# XNL-Soar: where we are and where we're headed

Deryle Lonsdale, Jeremiah McGhee,
Nathan Glenn, and Tory Anderson

# What is XNL-Soar?

- Update of NL-Soar
- Incremental language processor built on Soar
- Intended to help in language modeling tasks
  - Language/task integrations
  - Interfaces, ambiguity, parsing, lexicon, etc.
- More detail last year (and prior)

# Updated how?

- Soar 7.2 → Soar 9.3.1
- C → Tcl → Java
- Hand-coded sparse lexicon → WordNet 3.1
- P&P syntax → Minimalist Program
- Xwindows → Clig → GraphViz

# Basic linguistic functionality

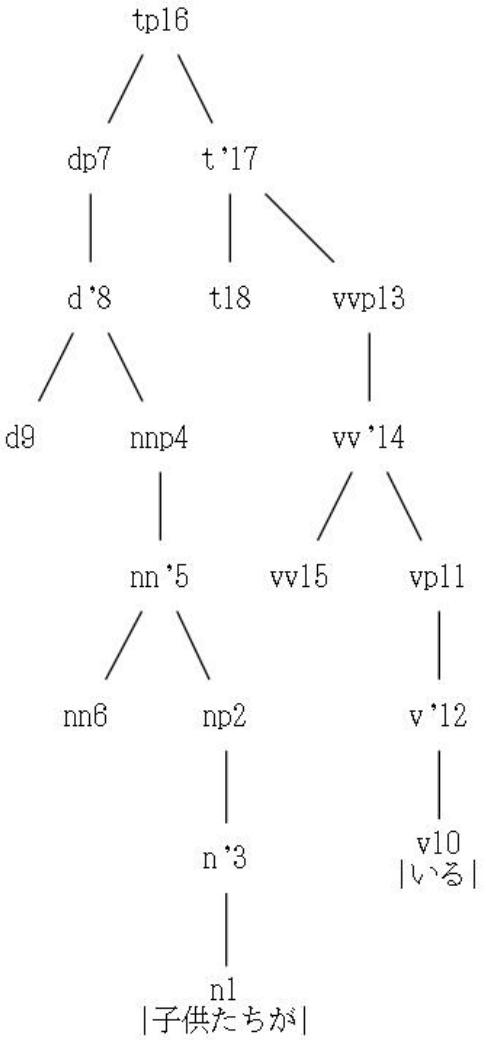
- Lexical input: word enters IO link
- Lexical access: retrieve all lexical, morphosyntactic, semantic info
- Incrementally build syn structure (parse trees) and sem structure (LCS)
- Snip operator when current hypothesis untenable
- Subcategorization, adjunction
- Empty categories, movements

