Bootstrapping a stochastic parse selection model via SVD-mapped semantics

DELPH-IN Summit 2013

St. Wendel, Germany

DELPH-IN

Glenn Slayden
University of Washington

Tamping down the fan-out

- Mitigating fan-out is critical at every stage of DELPH-IN processing scenarios
- Especially problematic is MT, where parser results are passed on as inputs to transfer and then yet further to generation
- Stochastic parse (and realization) selection models become absolutely crucial as a grammar gains competency
- Maximum Entropy parse selection is a mature, core DELPH-IN technology, available in all processing engines

Corpora for discriminative modeling

- DELPH-IN parse selection models are trained to discriminate between the desired *vs.* undesired derivations in a parse result
- Building these models requires a corpus of parse results annotated for the desired parse
- Developing these training resources is very labor-intensive
- Low-resource languages may not be able to support this type of sustained development effort

Selected Prior work

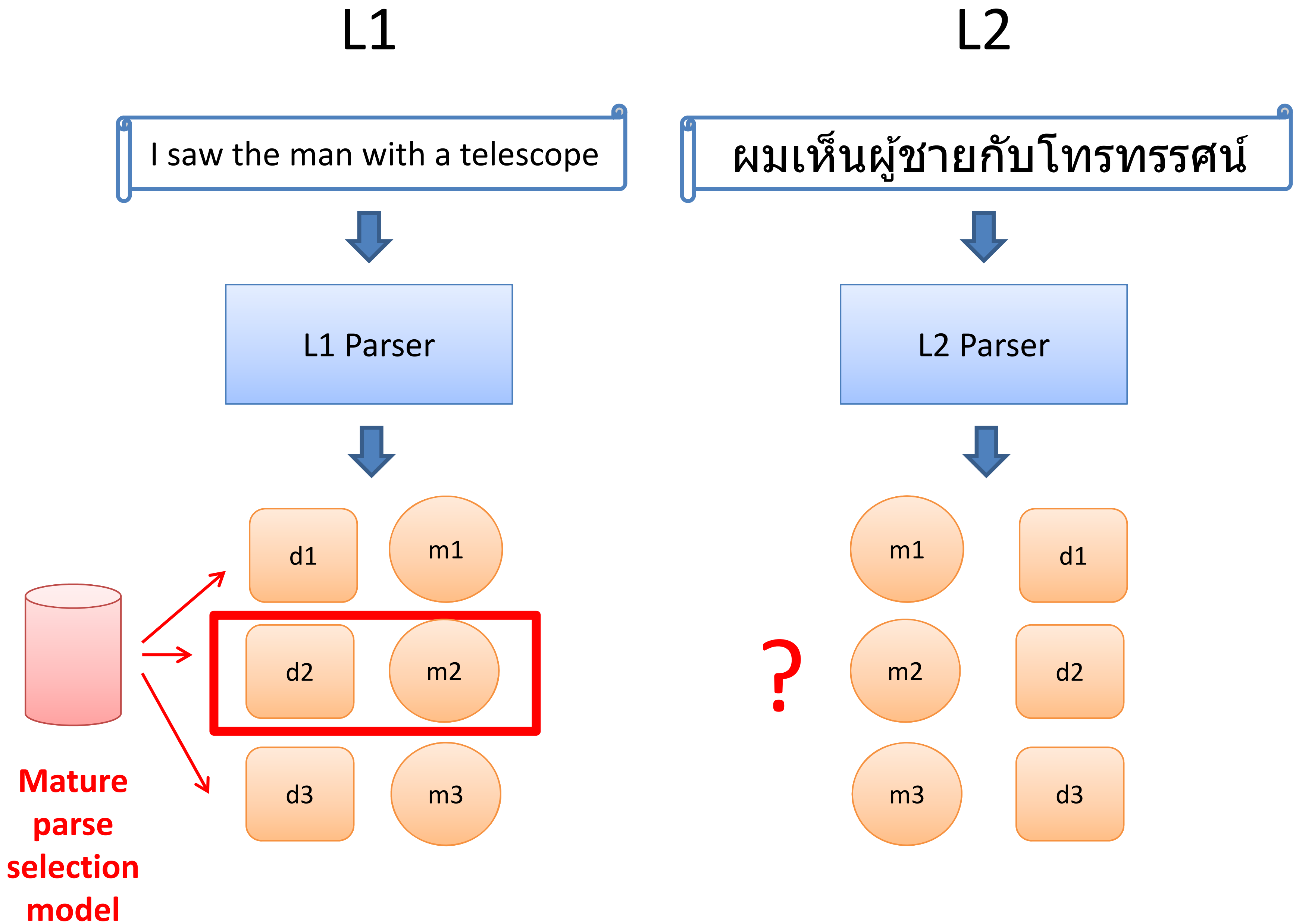
- Dridan & Oepen 2011. Parser evaluation using EDM
  - decomposing the MRS into elementary 'triples'
  - not concerned with setting triples in correspondence between disjoint MRSes
- Fujuta, Bond, Oepen & Tanaka 2010. Exploiting semantic information for HPSG parse selection

