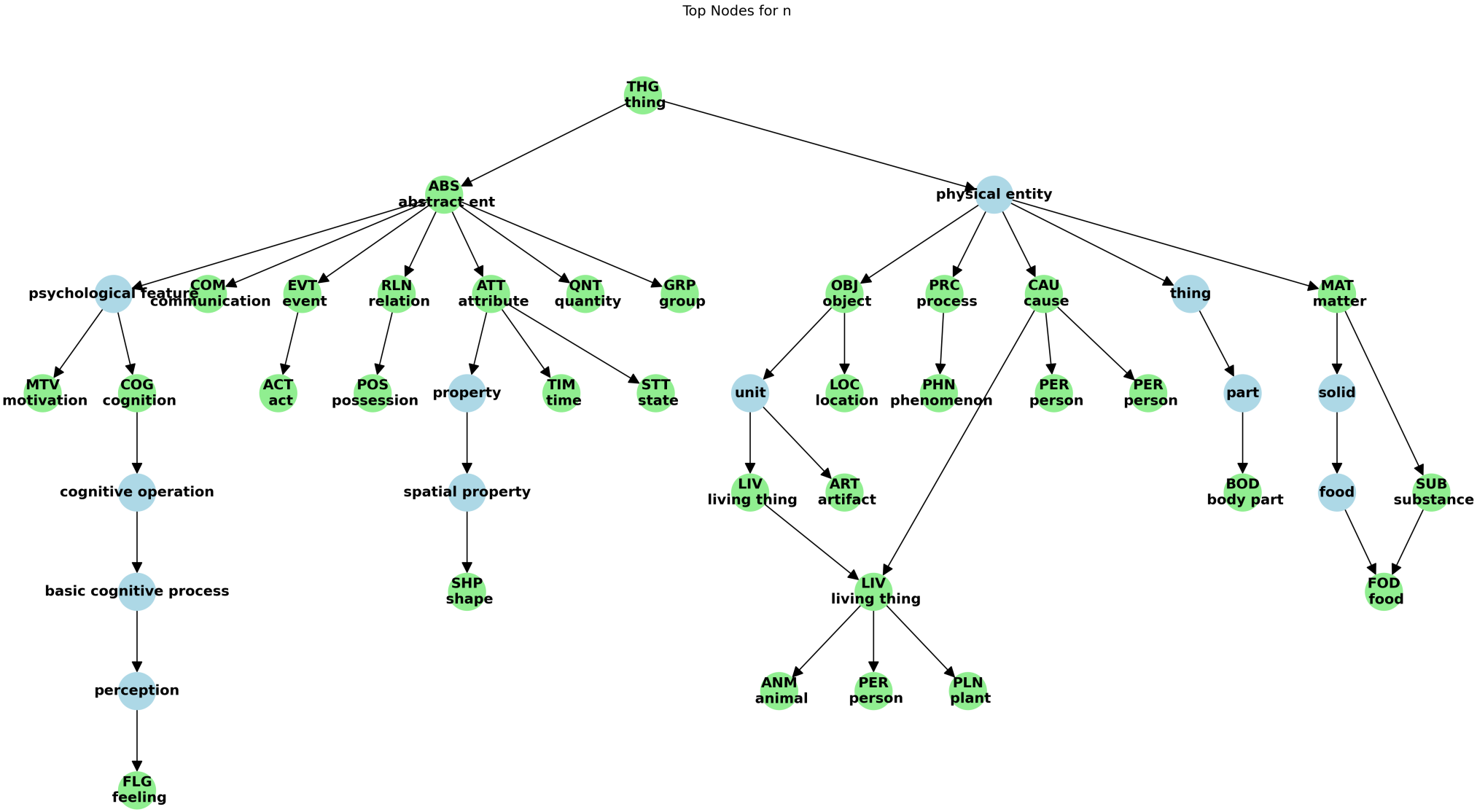
Motivations

- Increase validated lexical coverage of the ERG: 140,000 entries
- Include definitions for almost all lexical entries
- Provide more informative semantic output from the ERG
- Offer more detailed valence frames back to WordNet

