Semantic Operations for Automatically Extracting Transfer Rules for Machine Translation

Michael Wayne Goodman

February 21, 2018

Department of Linguistics

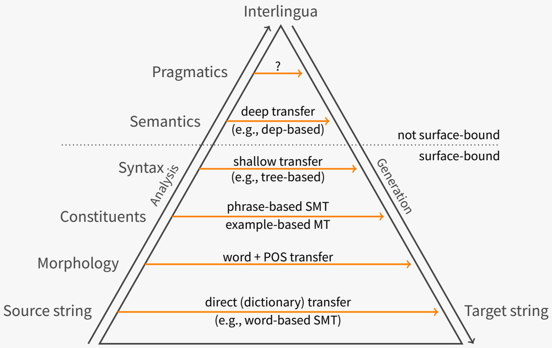
UNIVERSITY *of* WASHINGTON

Outline

Introduction

Introduction

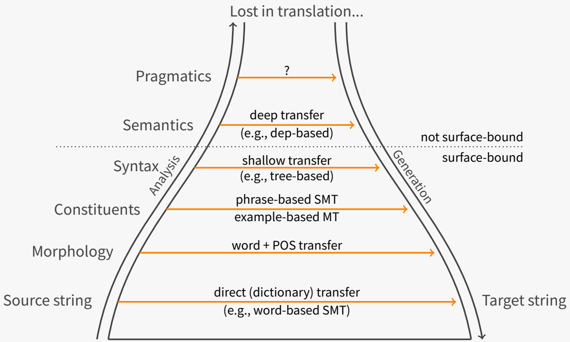
- Translation by Semantic transfer
  - Why semantics?
    - Goal of translation: another language, same meaning
    - Words, syntax, etc. are further removed from meaning
    - Semantics is not bound by surface order
  - Why transfer?
    - For that matter, what is transfer?
    - Semantic transfer; general-purpose interlinguas don't really exist



Vauquois Triangle

Introduction

- Translation by Semantic transfer
  - Why semantics?
    - Goal of translation: another language, same meaning
    - Words, syntax, etc. are further removed from meaning
    - Semantics is not bound by surface order
  - Why transfer?
    - For that matter, what is transfer?
    - Semantic transfer; general-purpose interlinguas don't really exist



Copestake Volcano
AKA Ikehara Discontinuity

AKA Vauquois inverted funnel with a very long spout

Introduction : Machine Translation

- Many paradigms of MT, e.g., RBMT, EBMT, SMT, NMT
- I work with RBMT—the LOGON transfer machinery (Lønning et al., 2004; Bond et al., 2011)
- RBMT is often considered hand-built, while EBMT, SMT, and NMT are data-driven
- At the core, there's not a big difference between modern data-driven transfer, EBMT, and phrase-based SMT (transfer rule store $\approx$ fragment base $\approx$ phrase table)

Introduction : Minimal Recursion Semantics

- Chosen semantic representation: MRS (Copestake et al., 2005)
- Represents the surface string; intentionally noncommittal toward unknown information
- Accounts for ambiguity with underspecification (lexical, scopal, ...)

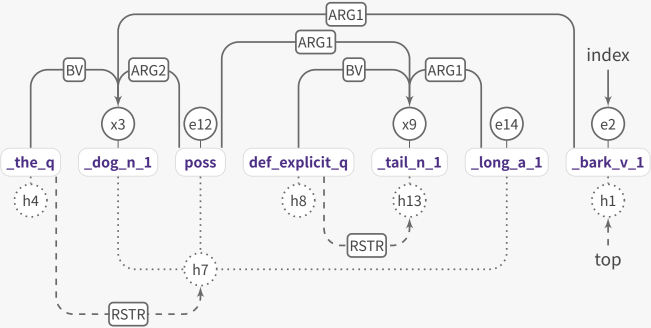
Example

(1)

長い	尻尾	の	犬	が	吠え	た。
nagai	shippo	no	inu	-ga	hoe	-ta
long	tail	GEN	dog	-NOM	bark	-PERF

"The dog whose tail is long barked." [jpn]

"The dog whose tail is long barked."