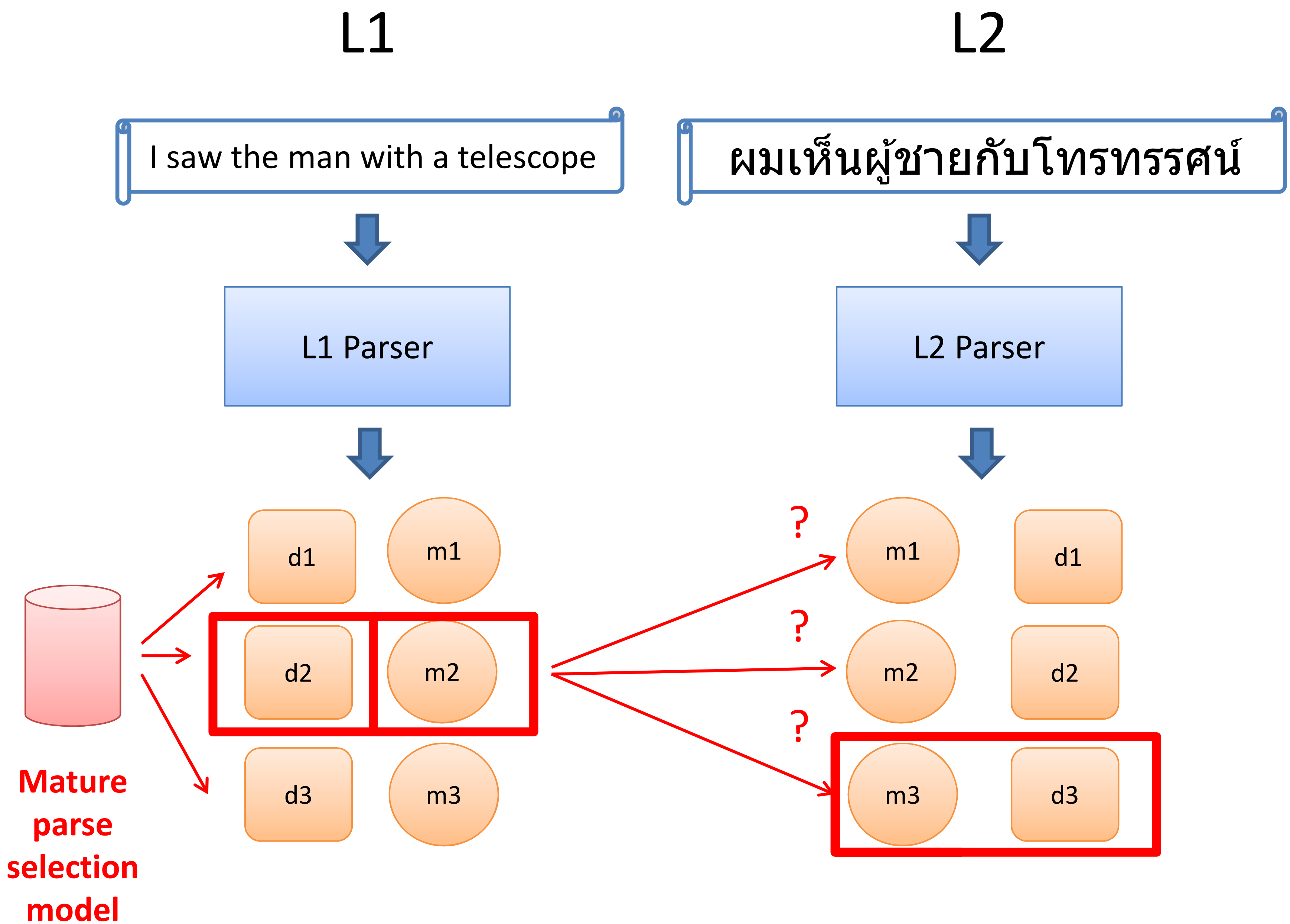
Motivation

- High-quality translation pairs are easier to obtain (and in volume) than discriminative derivation forests
- For these surface translation pairs, respective DELPH-IN grammars should produce similar semantics
  - modulo predicate names
  - as opposed to similar derivation trees
- Because each language independently pairs exactly one MRS with each derivation, MRS correspondence establishes one-to-one correspondence between bilingual derivations