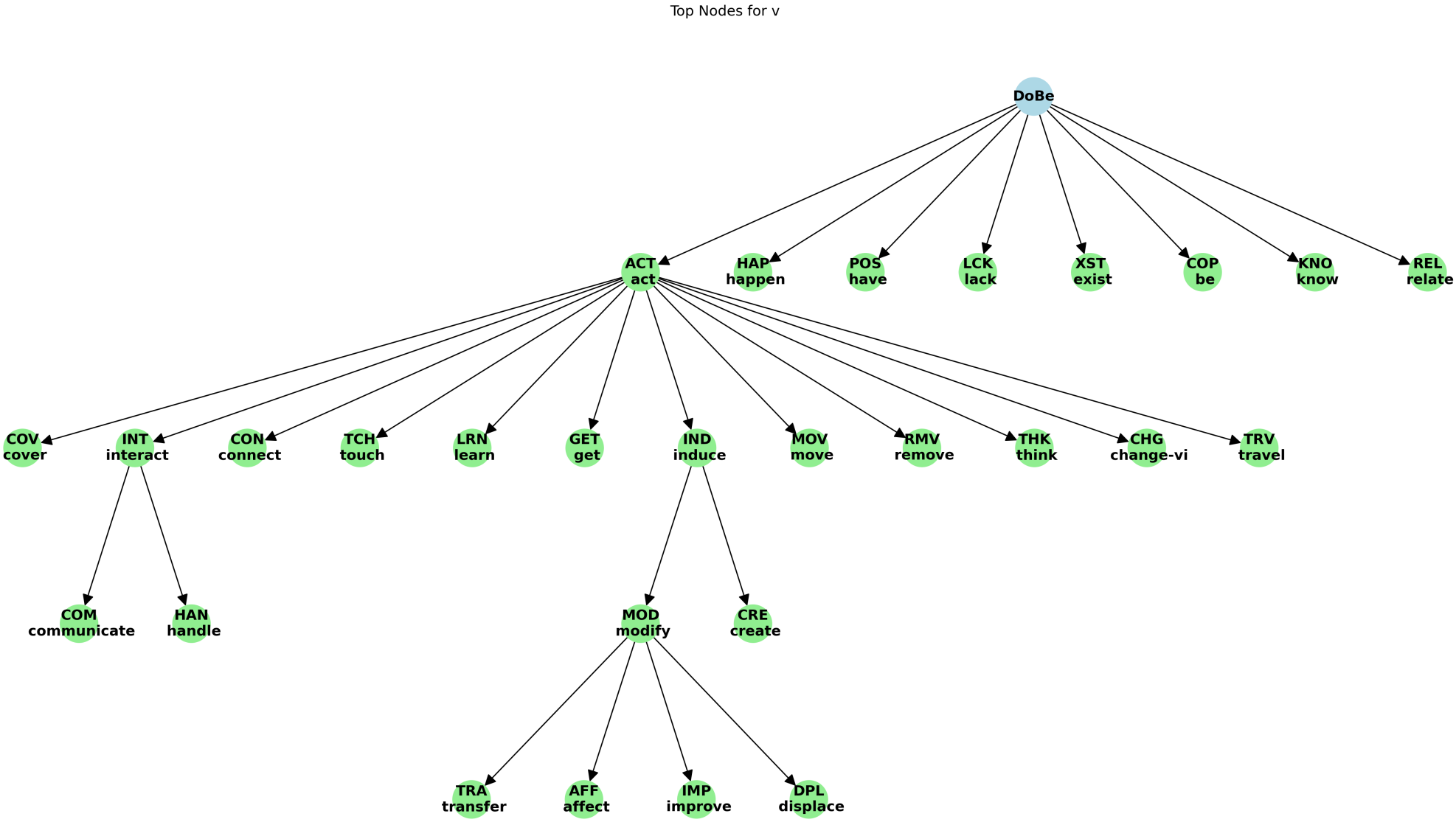
Strategy

- Choose broad clusters of WN noun and verb synsets (30 each)
  - Currently using Open English WordNet 2024
  - First pass used lexicographer files
  - Now using actual synsets to partition both nouns and verbs
- Auto-generate candidate enriched ERG lexical entries (55K)
  - (thanks, Francis, for successive refinements of this script)
- Manually curate these and finish linking what the script missed
  - Later, curate entries for words not yet in ERG (74K)
- Update the Redwoods treebank with these new sense distinctions
  - Started with Brown Corpus SemCor sentences, for sanity check
  - So far, updated first 40,000 words of 1.6M Redwoods treebank