Semantic Operations

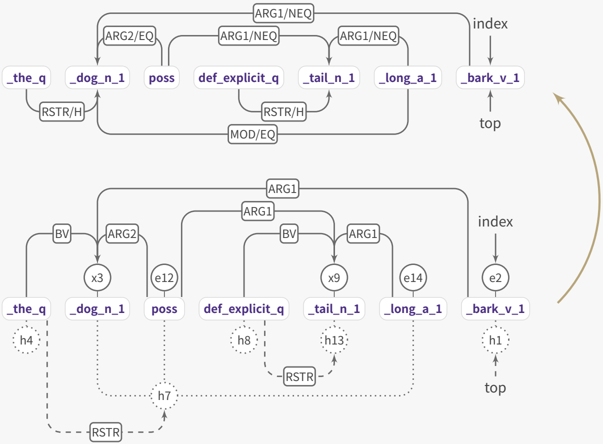
Semantic Operations

- MRS $\leftrightarrow$ DMRS conversion
- Semantic Tests
  - Connectedness
  - Isomorphism
  - Headedness
- PENMAN serialization
- DMRS simplifications
- Note: mostly not my invention, but my implementation

Semantic Operations : MRS to DMRS Conversion

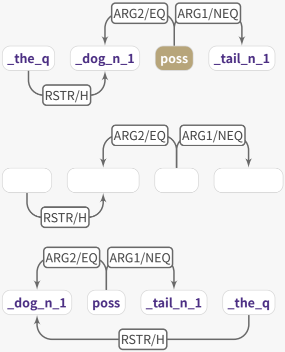
- Copestake 2009
- Variable-free
- EPs → nodes
- Arguments, qeqs → links
- Representative nodes allow MRS hypergraph to be DMRS DAG
- Leaner structure is easier to model
- (Near-)Lossless conversion with MRS

“The dog whose tail is long barked.”



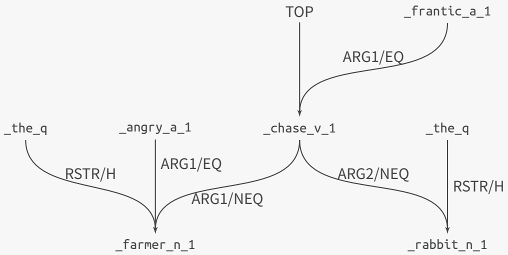
Semantic Operations : Tests

- Connectedness (MRS vs DMRS)
- Isomorphism
  - Full
  - Structural
  - Bilingual
- Headedness

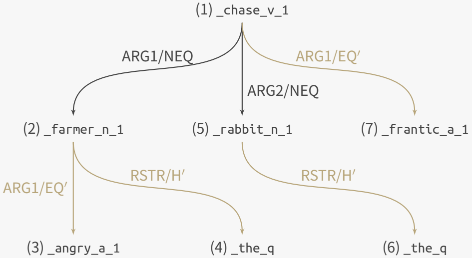


"The angry farmer frantically chased the rabbit."

Original orientation



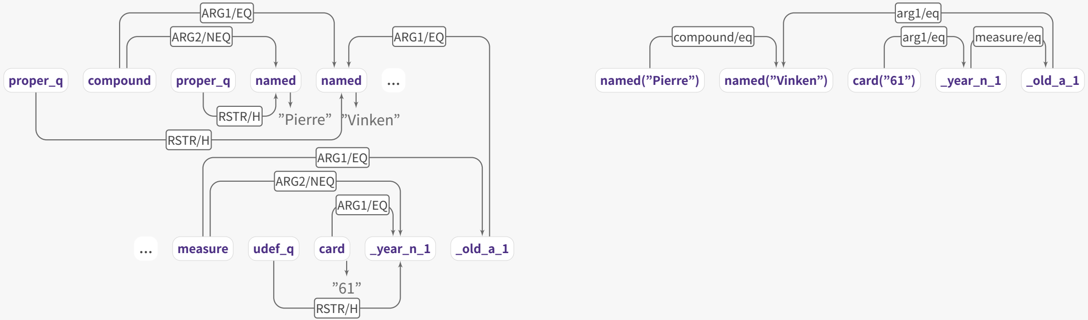
Headed Orientation



Semantic Operations : PENMAN Serialization

```
(e0 / _chase_v_1
    :ARG1-NEQ (x1 / _farmer_n_1
                 :ARG1-EQ-of (e2 / _angry_a_1)
                 :RSTR-H-of (u4 / _the_q))
    :ARG2-NEQ (x5 / _rabbit_n_1
                 :RSTR-H-of (u6 / _the_q))
    :ARG1-EQ-of (e7 / _frantic_a_1))
```