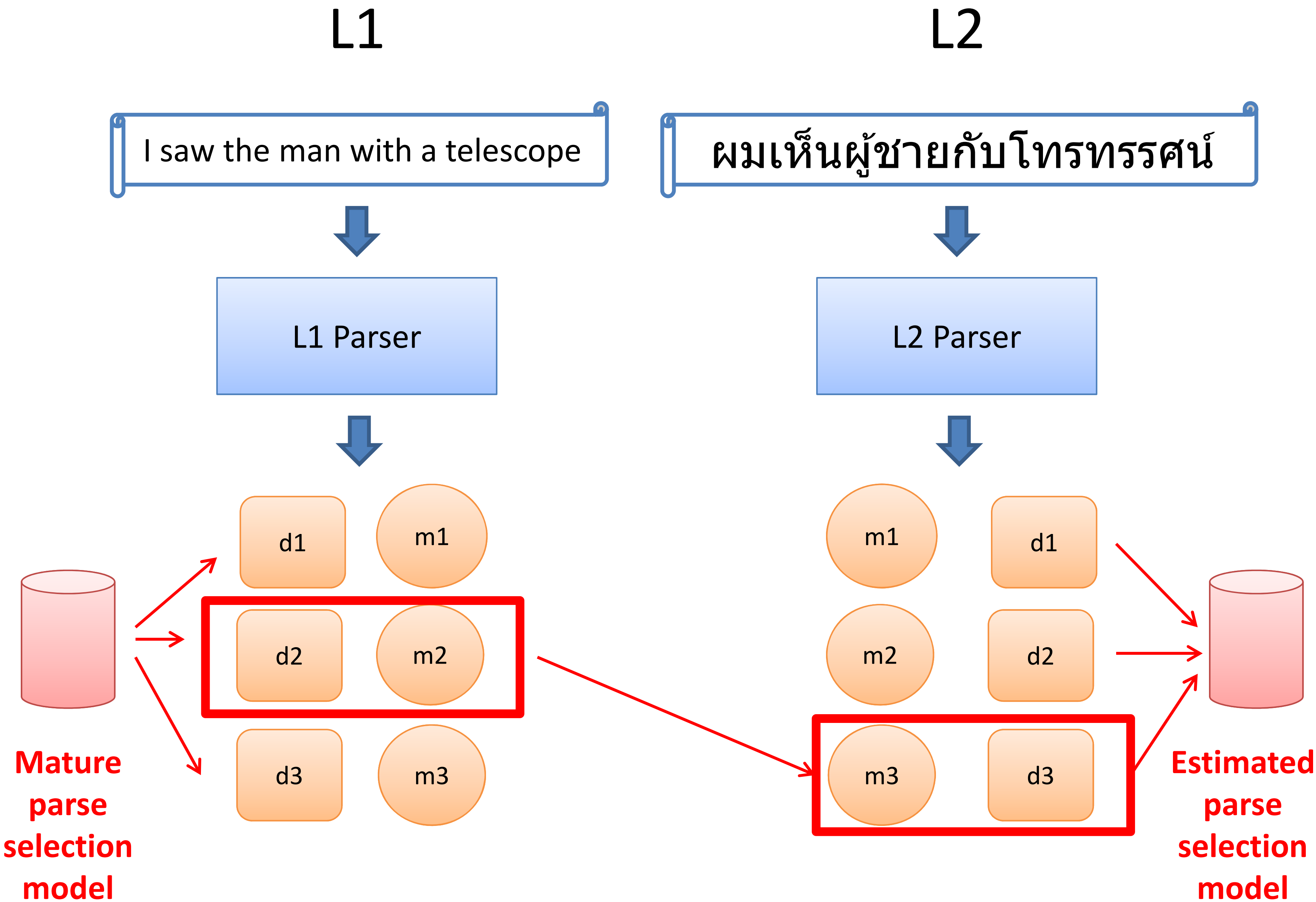
Semantic mediation

- This means that a rich and mature *syntactic* parse selection model from L1 can be used to estimate *syntactic* training data for L2
  - The estimation is mediated by *semantics*
- Given approximated L2 discriminations, a MaxEnt parse selection model is built for L2 in the normal way
  - TADM modeling toolkit (Malouf et al. 2005)


L1
L2
I saw the man with a telescope
ผมเห็นผู้ชายกับโทรทรรศน์
L1 Parser
L2 Parser
d1
m1
d2
m2
d3
m3
m1
d1
?
m2
d2
m3
d3
Mature parse selection model



L1
L2
I saw the man with a telescope
ผมเห็นผู้ชายกับโทรทรรศน์
L1 Parser
L2 Parser
d1
m1
d2
m2
d3
m3
Mature parse selection model
?
m1
d1
?
m2
d2
?
m3
d3



L1
L2
I saw the man with a telescope
ผมเห็นผู้ชายกับโทรทรรศน์
L1 Parser
L2 Parser
d1
m1
d2
m2
d3
m3
m1
d1
m2
d2
m3
d3
Mature parse selection model
Estimated parse selection model

What is this semantic mediation?

- What's needed is a robust, deterministic, grammar-agnostic metric of MRS similarity
- Since MRSes are formally DAGs, this is non-trivial
  - graph edit distance?
  - tree similarity? (but MRS is not a tree)
  - tree kernels? (but MRS is not a tree)

Desiderata for an isomorphism metric

- Proportional to the structural isomorphism between (abstract, arbitrary) directed graphs
  - do the MRSes have the same "shape?"
  - i.e. a similar structural signature as established by the occurrence of non-singleton variables
- Determinism guarantees
  - does the metric give an interpretable result for every MRS?
- Analytical power
  - does the metric maximize the use of available information?