Synset groupings for nouns



Synset groupings for verbs



WordNet verbs: 28 sets

act	02372362-v	act	2357
interact	02382049-v	int	130
communicate	00742582-v	com	1097
handle	02519853-v	han	126
move	01835473-v	mov	336
travel	01839438-v	trv	751
learn	00600349-v	lrn	204
think	00630153-v	thk	721
change	00109468-v	chg	1441
remove	00173351-v	rmv	201
touch	01208838-v	tch	197
cover	01335412-v	cov	189
connect	01357376-v	con	267
induce	00772482-v	ind	193
create	01620211-v	cre	754
modify	00126072-v	mod	1779
displace	01854282-v	dpl	1242
affect	00137133-v	aff	151
improve	00206293-v	imp	129
transfer	02225243-v	tra	469
get	02215637-v	get	242
happen	00340744-v	hap	
exist	02609706-v	xst	
have	02208144-v	pos	
know	00596016-v	kno	
relate	02681865-v	rel	
lack	02638434-v	lck	
be	02610777-v	cop	

WordNet nouns: 29 sets

substance	00020270-n	sub
food	00021445-n	fod
body part	05227735-n	bod
object	00002684-n	obj
location	00027365-n	loc
artifact	00022119-n	art
person	10398111-n	per
animal	00015568-n	anm
plant	00017402-n	pln
cause	00007347-n	cau
matter	00021007-n	mat
process	00029976-n	prc
phenomenon	00034512-n	phn
cognition	00023451-n	cog
feeling	05730374-n	flg
motivation	00023953-n	mtv
attribute	00024444-n	att
shape	05071206-n	shp
state	00024900-n	stt
time	00028468-n	tim
event	00029677-n	evt
act	00030657-n	acn
group	00031563-n	grp
relation	00032220-n	rln
possession	00032912-n	psn
communication	00033319-n	cmn
quantity	00033914-n	qnt
thing	00001740-n	thg

Lexical entry enrichment: *draft*

```
draft_v1 := v_np_le &
 [ ORTH < "draft" >,
   SYNSEM [ LKEYS.KEYREL.PRED "_draft_v_1_rel",
            PHON.ONSET con ] ].
```

```
draft_v1-cre := v_np_cre_le &
  [ ORTH < "draft" >,
    SYNSEM [ LKEYS.KEYREL.PRED "_draft_v_cre_rel",
             PHON.ONSET con ] ].
```

Lexical entry enrichment: *draft*

```
draft_v1 := v_np_le &
 [ ORTH < "draft" >,
   SYNSEM [ LKEYS.KEYREL.PRED "_draft_v_1_rel",
            PHON.ONSET con ] ].
```

```
draft_v1-cre := v_np_cre_le &
  [ ORTH < "draft" >,
    SYNSEM [ LKEYS.KEYREL.PRED "_draft_v_cre_rel",
             PHON.ONSET con ],
    SYNSEM.LKEYS.KEYREL.CONCEPTS < "01687477-v", "01705493-v" > ].
;; DEF make a blueprint of
;; DEF draw up an outline or sketch for something
;; EX draft a speech
```

Lexical entry enrichment: *draft*

```
draft_v1 := v_np_le &
 [ ORTH < "draft" >,
   SYNSEM [ LKEYS.KEYREL.PRED "_draft_v_1_rel",
            PHON.ONSET con ] ].
```

```
draft_v1-cre := v_np_cre_le &
  [ ORTH < "draft" >,
    SYNSEM [ LKEYS.KEYREL.PRED "_draft_v_cre_rel",
             PHON.ONSET con ],
    SYNSEM.LKEYS.KEYREL.CONCEPTS < "01687477-v", "01705493-v" > ].
;; DEF make a blueprint of
;; DEF draw up an outline or sketch for something
;; EX draft a speech

draft_v1-pos := v_np_pos_le &
  [ ORTH < "draft" >,
    SYNSEM [ LKEYS.KEYREL.PRED "_draft_v_pos_rel",
             PHON.ONSET con ],
    SYNSEM.LKEYS.KEYREL.CONCEPTS < "01099911-v" > ].
;; DEF engage somebody to enter the army
```

Treebanking example: *draft*

Hamilton was bent over his desk , drafting a legal paper by the light of a candle .

2 remaining. Gold tree is out.