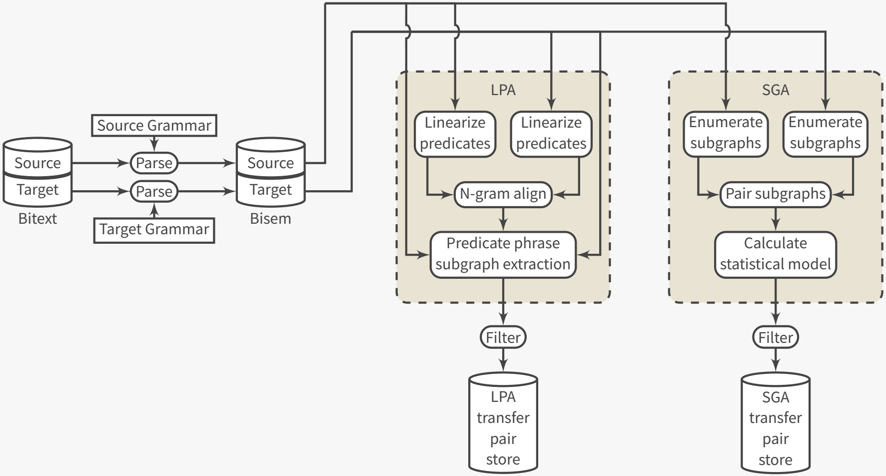
Semantic Operations : DMRS Simplification



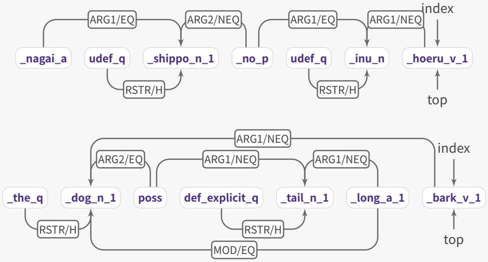
- Remove predictable nodes (e.g., default quantifiers)
- Merge predicates with constants
- Change 2-argument nodes to edges
- (Copestake et al., 2016; Ivanova et al., 2012; Oepen et al., 2015)

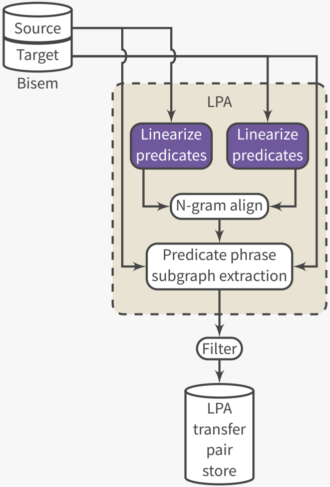
Bilingual Subgraph Alignment

Bilingual Subgraph Alignment Methods



LPA : Predicate Phrase Projection : Linearization






```
_nagai_a udef_q _shippo_n_1 _no_p udef_q _inu_n _hoeru_v_1
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
_the_q _dog_n_1 poss def_explicit_q _tail_n_1 _long_a_1 _bark_v_1
```

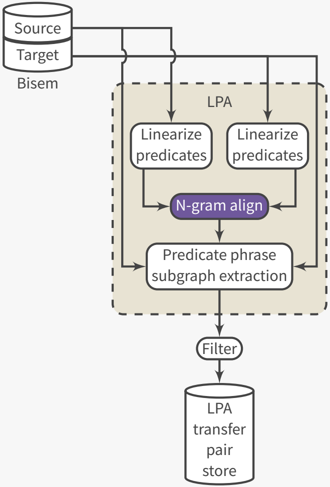
- Predicates taken in original (surface) order
- Predictable predicates (e.g., default quantifiers) are noise to the aligner; remove them

LPA : Predicate Phrase Projection : N-gram Alignment

```
_nagai_a _shippo_n_1 _no_p _inu_n _hoeru_v_1
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
_the_q _dog_n_1 poss def_explicit_q _tail_n_1 _long_a_1 _bark_v_1
```

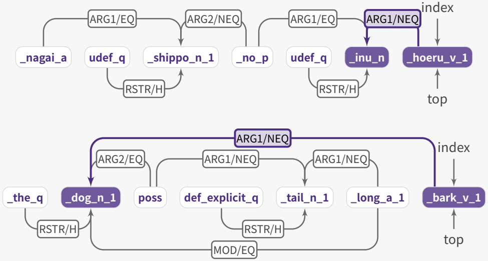
- Anymalign (Lardilleux et al., 2012)

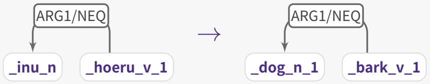
Source		Target	$P(t\|s)$	$P(s\|t)$
_nagai_a	→	_long_a_1	0.874	0.814
_shippo_n_1	→	_tail_n_1	0.459	0.041
_inu_n	→	_dog_n_1	0.949	0.944
_hoeru_v_1	→	_bark_v_1	0.866	0.628
_nagai_a _shippo_n_1	→	_tail_n_1 _long_a_1	0.953	0.877
_inu_n _hoeru_v_1	→	_long_a_1 _bark_v_1	0.023	0.003
_inu_n _hoeru_v_1	→	_dog_n_1 _bark_v_1	0.637	0.755
...	→	...		