  - can formally-defined aspects of MRS be fully exploited?
- Ignore grammar-specific types and predicates?

Singular value decomposition (SVD)

- SVD is a two-mode factor analysis which simultaneously achieves:
  - noise attenuation
  - redundancy detection (Schutze, 1992)
  - a similarity retrieval metric (Kontostathis and Pottenger, 2002)
- The well-known NLP application is in information retrieval (IR)
  - terms (rows) by documents (columns)

SVD definition

$$A_{m \times n} = U_{m \times d} \Sigma_{d \times d} (V_{n \times d})^T$$

$$d = \min(m, n)$$

A: (input matrix)

m: (columns) $\langle MRS, role, relation \rangle$

n: (rows) $roles \cup \langle MRS, variable \rangle$

MRS-SVD embedding

- How to embed MRS—formally a DAG—into matrix form?
- MRS has two structural levels:
  - relations, which group
  - role/variable assignments
- Solution: use *special rows* to tie together the role/variable assignments for each relation

	e:00-LTOP	e:00-XARG	e:00-INDEX	e:R0-LBL	e:R0-ARG0	e:R0-RSTR	e:R0-BODY	e:R1-LBL	e:R1-ARG0	e:R2-LBL	e:R
LTOP	1										
XARG		1									
INDEX			1								
LBL				1				1		1	
ARG0					1				1		
RSTR						1					
BODY							1				
ARG1											
ARG2											
HARG											
LARG											
x1-en		1			1				1		
h2-en						1					
h4-en								1			
h5-en	1									1	
e6-en			1								
h0-th											
x1-th											
h3-th											
h5-th											
e6-th											

http://www.computational-semantics.com/svd-align/mrs-svd.png

เด็ก กิน ข้าว

h5 e6 { SF prop }
{ h0 : *_child_n_1_rel*(x1 { PERS 3 })
h2 : *exist_q_rel*(x1, h3, h4)
h5 : *_eat_v_2_rel*(e6, x1) }
{ h3 *qeq* h0 }