/ 40607059 -- accepted
prev | next | accept | reject | list | exit [show | hide ignored text]

Exact Unary Chain

v_prp_olr | v_np_pos_le drafting **1 trees**	v_prp_olr | v_np_cre_le drafting **1 trees**

9 new manual

hdn_bnp-pn_c @n_sg_ilr @n_-_pn-per_le	=	0 to 1	[x]
hd-aj_scp-pr_c	+	2 to 18	[x]
hd-cmp_u_c	+	11 to 18	[x]
hd-cmp_u_c	+	7 to 11	[x]
n_sg_ilr @n_-_c-art_le	=	16 to 17	[x]
n_sg_ilr @n_-_c-cmn_le	=	10 to 11	[x]
hd_optcmp_c @v_pas_odlr @v_np*_chg_le	=	2 to 3	[x]
aj_-_i-prt_le	=	9 to 10	[x]
n_ms-cnt_ilr @n_-_mc-phn_le	=	13 to 14	[x]

27 new inferred

pt_-_comma-informal_le	-	6 to 7	[x]
hd-aj_scp-pr_c	-	1 to 18	[x]

Treebanking example: *paper*

Hamilton was bent over his desk , drafting a legal paper by the light of a candle .

3 remaining. Gold tree is out.

/ 40607059 -- accepted
prev | next | accept | reject | list | exit [show | hide ignored text]

Exact Unary Chain

n_sg_ilr	n_sg_ilr	n_sg_ilr
|	|	|
n_-_c-art_le	n_-_c-cmn_le	n_-_c-grp_le
paper	paper	paper
1 trees	1 trees	1 trees

9 new manual

hdn_bnp-pn_c @n_sg_ilr @n_-_pn-per_le	=	0 to 1	[x]
hd_optcmp_c @v_pas_odlr @v_np*_chg_le	=	2 to 3	[x]
v_prp_olr @v_np_cre_le	=	7 to 8	[x]
aj_-_i-prt_le	=	9 to 10	[x]
n_ms-cnt_ilr @n_-_mc-phn_le	=	13 to 14	[x]
hd-aj_scp-pr_c	+	2 to 18	[x]
hd-cmp_u_c	+	11 to 18	[x]
hd-cmp_u_c	+	7 to 11	[x]
n_sg_ilr @n_-_c-art_le	=	16 to 17	[x]

27 new inferred

pt_-_comma-informal_le	-	6 to 7	[x]
hd-aj_scp-pr_c	-	1 to 18	[x]
hd-cmp_u_c	-	2 to 18	[x]

Treebanking example: *light*

Hamilton was bent over his desk , drafting a legal paper by the light of a candle .

8 remaining. Gold tree is out.

/ 40607059 -- accepted
prev | next | accept | reject | list | exit [show | hide ignored text]

Exact Unary Chain

n_ms-cnt_ilr | n_-_mc-art_le light **1 trees**	n_ms-cnt_ilr | n_-_mc-att_le light **1 trees**	n_ms-cnt_ilr | n_-_mc-cmn_le light **1 trees**	n_ms-cnt_ilr | n_-_mc-cog_le light **1 trees**	n_ms-cnt_ilr | n_-_mc-loc_le light **1 trees**	n_ms-cnt_ilr | n_-_mc-per_le light **1 trees**
n_ms-cnt_ilr | n_-_mc-phn_le light **1 trees**	n_ms-cnt_ilr | n_-_mc-stt_le light **1 trees**				

9 new manual

hdn_bnp-pn_c @n_sg_ilr @n_-_pn-per_le	=	0 to 1	[x]
v_prp_olr @v_np_cre_le	=	7 to 8	[x]
hd-aj_scp-pr_c	+	2 to 18	[x]
hd-cmp_u_c	+	11 to 18	[x]
hd-cmp_u_c	+	7 to 11	[x]
n_sg_ilr @n_-_c-art_le	=	16 to 17	[x]
n_sg_ilr @n_-_c-cmn_le	=	10 to 11	[x]
hd_optcmp_c @v_pas_odlr @v_np*_chg_le	=	2 to 3	[x]
aj_-_i-prt_le	=	9 to 10	[x]

27 new inferred

pt_-_comma-informal_le	-	6 to 7	[x]
hd-aj_scp-pr_c	-	1 to 18	[x]
hd-cmp_u_c	-	2 to 18	[x]

Treebanking example: *legal*

Hamilton was bent over his desk , drafting a legal paper by the light of a candle .

2 remaining. Gold tree is out.