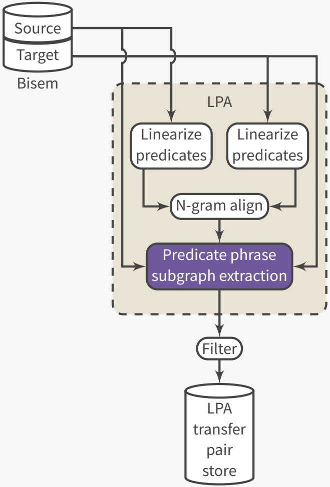
LPA : Predicate Phrase Projection : Projection and Extraction

_inu_n _hoeru_v_1 → _dog_n_1 _bark_v_1 0.637 0.755





0.637 0.755



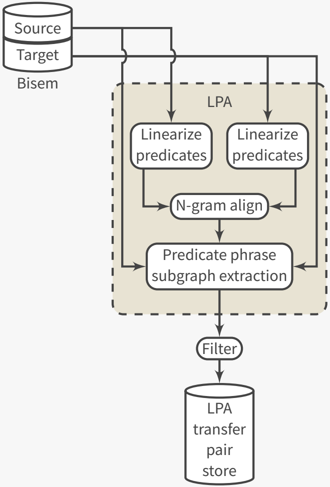
LPA : Predicate Phrase Projection : Summary

Translation pair extraction via pred phrase alignments

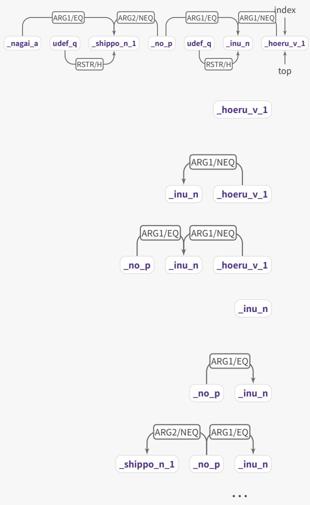
- Predicate phrases and probabilities from aligner
- Filtered by graph properties

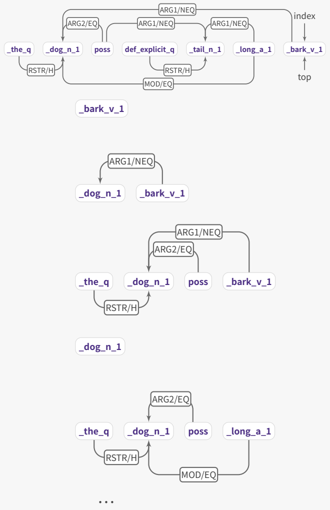
Compare: Haugereid and Bond 2011, 2012

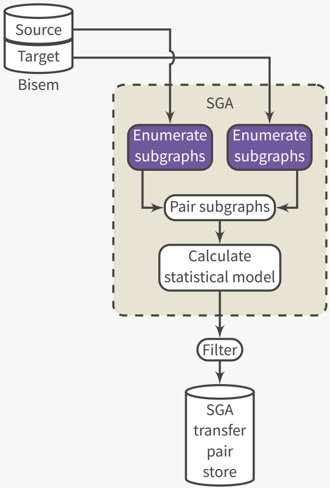
- Extracts when predicate phrases match templates
- MRS structure stored in rules, not extracted from training data
- Results in more constrained rules
- Templates are a resource that must be maintained WRT two grammars



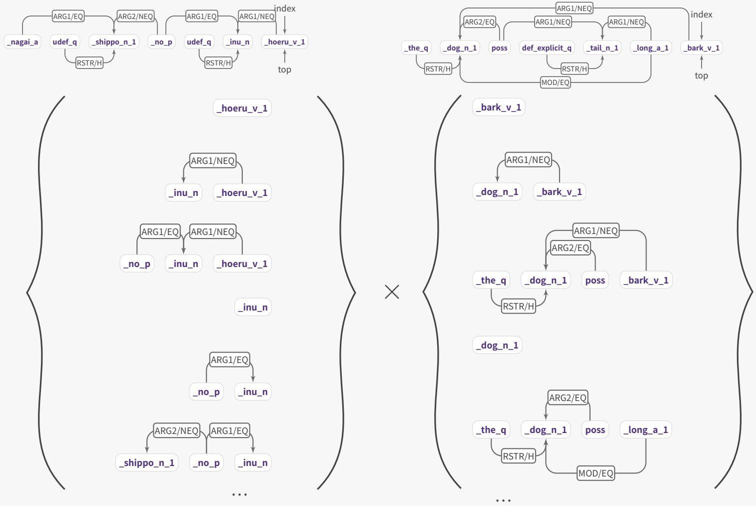
SGA : Subgraph Enumeration : Enumeration