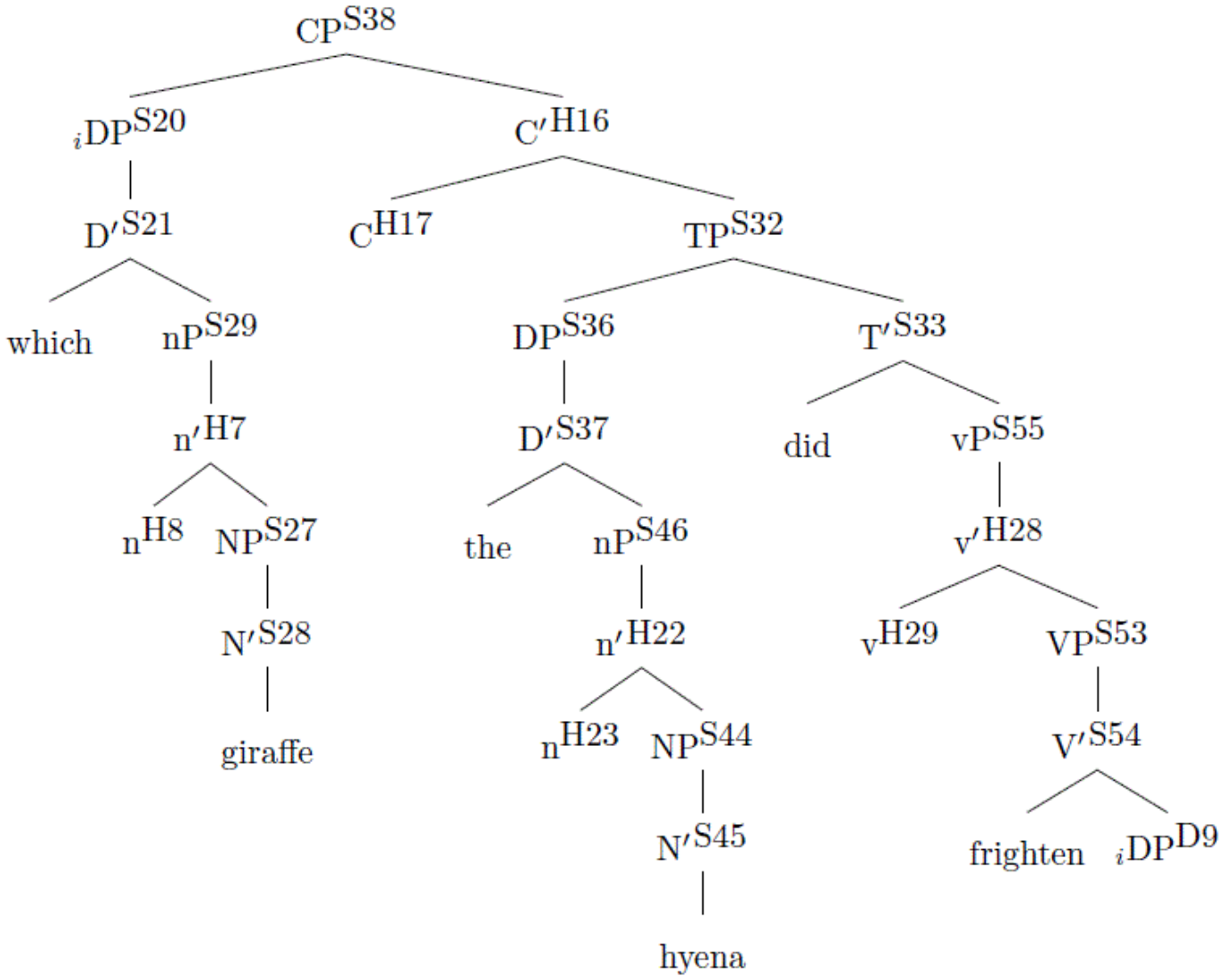
# New functionality via transition

- SVN repository
- Eclipse IDE
- A/R set graphing
- Smem, Epmem, RL (only Smem so far)
- Batch tester
- Various parsing strategies
- Japanese syntax

# New functionality via transition

| | |
|---|---|
| assigners | comp-C': H16[C' ]<br>comp-D': S21[D' which] , S37[D' the]<br>spec-DP: S20[DP which] , S36[DP the]<br>comp-N': S28[N' giraffe] , S45[N' hyena]<br>spec-NP: S27[NP giraffe] , S44[NP hyena]<br>comp-T': S33[t' did]<br>spec-TP: S32[tP did]<br>comp-V': S54[N' frighten]<br>spec-VP: S53[NP frighten] |
| receivers | comp-DP: S20[DP which] , S36[DP the]<br>comp-NP: S27[NP giraffe] , S44[NP hyena]<br>comp-TP: S32[tP did]<br>comp-VP: S53[NP frighten] |




CP S38
iDP S20
C′ H16
D′ S21
C H17
TP S32
which
nP S29
DP S36
T′ S33
n′ H7
D′ S37
did
vP S55
n H8
NP S27
the
nP S46
v′ H28
N′ S28
n′ H22
v H29
VP S53
giraffe
n H23
NP S44
V′ S54
N′ S45
frighten
iDP D9
hyena