The child is eating.

h5 e6 { SF prop, TENSE pres, MOOD indicative, PROG +, PERF - }
{ h0 : *_the_q_rel*(x1 { PERS 3, NUM sg, IND + }, h2, h3)
h4 : *_child_n_1_rel*(x1)
h5 : *_eat_v_1_rel*(e6, x1, p7) }
{ h2 *qeq* h4 }

MRS SVD embedding

Test scenario

- ERG (Flickinger 2000) trunk 13169
- Grammar of Thai based on Matrix (Bender et al. 2002)
- 187 Sentences parsed by both grammars
- pair-up one MRS from each grammar; embed both in a single matrix
- Reduce this matrix with SVD; see if the result says anything interesting about the isomorphism of the disjoint MRSes

Investigations

- What is the formal mathematical status of the MRS embedding proposed here?
- Are the singular values predictive?
  - initial excitement over w[0] now turns out to be a null result
- Excellent suggestions of Woodley and Guy (thanks!):
  - consider the distribution of singular values
  - compress each MRS individually first, then compare singular value vectors
- Further work on how to aggregate the multiple column vectors for a relation to obtain relation alignment
- much more…

latest results (1:47pm)

- Now studying 8 sentences
- http://www.computational-semantics.com/new-align/new-align.html
  - เขา ไป ซื้อ ดอกไม้ ที่ ตลาด และ ไป เยี่ยม เพื่อน
  - “She bought flowers at the market and went to visit a friend.”
  - see id ‘th219441’ (19 Thai parses) (select Thai #15?)

study subset

n-th	n-en			Maxent
6	1	The man can go.	root_strict	2.608923
6	1	The man went.	root_strict	0.813792
19	6	She bought flowers at the market and went to visit a friend.	root_strict	5.362326
2	7	Give way to passengers.	root_strict	2.582978
2	6	The cats and dogs are chasing cars.	root_strict	3.439535
10	2	The servant has returned.	root_strict	4.935633
2	1	He is reading a book.	root_strict	6.742530
4	2	I’m not the doctor.	root_informal	7.203028

ERG

```
h8 e6 { SF prop, TENSE pres, MOOD indicative, PROG -, PERF - }
{ h0 : pron_rel( x1 { PERS 1, NUM sg, PRONTYPE std_pron })
  h2 : pronoun_q_rel( x1, h3, h4 )
  h5 : _be_v_id_rel( e6, x1, x7 { PERS 3, NUM sg, IND + })
  h8 : neg_rel( e9 { SF prop, TENSE untensed, MOOD indicative, PROG -, PERF - }, h10 )
  h11 : _the_q_rel( x7, h12, h13 )
  h14 : _doctor_n_1_rel( x7 ) }
{ h3 qeq h0, h10 qeq h5, h12 qeq h14 }
```

I’m not the doctor - ผม ไม่ ได้ เป็น หมอ

1

```
h13 e15 { SF prop }
{ h0 : pron_rel( x1 { PERS 1, NUM sg, GEND m, SPECI + })
  h2 : exist_q_rel( x1, h3, h4 )
  h5 : neg_rel( e6, h7 )
  h8 : _can_v_rel( e9, x1, x10 { PERS 3 })
  h11 : _be_v_id_rel( e12, x1, x10 )
  h13 : _and_c_rel( h5, h14, e15, e9, e12 )
  h16 : _doctor_n_1_rel( x10 )
  h17 : exist_q_rel( x10, h18, h19 ) }
{ h3 qeq h0, h7 qeq h8, h18 qeq h16 }
```

2

```
h5 e15 { SF prop }
{ h0 : pron_rel( x1 { PERS 1, NUM sg, GEND m, SPECI + })
  h2 : exist_q_rel( x1, h3, h4 )
  h5 : neg_rel( e6, h7 )
  h8 : _can_v_rel( e9, x1, x10 { PERS 3 })
  h11 : _be_v_id_rel( e12, x1, x10 )
  h13 : _and_c_rel( h8, h14, e15, e9, e12 )
  h16 : _doctor_n_1_rel( x10 )
  h17 : exist_q_rel( x10, h18, h19 ) }
{ h3 qeq h0, h7 qeq h13, h18 qeq h16 }
```