/ 40607059 -- accepted
prev | next | accept | reject | list | exit [show | hide ignored text]

aj_-_i-all_le	aj_-_i-prt_le
legal	legal
1 trees	1 trees

Exact Unary Chain

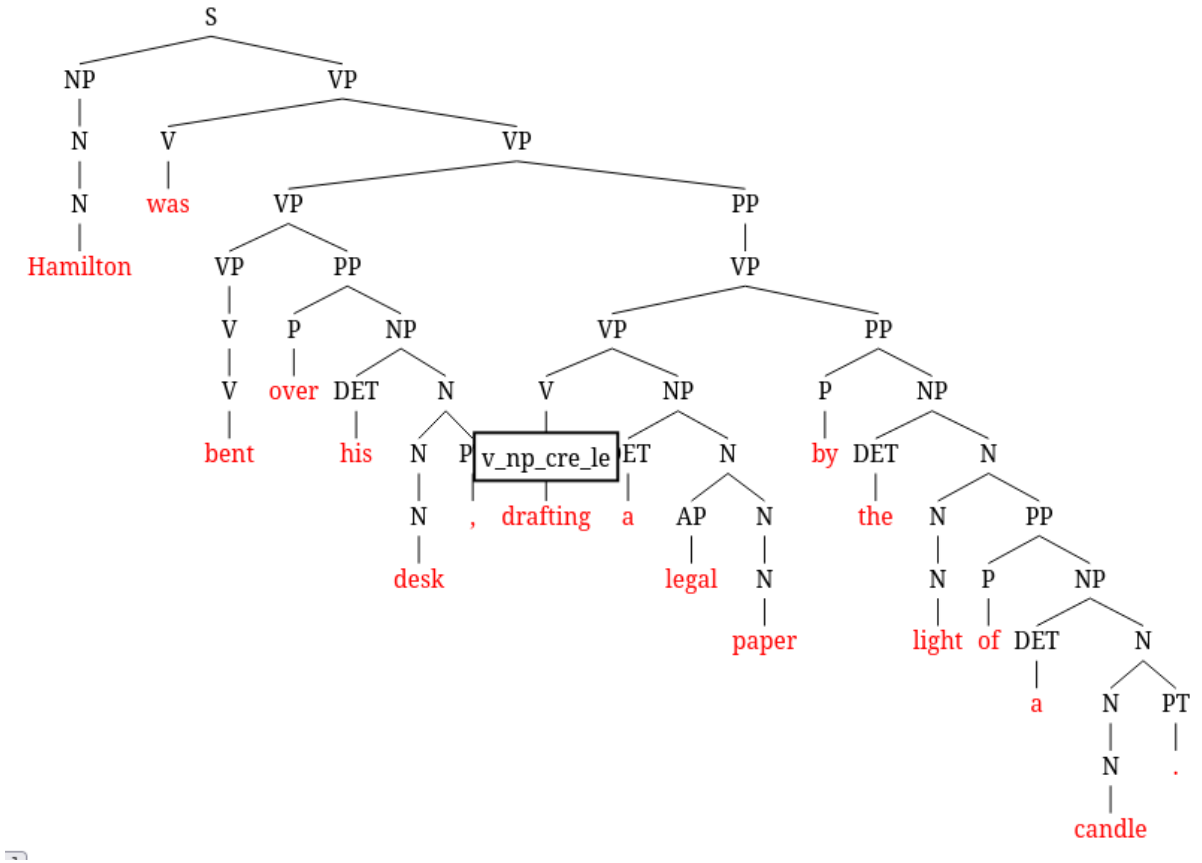
9 new manual

hdn_bnp-pn_c @n_sg_ilr @n_-_pn-per_le	=	0 to 1	[x]
v_prp_olr @v_np_cre_le	=	7 to 8	[x]
n_ms-cnt_ilr @n_-_mc-phn_le	=	13 to 14	[x]
hd-aj_scp-pr_c	+	2 to 18	[x]
hd-cmp_u_c	+	11 to 18	[x]
hd-cmp_u_c	+	7 to 11	[x]
n_sg_ilr @n_-_c-art_le	=	16 to 17	[x]
n_sg_ilr @n_-_c-cmn_le	=	10 to 11	[x]
hd_optcmp_c @v_pas_odlr @v_np*_chg_le	=	2 to 3	[x]