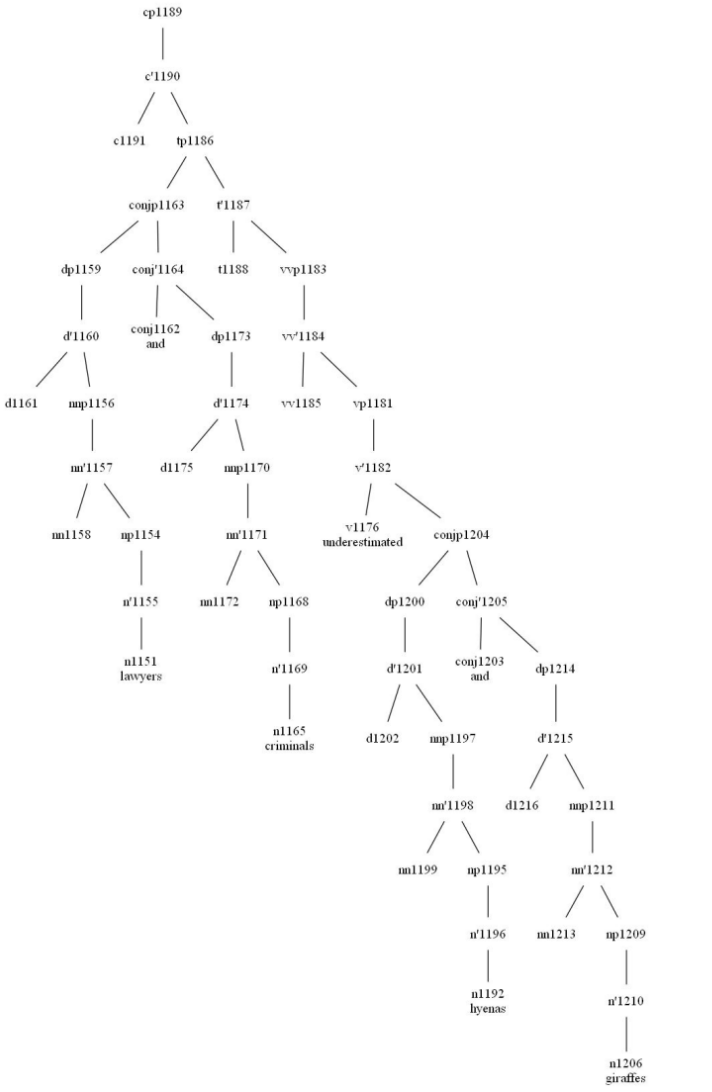
# Noteworthy milestones

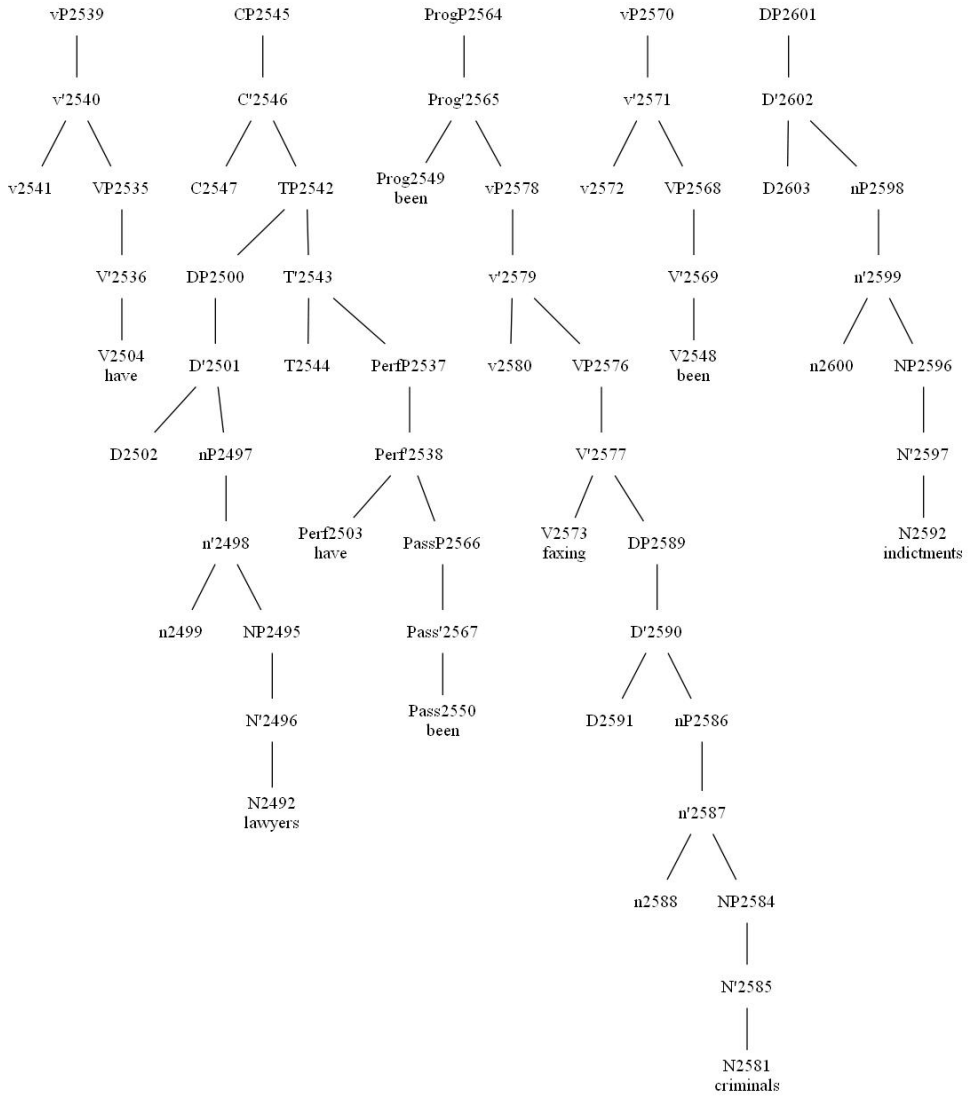
- Research award lecture
  - 45-minute-long live demo in front of whole college
  - English and Japanese sentences (syntax only)
  - Wide array of structural types, moderate complexity
- Japanese parsing
- Snip operators