3

```
h5 e9 { TENSE past, SF prop }
{ h0 : pron_rel( x1 { PERS 1, NUM sg, GEND m, SPECI + })
  h2 : exist_q_rel( x1, h3, h4 )
  h5 : neg_rel( e6, h7 )
  h8 : _be_v_id_rel( e9, x1, x10 { PERS 3 })
  h11 : _doctor_n_1_rel( x10 )
  h12 : exist_q_rel( x10, h13, h14 ) }
{ h3 qeq h0, h7 qeq h8, h13 qeq h11 }
```

4

```
h5 e15 { SF prop }
{ h0 : pron_rel( x1 { PERS 1, NUM sg, GEND m, SPECI + })
  h2 : exist_q_rel( x1, h3, h4 )
  h5 : neg_rel( e6, h7 )
  h8 : _can_v_rel( e9, x1, x10 { PERS 3 })
  h11 : _be_v_id_rel( e12, x1, x10 )
  h13 : _and_c_rel( h8, h14, e15, e9, e12 )
  h16 : _doctor_n_1_rel( x10 )
  h17 : exist_q_rel( x10, h18, h19 ) }
{ h3 qeq h0, h7 qeq h13, h18 qeq h16 }
```

Alignment # 3
from previous slide

role accuracy: 1.0000
const-type precision: 1.0000
const-type recall: 1.0000
const-value accuracy: 0.9091
var-subtype accuracy: 0.9333
variable precision: 0.5625
variable recall: 0.6000

[0] T230338-3 00 LTOP	h5	[0] E230338-1 00 LTOP	h8
[0] T230338-3 00 XARG	i15	[0] E230338-1 00 XARG	x1
[0] T230338-3 00 INDEX	e9	[0] E230338-1 00 INDEX	e6
[1] T230338-3 e9 TENSE	past	[8] E230338-1 e9 TENSE	untensed
[1] T230338-3 e9 SF	prop	[8] E230338-1 e9 SF	prop
[2] T230338-3 R0 PRED	pron_rel	[3] E230338-1 R0 PRED	pron_rel
[2] T230338-3 R0 LBL	h0	[7] E230338-1 R3 LBL	h8
[2] T230338-3 R0 ARG0	x1	[9] E230338-1 R4 ARG0	x7
[3] T230338-3 x1 PERS	1	[1] E230338-1 x1 PERS	1
[3] T230338-3 x1 NUM	sg	[6] E230338-1 x7 NUM	sg
[4] T230338-3 R1 PRED	exist_q_rel	[4] E230338-1 R1 PRED	exist_q_rel
[4] T230338-3 R1 LBL	h2	[9] E230338-1 R4 LBL	h11
[4] T230338-3 R1 ARG0	x1	[4] E230338-1 R1 ARG0	x1
[4] T230338-3 R1 RSTR	h3	[4] E230338-1 R1 RSTR	h3
[4] T230338-3 R1 BODY	h4	[4] E230338-1 R1 BODY	h4
[5] T230338-3 R2 PRED	neg_rel	[7] E230338-1 R3 PRED	neg_rel
[5] T230338-3 R2 LBL	h5	[7] E230338-1 R3 LBL	h8
[5] T230338-3 R2 ARG0	e6	[7] E230338-1 R3 ARG0	e9
[5] T230338-3 R2 ARG1	h7	[7] E230338-1 R3 ARG1	h10
[6] T230338-3 R3 PRED	_be_v_id	[5] E230338-1 R2 PRED	_be_v_id
[6] T230338-3 R3 LBL	h8	[7] E230338-1 R3 LBL	h8
[6] T230338-3 R3 ARG0	e9	[5] E230338-1 R2 ARG0	e6
[6] T230338-3 R3 ARG1	x1	[7] E230338-1 R3 ARG1	h10
[6] T230338-3 R3 ARG2	x10	[5] E230338-1 R2 ARG2	x7
[7] T230338-3 x10 PERS	3	[6] E230338-1 x7 PERS	3
[8] T230338-3 R4 PRED	_doctor_n_1	[10] E230338-1 R5 PRED	_doctor_n_1
[8] T230338-3 R4 LBL	h11	[3] E230338-1 R0 LBL	h0
[8] T230338-3 R4 ARG0	x10	[4] E230338-1 R1 ARG0	x1
[9] T230338-3 R5 PRED	exist_q_rel	[4] E230338-1 R1 PRED	exist_q_rel
[9] T230338-3 R5 LBL	h12	[9] E230338-1 R4 LBL	h11
[9] T230338-3 R5 ARG0	x10	[9] E230338-1 R4 ARG0	x7
[9] T230338-3 R5 RSTR	h13	[9] E230338-1 R4 RSTR	h12
[9] T230338-3 R5 BODY	h14	[9] E230338-1 R4 BODY	h13
[10] T230338-3 Q0 HARG	h3	[11] E230338-1 Q0 HARG	h3
[10] T230338-3 Q0 LARG	h0	[11] E230338-1 Q0 LARG	h0
[11] T230338-3 Q1 HARG	h7	[12] E230338-1 Q1 HARG	h10
[11] T230338-3 Q1 LARG	h8	[11] E230338-1 Q0 LARG	h0
[12] T230338-3 Q2 HARG	h13	[12] E230338-1 Q1 HARG	h10
[12] T230338-3 Q2 LARG	h11	[13] E230338-1 Q2 LARG	h14