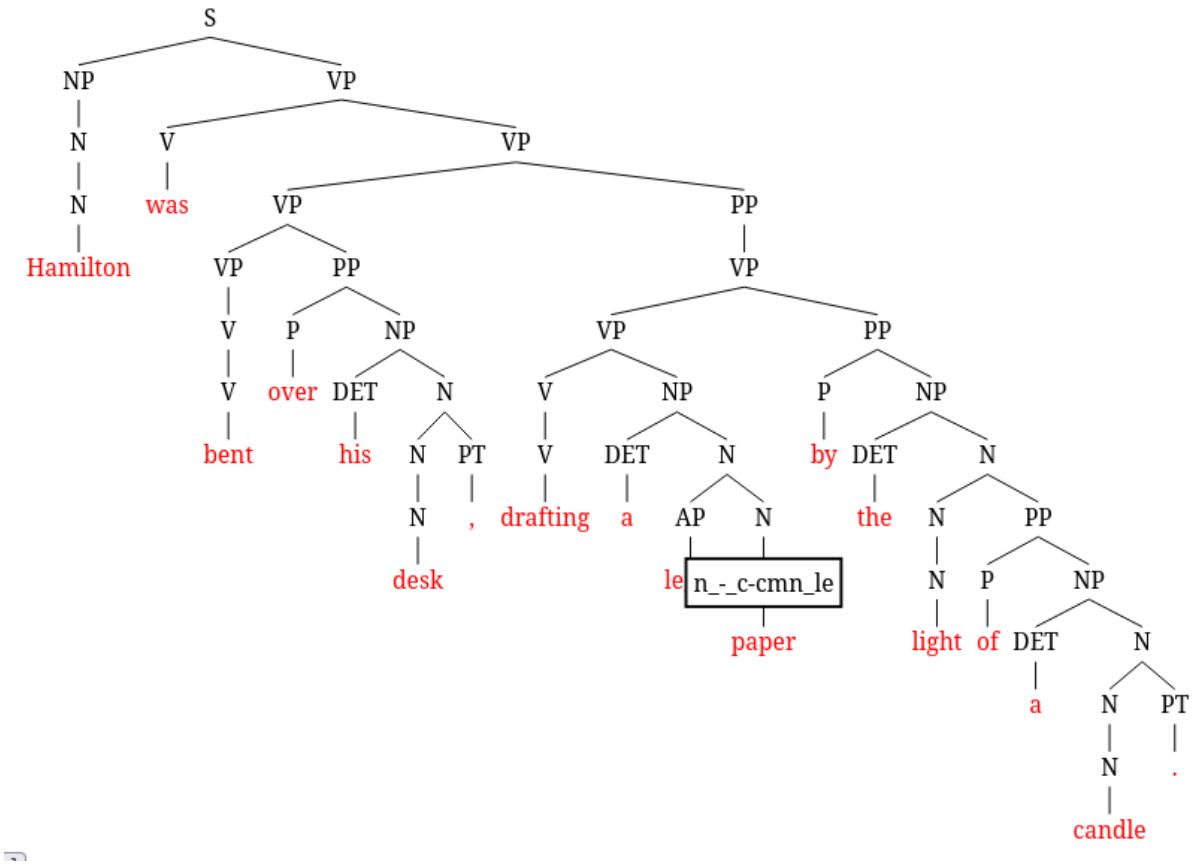
27 new inferred

pt_-_comma-informal_le	-	6 to 7	[x]
hd-aj_scp-pr_c	-	1 to 18	[x]
hd-cmp_u_c	-	2 to 18	[x]