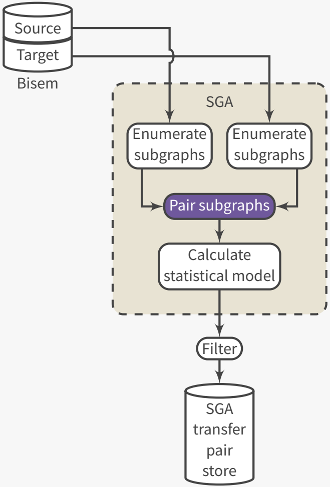






SGA : Subgraph Enumeration : Pairing



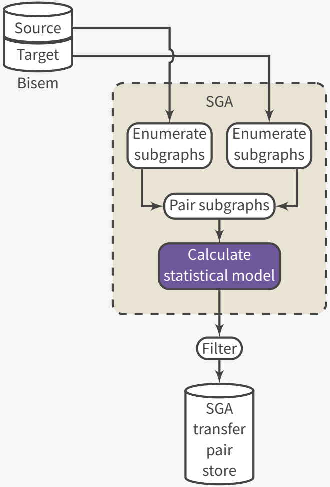


SGA : Subgraph Enumeration : Filtering

- Count subgraph pairs to get corpus statistics

Subgraph pair count (src,tgt)	$X_{s,t}$
Marginal pair counts	X_s, X_t
Count of all subgraphs	$N = \sum_{s,t} X_{s,t}$
Forward translation probability	$P(t\|s) = \frac{P(s,t)}{P(s)} = \frac{X_{s,t}}{X_s}$
Backward translation probability	$P(s\|t) = \frac{P(s,t)}{P(t)} = \frac{X_{s,t}}{X_t}$
Symmetric translation probability	$P(t\|s) * P(s\|t)$

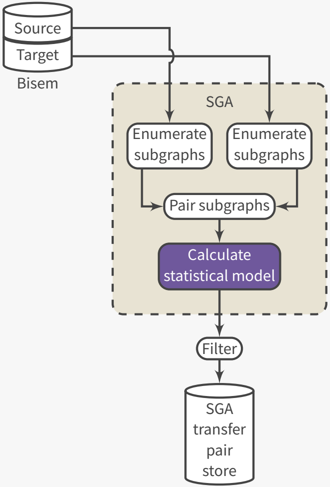
- ...but filter unlikely pairs before calculating probabilities



SGA : Subgraph Enumeration : ϕ^2 (Church and Gale, 1991)

	target	¬target	total
source	$a = X_{s,t}$	b	$a + b = X_s$
¬source	c	d	$c + d = N - X_s$
total	$a + c = X_t$	$b + d = N - X_t$	N

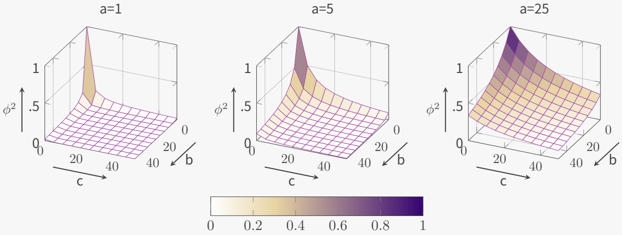
$$\phi^2 = \frac{\chi^2}{N} = \frac{(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

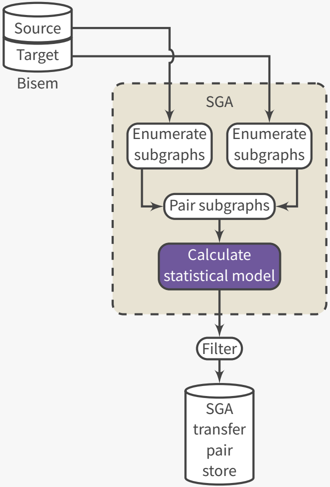


SGA : Subgraph Enumeration : ϕ^2 (Church and Gale, 1991)

	target	¬target	total
source	$a = X_{s,t}$	b	$a + b = X_s$
¬source	c	d	$c + d = N - X_s$
total	$a + c = X_t$	$b + d = N - X_t$	N

$$\phi^2 = \frac{\chi^2}{N} = \frac{(ad - bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