Evaluation

- This technique quickly outpaced the ability of the Thai grammar to challenge its merits.
  - The limited competency of the Thai grammar means it generates few derivations for the sentences it does parse.
  - Thu, evaluation of this work became hampered by insufficient stress.
  - This is a good thing; SVD shows promise for bootstrapping complex models.

Applicability

- This work is mostly applicable to grammars that have significantly developed past ‘toy’ status
  - because off-the-shelf ‘Matrix’ grammars constrain ambiguity pretty well
  - Ambiguity-generating extensions in the Thai grammar include:
    - verb serialization which is handled as asyndetic coordination
    - subject or pronoun drop

Future work

- Extend the Thai grammar so that this bootstrapping method can face realistic challenges
- Evaluate alternative VSM distance interpretations
- Better understanding of the linear algebra which underlies this embedding

References

- Emily M Bender, Dan Flickinger and Stephan Oepen. (2002). The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In P*roceedings of the 2002 workshop on Grammar engineering and evaluation-Volume 15* (pp. 1-7). Association for Computational Linguistics.
- Ann Copestake, Dan Flickinger, Carl Pollard and Ivan A. Sag. (2005). Minimal recursion semantics: An introduction. *Research on Language and Computation*, 3(2-3), 281-332.
- Rebecca Dridan and Stephan Oepen. (2011). Parser evaluation using elementary dependency matching. In *Proceedings of the 12th International Conference on Parsing Technologies* (pp. 225-230). Association for Computational Linguistics.
- Dan Flickinger. (2000). On building a more efficient grammar by exploiting types. Natural Language Engineering, 6(1), 15-28.
- S. Fujita, Francis Bond, Stephan Oepen and T. Tanaka. (2010). Exploiting semantic information for HPSG parse selection. *Research on Language and Computation*, 8(1), 1-22.

References, cont.

- Gene H. Golub and Charles F. Van Loan. (1996). *Matrix computations (3rd ed.)*. Johns Hopkins University Press, Baltimore, MD, USA.
- April Kontostathis and William M. Pottenger. (2002). *Transitivity and the co-occurrence relation in LSI*. Technical Report LU-CSE-02-005, Lehigh University.
- Stephan Oepen, Dan Flickinger, Kristina Toutanova, and Christoper D. Manning. (2004). LinGO Redwoods: A Rich and Dynamic Treebank for HPSG. *Research on Language and Computation* 2(4):575-596.
- Stephan Oepen and J. T. Lønning. (2006). Discriminant-based MRS banking. In Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006).
- Hinrich Schütze. (1992). Dimensions of meaning. In *Proceedings of Supercomputing*.
- Glenn Slayden. (2010) *Array TFS storage for unification grammars*. University of Washington Master's Thesis, 2010.
- Kristina Toutanova and Christopher D. Manning (2002). Feature Selection for a Rich HPSG Grammar Using Decision Trees. *In Proceedings of the Sixth Conference on Natural Language Learning (CoNLL 2002)*, Taipei, Taiwan.

Thank you!