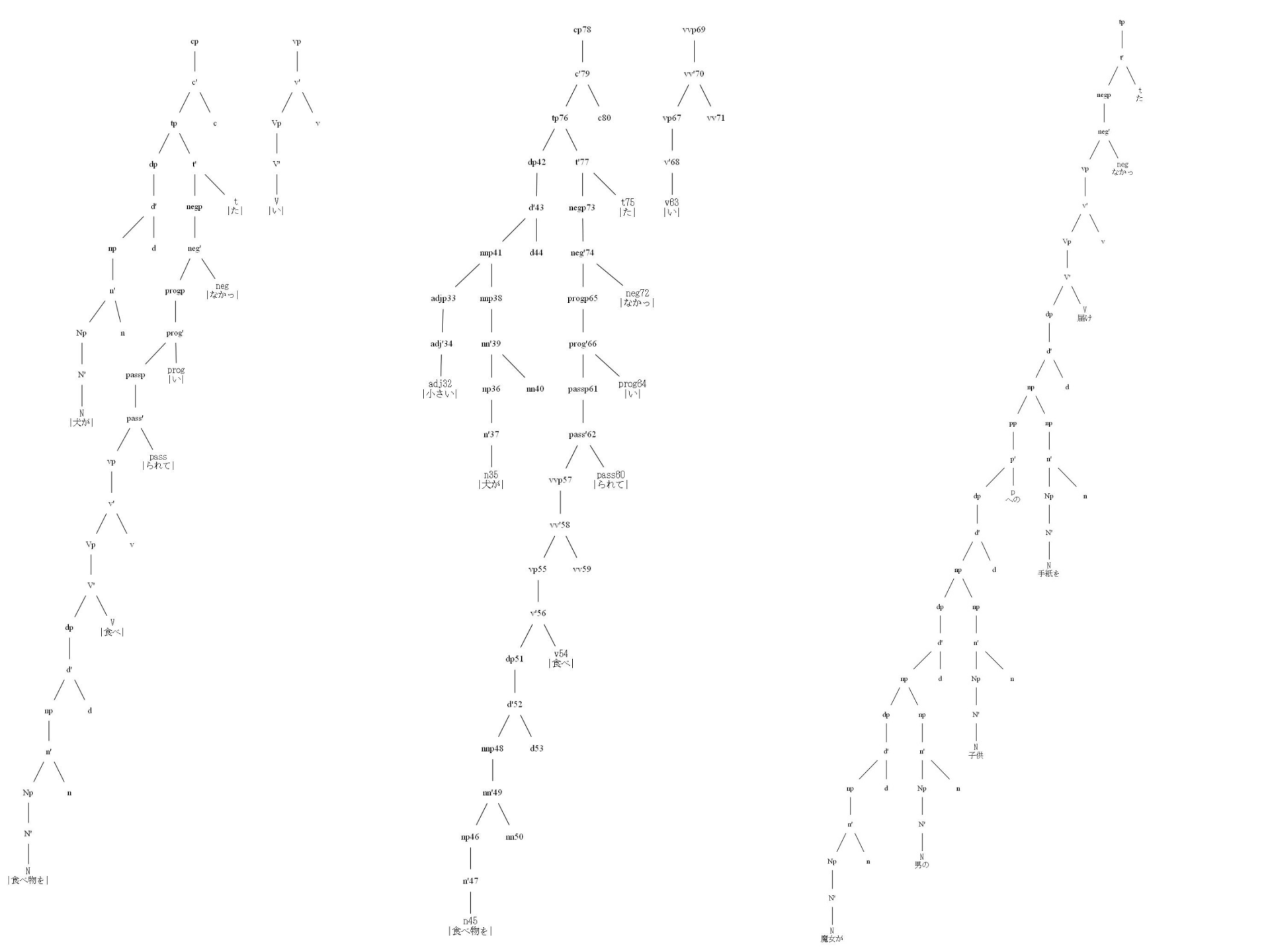
# Sample demo sentences

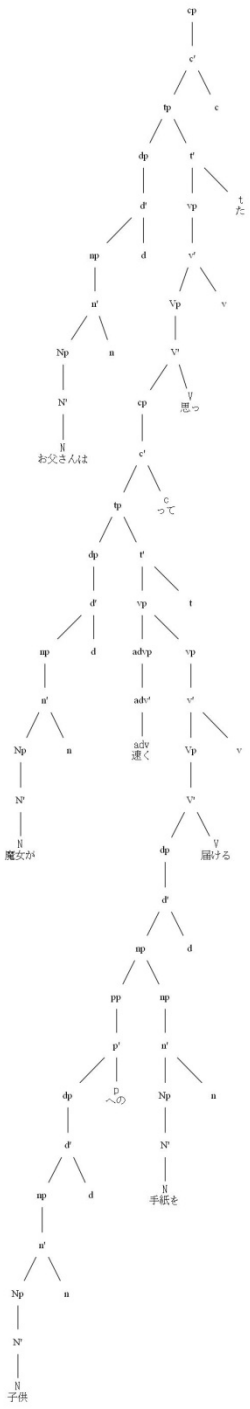
a wicked demon casts spells on disobedient children .#en
our resolution for redevelopment should not be underestimated .#en
we say a prayer of gratitude to the salmon .#en
猫が鼠を食べた。#JP
さわやかな川が流れていた。#JP
seagulls are gluttonous creatures .#en
Lushootseed tribes have been displaced .#en
revitalization is a component of our strategic direction .#en
the Salish respect matriarchal lineage .#en
ガールフレンドが新しい彼氏を見つけていた。#JP
the pebbles retain memories of important events .#en
the president has been orchestrating a political process .#en
宣教師は新しいガールフレンドを見つけなかった。#JP
they ceded their territories in the treaty of PointElliott .#en
お母さんが薬を飲んだ。#JP
revitalization is important to our community .#en
新しい彼氏は逃げた。#JP
outlandish stories frightened naughty children .#en
nobody should question our determination .#en
the reservation represents rural life .#en
she wanted to skin a deer .#en
the bureau maintains educational facilities on the reservation .#en