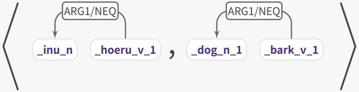




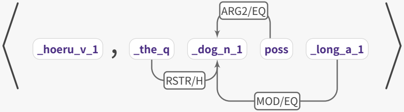
SGA : Subgraph Enumeration : Weighted ϕ^2

- Very rare predicates can be useful for rounding out the model, but ϕ^2 gives equal weight to all subgraphs with the rare pred
- Generally, balanced translations are preferred, so I discount ϕ^2 score by difference in graph order

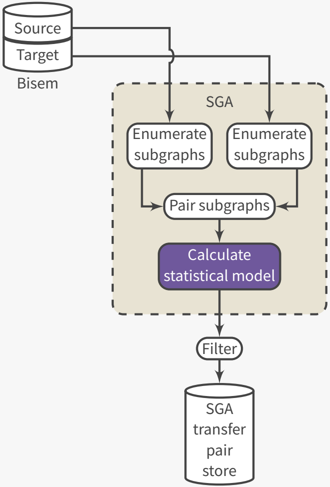
$$W_{s,t} = \frac{1}{||V(s)|-|V(t)||+1}$$



$$w = \frac{1}{|2-2|+1} = 1$$



$$w = \frac{1}{|1-4|+1} = 0.25$$



SGA : Subgraph Enumeration : Summary

Translation pair enumeration and modeling

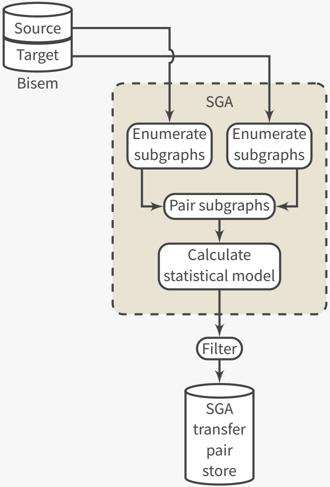
- Graph-native bilingual alignment
- Prefilters to avoid bad graphs
- Weighted ϕ^2 filters to avoid bad translations
- Calculate probabilities to rank pairs

Compare: Jellinghaus 2007

- Translation pairs by top-down traversal
- Traversal done on both sides synchronously

Compare: Hearne and Way 2003; Graham et al. 2009

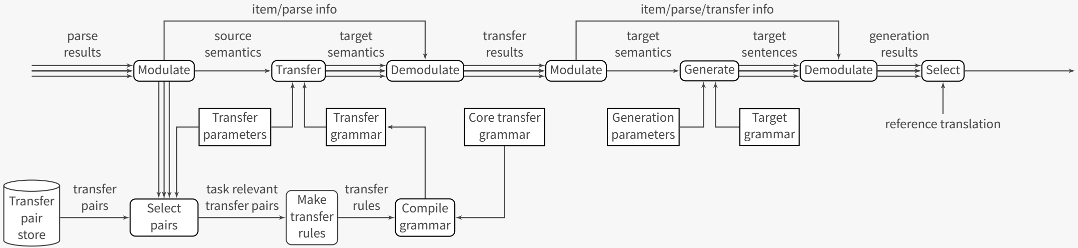
- Enumeration of LFG f-structure fragments
- Fragment base is consulted during translation (EBMT)



Experimental Design

Experimental Design : Tasks

- Translate
- For JaEn-based systems:
  - Parse items (done once for all systems)
  - Transfer with each system
  - Generate for each system's outputs

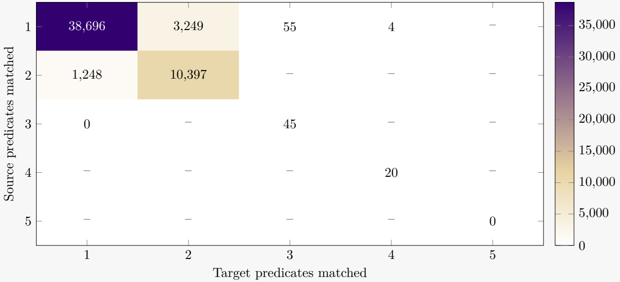


Experimental Design : Baselines

Two baselines

- Moses (see paper for settings)
- H&B: JaEn with Haugereid and Bond (2011, 2012) rules

H&B rule distribution by matched predicates*



* Subgraph order may be higher due to rule types introducing predicates.