Derivation tree added distinction: *draft*



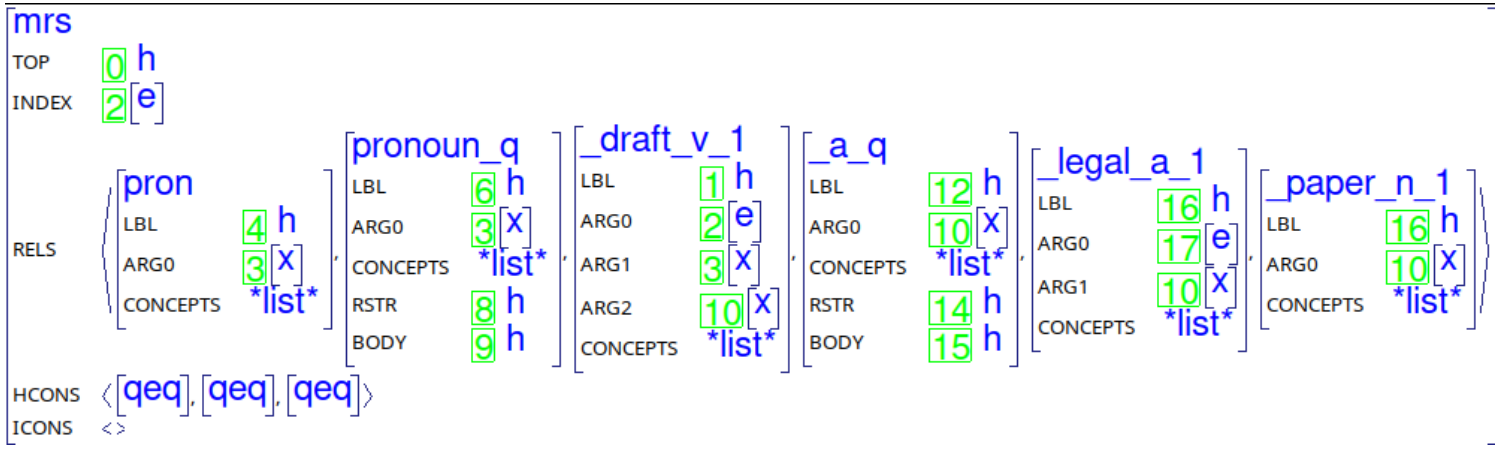
Derivation tree added distinction: *paper*



Sample enriched MRS

He was drafting a legal paper.