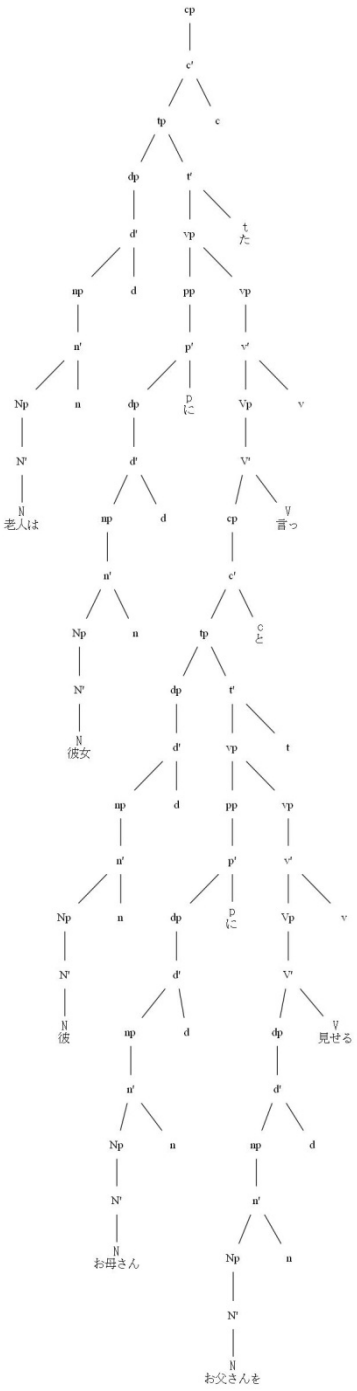
# Dealing with greater complexity

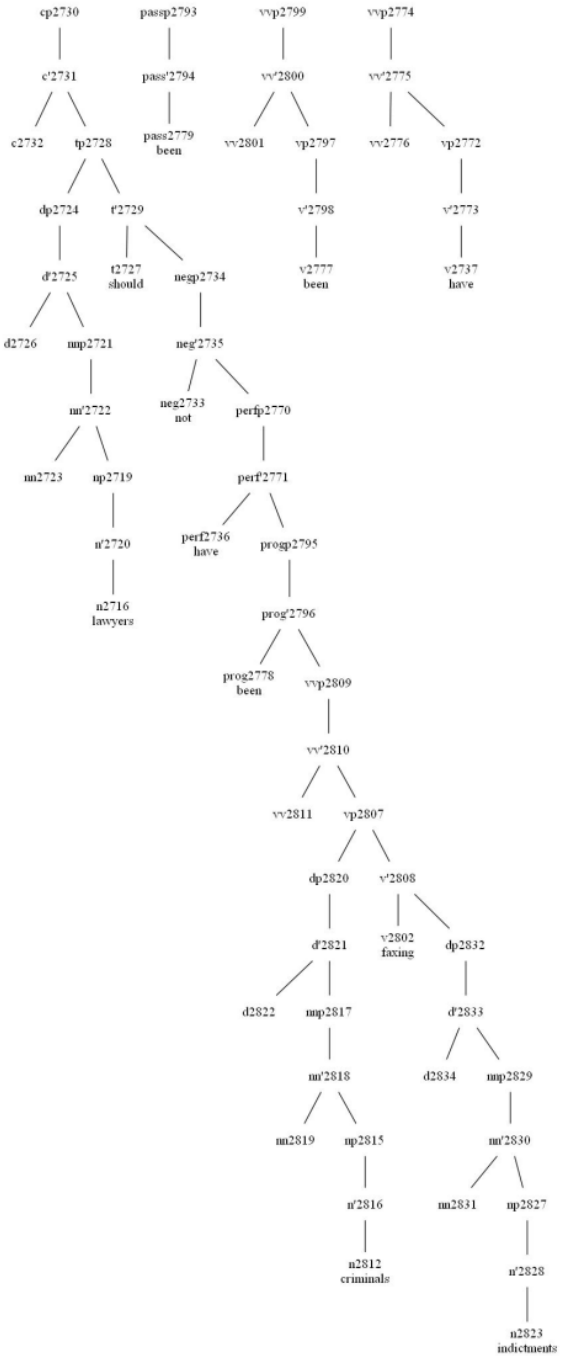



cp
c'
tp
c
dp
t'
d'
negp
t
|た|
np
d
neg'
n'
progp
neg
|なかっ|
Np
n
prog'
N'
passp
prog
|い|
N
|犬が|
pass'
vp
pass
|られて|
v'
Vp
v
V'
dp
V
|食べ|
d'
np
d
n'
Np
n
N'
N
|食べ物を|
vp
v'
Vp
v
V'
V
|い|
cp78
c'79
tp76
c80
dp42
t'77
d'43
negp73
t75
|た|
nnp41
d44
neg'74
adjp33
nnp38
progp65
neg72
|なかっ|
adj'34
nn'39
prog'66
adj32
|小さい|
np36
nn40
passp61
prog64
|い|
n'37
pass'62
n35
|犬が|
vvp57
pass60
|られて|
vv'58
vp55
vv59
v'56
dp51
v54
|食べ|
d'52
nnp48
d53
nn'49
np46
nn50
n'47
n45
|食べ物を|
vvp69
vv'70
vp67
vv71
v'68
v83
|い|
tp
t'
negp
t
た
neg'
vp
neg
なかっ
v'
Vp
v
V'
dp
V
届け
d'
np
d
pp
np
p'
n'
dp
P
への
Np
n
d'
N'
np
d
N
手紙を
dp
np
d'
n'
np
d
Np
n
dp
np
N'
d'
n'
N
子供
np
d
Np
n
n'
N'
Np
n
N
男の
N'
N
魔女が


cp
c'
tp
c
dp
t'
d'
vp
t
た
np
d
v'
n'
Vp
v
Np
n
V'
N'
cp
V
思っ
N
お父さんは
c'
c
って
tp
dp
t'
d'
vp
t
np
d
advp
vp
n'
adv'
v'
Np
n
adv
速く
Vp
v
N'
V'
N
魔女が
dp
V
届ける
d'
np
d
pp
np
p'
n'
D
への
dp
Np
n
d'
N'
np
d
N
手紙を
n'
Np