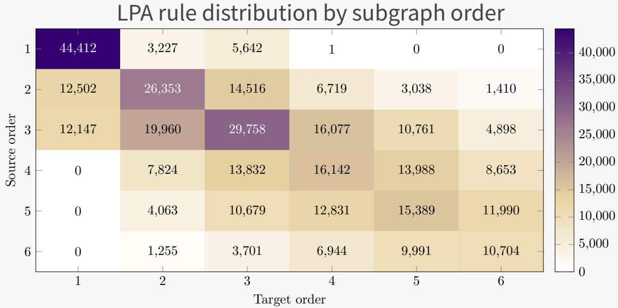
Experimental Design : Configurations

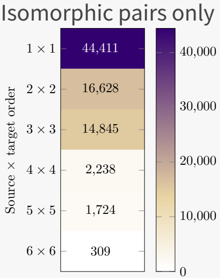
Two experimental systems

- 2 rule sets in transfer grammars
  - Single rules: order $1 \times 1\text{–}n$
  - MWE rules: order $2\text{–}n \times 1\text{–}n$
- 21 configurations for each system

S	$(n = \{1 \ldots 6\})$	my single rules, H&B MWE rules
M	$(n = \{2 \ldots 6\})$	my MWE rules, H&B single rules
P	$(n = \{2 \ldots 6\})$	my single and MWE rules
O	$(n = \{2 \ldots 6\})$	my single and MWE rules, only isomorphic

Experimental Design : LPA





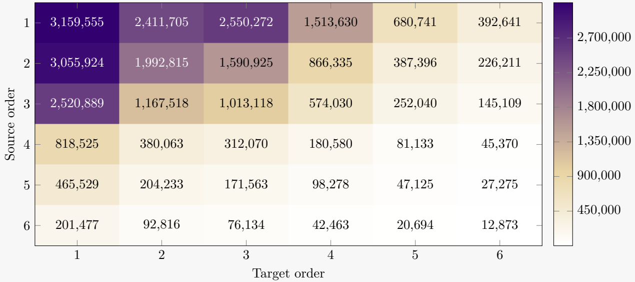
- No 1×4–6, so LPA only has 18 configurations

Experimental Design : LPA

✓	`(x0 / _ラクトース_n_unknown)`	`(x0 / _lactose_n_1)`
✓	`(e0 / _chakuriku_s_1 :ARG1-EQ-of (e1 / _ni_p :ARG2-NEQ (x2 / named :carg "isutanbu-ru_1")))`	`(e0 / _land_v_1 :ARG3-H (e1 / _in_p :ARG2-NEQ (x2 / named :carg "Istanbul")))`
✗	`(e0 / _hatsumei_s_1 :ARG1-NEQ (x1 / _beru_n_1) :ARG2-NEQ (x2 / nominalization :ARG1-HEQ (e3 / _denwa_s_1)))`	`(e0 / _invent_v_1 :ARG1-NEQ (x1 / named :carg "Bell"))`

Experimental Design : SGA

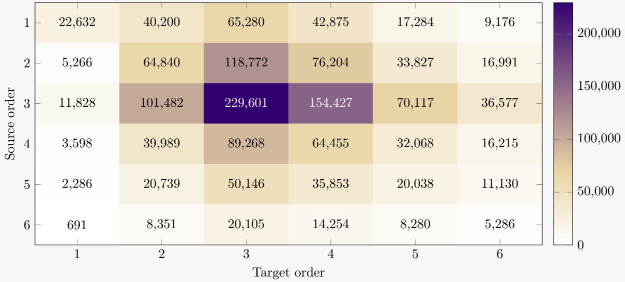
SGA rule distribution by subgraph order; all rules



- Including all pairs
- Smaller subgraphs are repeated (and paired) more frequently

Experimental Design : SGA

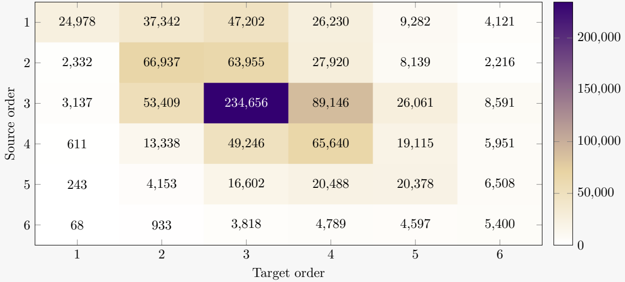
SGA rule distribution by subgraph order; ϕ^2-filtered