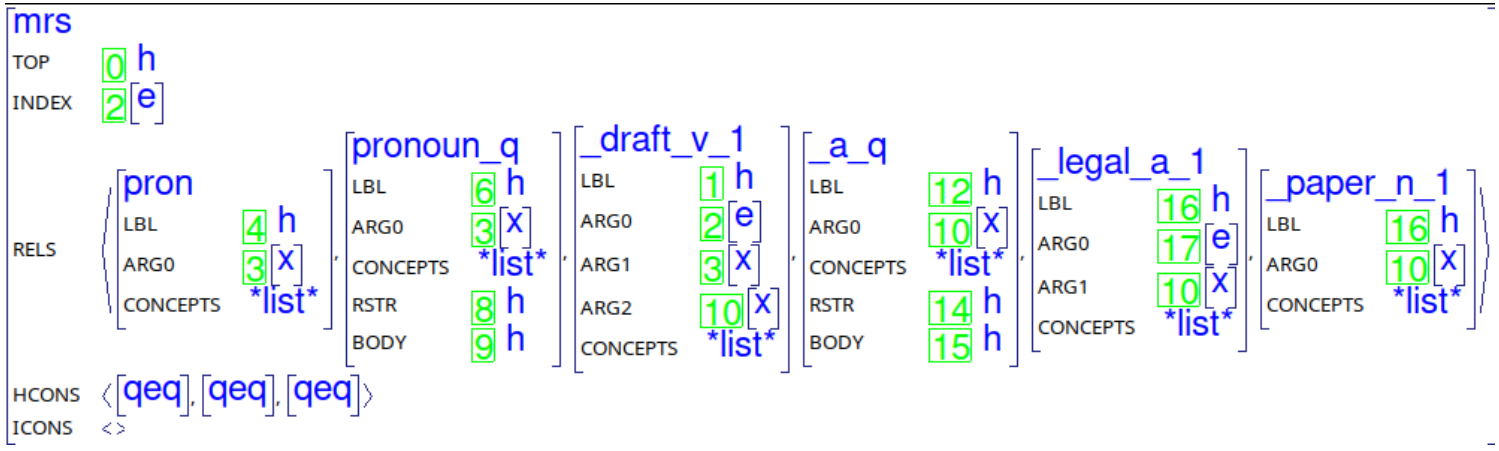
old



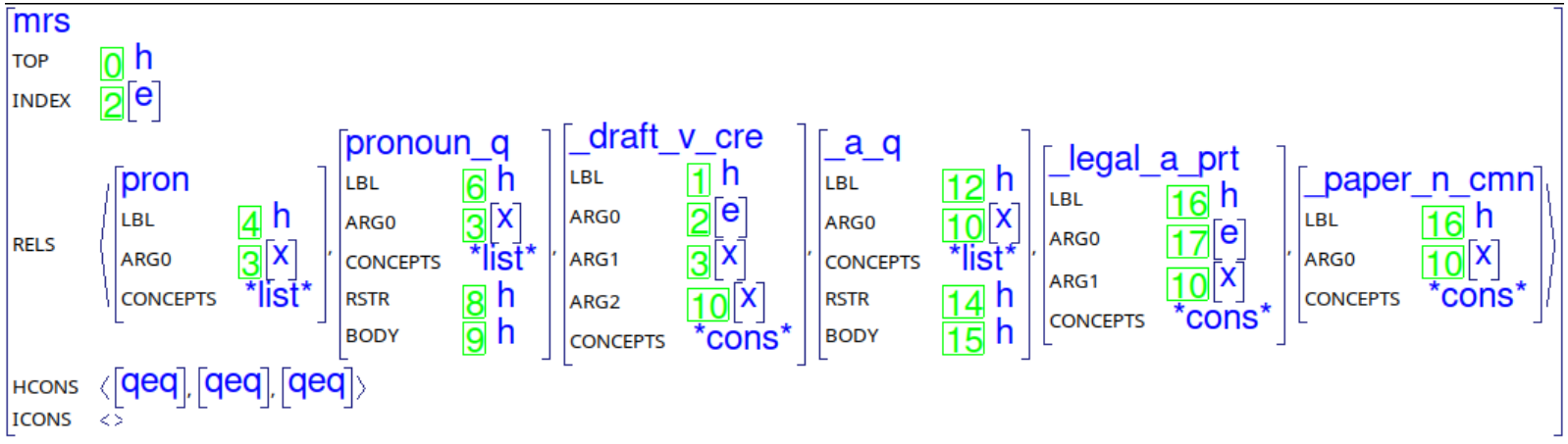
Sample enriched MRS

He was drafting a legal paper.

old



new



Modest increased parsing cost

Brown corpus sample profile: 2578 items, 16 tokens/item average

Length	(g)old			new			reduction		
	tasks	time	space	tasks	time	space	tasks	time	space
	ϕ	ϕ (s)	ϕ (kb)	ϕ	ϕ (s)	ϕ (kb)	%	%	%
$70 \leq 75$	-1	104.21	21504029	-1	116.35	21448633	-1.0	-11.7	0.3
$65 \leq 70$	-1	111.45	21504025	-1	129.18	21504026	-1.0	-15.9	-0.0
$60 \leq 65$	-1	64.49	13513758	-1	64.80	13678905	-1.0	-0.5	-1.2
$55 \leq 60$	-1	27.66	5830745	-1	25.71	5937781	-1.0	7.1	-1.8
$50 \leq 55$	-1	26.43	5384269	-1	29.03	5572079	-1.0	-9.8	-3.5
$45 \leq 50$	-1	25.14	4790197	-1	28.84	5485427	-1.0	-14.7	-14.5
$40 \leq 45$	-1	13.30	2677540	-1	13.79	2815153	-1.0	-3.6	-5.1
$35 \leq 40$	-1	8.68	1813070	-1	9.33	1878766	-1.0	-7.5	-3.6
$30 \leq 35$	-1	4.91	1008682	-1	5.22	1076423	-1.0	-6.3	-6.7
$25 \leq 30$	-1	2.67	549304	-1	2.95	592936	-1.0	-10.2	-7.9
$20 \leq 25$	-1	1.38	288606	-1	1.49	312715	-1.0	-7.9	-8.4
$15 \leq 20$	-1	0.57	125227	-1	0.66	140141	-1.0	-16.5	-11.9
$10 \leq 15$	-1	0.20	47315	-1	0.25	56846	-1.0	-26.6	-20.1
$5 \leq 10$	-1	0.07	16316	-1	0.09	21131	-1.0	-30.7	-29.5
$0 \leq 5$	-1	0.02	4519	-1	0.03	6151	-1.0	-13.3	-36.1
Total	**-1**	**1.85**	**380917**	**-1**	**2.01**	**407259**	**-1.0**	**-8.7**	**-6.9**

(generated by [incr tsdb()] at 6-jul-2025 (09:41 h))

Number of WN synsets per ERG lexical entry

synsets	entries
1	37936
2	9472
3	3041
4	1283
5	561
6	304
7	196
8	102
9	75
10	93
11	28
12	29
13	51
14	24
15	19
16	7
17	9
18	5
19	5
21	13
25	4
27	3
Total	53260