n
N'
N
子供


cp
c'
tp
c
dp
t'
d'
vp
t
た
np
d
pp
vp
n'
p'
v'
Np
n
dp
D
に
Vp
v
N'
d'
V'
N
老人は
np
d
cp
V
言っ
n'
c'
Np
n
tp
こと
N'
dp
t'
N
彼女
d'
vp
t
np
d
pp
vp
n'
p'
v'
Np
n
dp
D
に
Vp
v
N'
d'
V'
N
彼
np
d
dp
V
見せる
n'
d'
Np
n
np
d
N'
n'
N
お母さん
Np
n
N'
N
お父さんを


cp2730
passp2793
vvp2799
vvp2774
c'2731
pass'2794
vv'2800
vv'2775
c2732
tp2728
pass2779
been
vv2801
vp2797
vv2776
vp2772
dp2724
t'2729
v'2798
v'2773
d'2725
t2727
should
negp2734
v2777
been
v2737
have
d2726
mp2721
neg'2735
neg2733
not
perfp2770
mn'2722
mn2723
np2719
perf'2771
perf2736
have
progp2795
n'2720
n2716
lawyers
prog'2796
prog2778
been
vvp2809
vv'2810
vv2811
vp2807
dp2820
v'2808
d'2821
v2802
faxing
dp2832
d2822
mp2817
d'2833
mn'2818
d2834
mp2829
mn2819
np2815
mn'2830
n'2816
mn2831
np2827
n2812
criminals
n'2828
n2823
indictments

# Haven't yet made the transition

- Nominal compounding
- Chunking
- "Careful mode"
- Generation: sem → sentence
- Mapping: $lang_1$ sem → $lang_2$ sem
- Discourse/dialogue processing
- SimTime: buffered input
- French

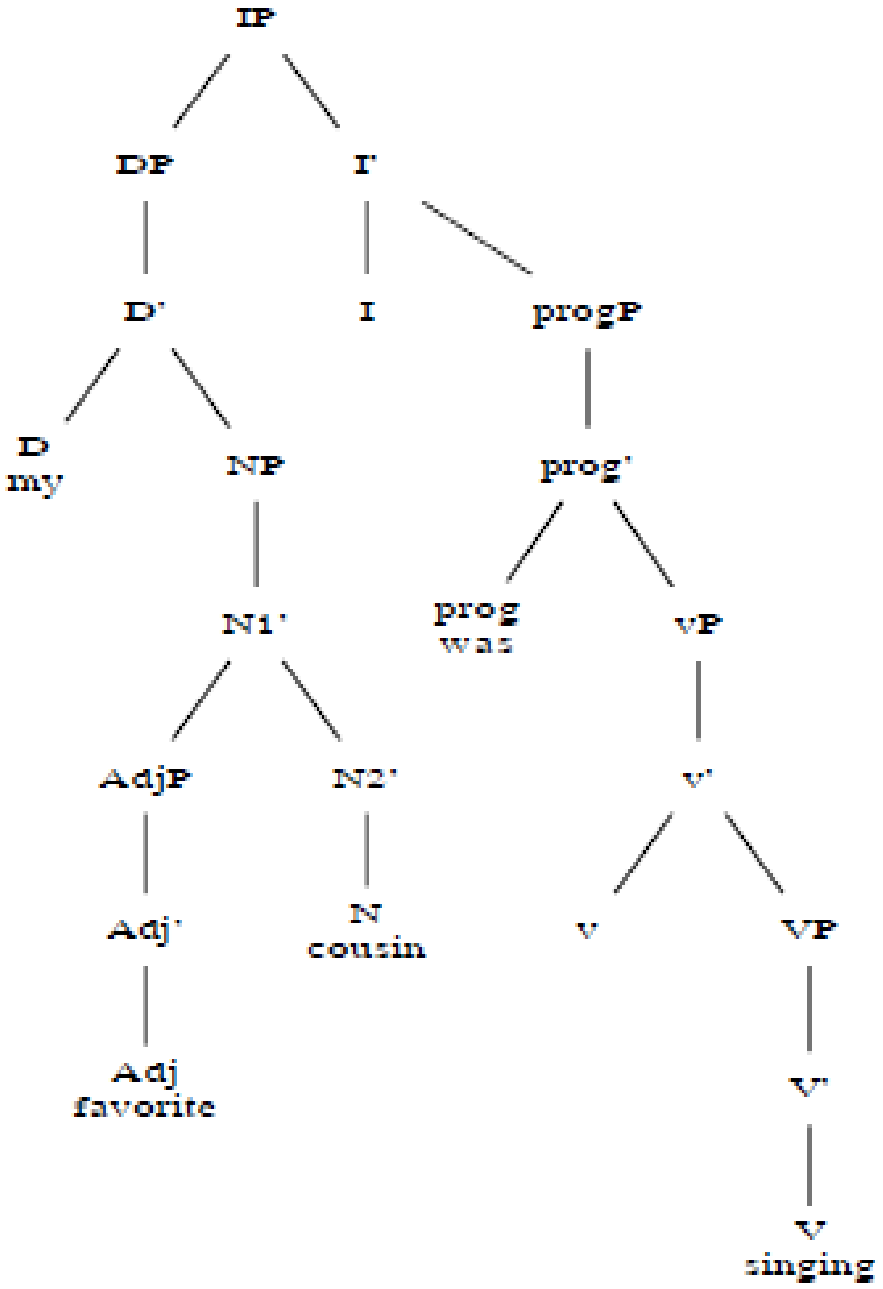
# Participle/gerund ambiguity

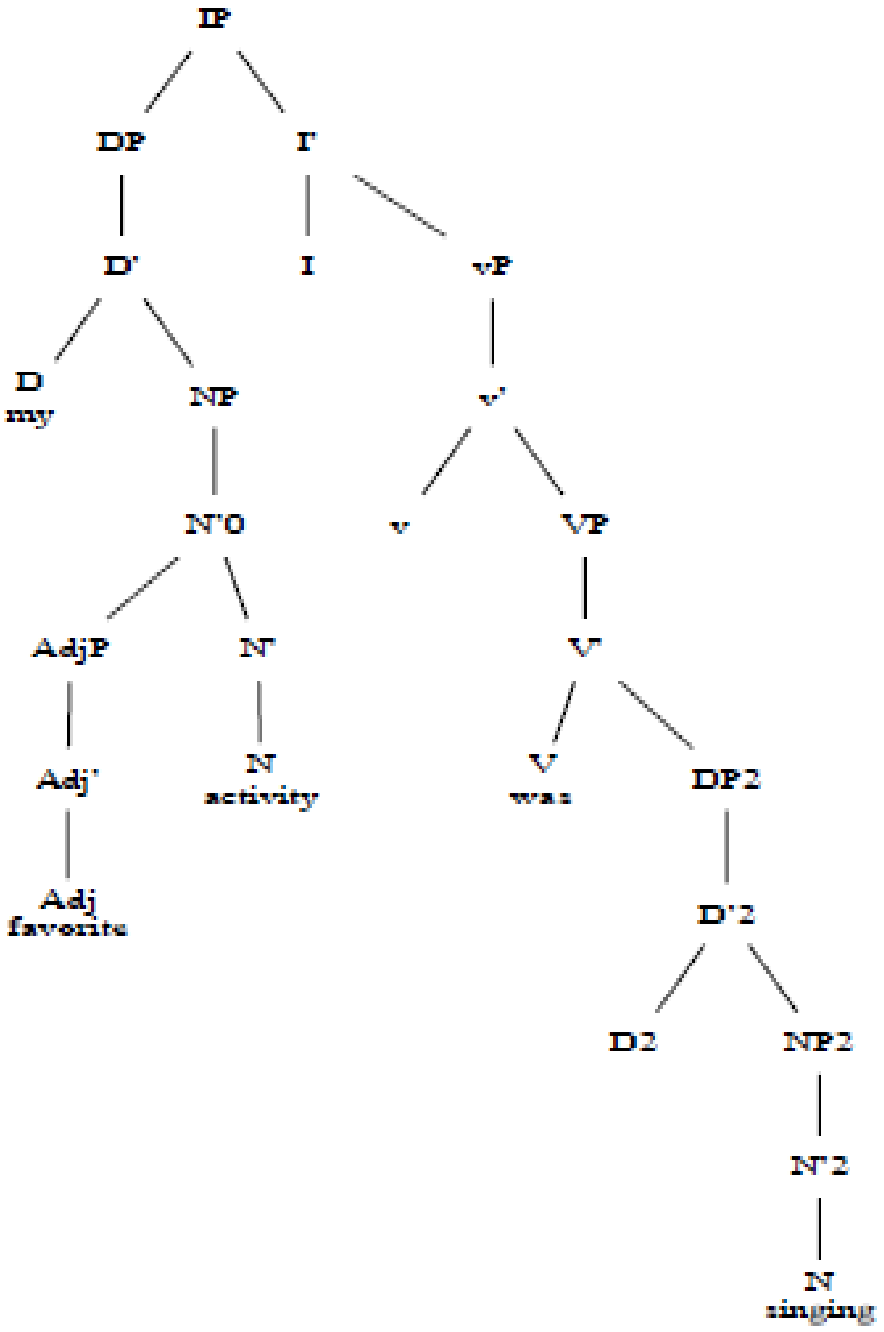
- My favorite cousin was singing. (participle)
- My favorite activity was singing. (gerund)
- His business is advertising. (both possible)

- Corpus-based evaluation of several strategies:
  - Force verb or noun
  - Random
  - Use Smem
  - Use analogy
  - Use sem relatedness metric

| Strategy | Correct | Incorrect |
|---|---|---|
| *forcenoun* | 39 | 49 |
| *forceverb* | 52 | 36 |
| *random(avg.)* | 44.0 | 43.7 |
| *Smem* | 46 | 42 |
| $\text{analogy}_{lex}$ | 47 | 41 |
| $\text{analogy}_{synsets}$ | 48 | 40 |
| *Resnik* | 68 | 20 |

Figure 7: System performance on test set of sentences (n=92, failed=4).


IP
DP
I'
D'
I
progP
D
my
NP
prog'
N1'
prog
was
vP
AdjP
N2'
v'
Adj'
N
cousin
v
VP
Adj
favorite
V'
V
singing


IP
DP
I'
D'
I
vP
D
my
NP
v'
N'0
v
VP
AdjP
N'
V'
Adj'
N
activity
V
was
DP2
Adj
favorite
D'2
D2
NP2
N'2
N
singing

# Sample syntactic snips

- He was ugly. ok: main verb "be"
- He was | singing. snip: main → prog
- He was | tricked. snip: main → pass

- He has a cold. ok: main verb "have"
- He has | died. snip: main → perf
- He had | been | singing. 2 snips

- He sent the man.
- He sent the man | an email.

# Sample syntactic snips (2)

- Lawyers suspected the man | was guilty.
- He eats spaghetti | and meatballs.
- I like green | vegetables.
- She believes the man | burped.

# Underway

- Items mentioned above in "Haven't yet…"

- Massively ambiguous parses
- Finish semantics
- Semsnips: semantic snips

- Epmem and language

- Project codebase recovery ☹

# Down the road

- Better morpho-syntax-synsem know-ledge (e.g. VerbNet interface)
- Interface with Sphinx
  - Modeling spoken language
- Multilingual agent
  - Language control
- Language acquisition modeling
- Task integrations (again)
- Careful reading (e.g. of web content)

# Conclusions

## NUGGETS

- Hundreds (more) sentences parsed
- Development/testing tools, IDE
- Interest despite hiatus
- Quantitative evaluations now possible (on a small scale)

## COAL

- Still not public-ready
- Juggling lots of functionality increments simultaneously
- Difficulty reaching "big data"-oriented computational community
- Temporary repo setback