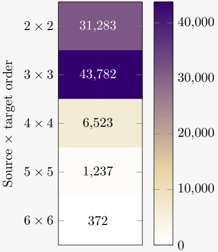
- Including pairs after ϕ^2 filtering and symmetric probability filtering
- Unlike LPA's diagonal trend, SGA radiates
- Weighted ϕ^2 filtering (next slide) changes probabilistic model, affecting the distribution

Experimental Design : SGA

SGA rule distribution by subgraph order; weighted-ϕ^2-filtered



Isomorphic pairs only



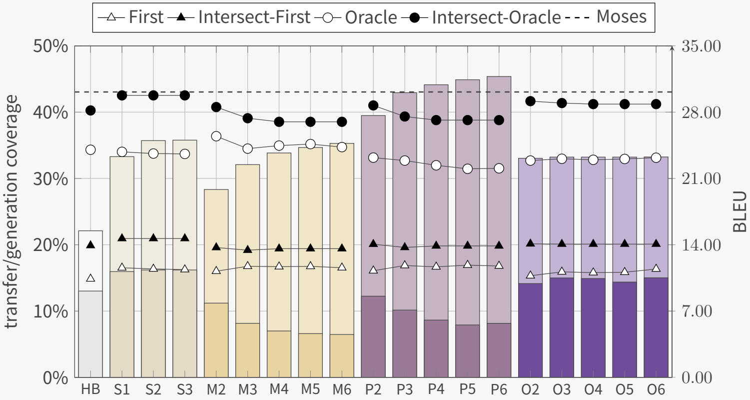
Experimental Design : SGA

✓	`(x0 / nominalization`
	`    :ARG1-HEQ (e1 / _houkoku_s_4)`
	`    :ARG1-EQ-of (e2 / _shoujiki_a_3))`
		`(x0 / nominalization`
		`    :ARG1-HEQ (e1 / _report_v_to)`
		`    :ARG1-EQ-of (e2 / _honest_a_1))`

✓	`(x0 / _gengo_n_1`
	`    :ARG1-EQ-of (e1 / compound`
	`        :ARG2-NEQ (x2 / _shizen_n)))`
		`(x0 / _language_n_1`
		`    :ARG1-EQ-of (e1 / _natural_a_for))`

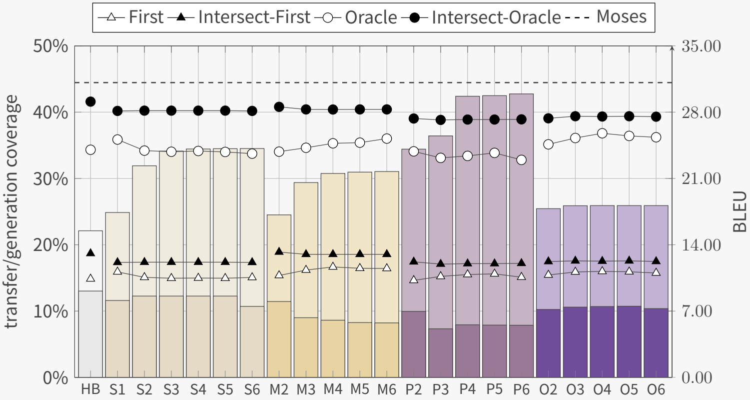
✗	`(x0 / _kureetaa_n`
	`    :ARG1-EQ-of (e1 / _no_p`
	`        :ARG2-NEQ (x2 / _kazan_n_1))`
	`    :ARG1-EQ-of (e3 / _kyodai_a_2))`
		`(x0 / _volcano_n_1`
		`    :ARG1-EQ-of (e1 / _collapse_v_1`
		`        :ARG1-EQ-of (e2 / _into_p))`
		`    :RSTR-H-of (u3 / _a_q))`

Results and Analysis



Light bar=transfer coverage; Dark bar=generation coverage

Results : SGA : Development



Light bar=transfer coverage; Dark bar=generation coverage

Results : LPA and SGA : Test

	BLEU		NIST		METEOR	
System	First	Oracle	First	Oracle	First	Oracle
Moses	**35.05**		**5.95**		**35.80**	
H&B	13.12	**30.40**	3.83	**5.56**	29.74	**35.77**
LPA-O2	11.68	24.86	3.48	4.94	28.26	32.72
SGA-O5	**12.58**	**29.74**	**3.80**	**5.47**	**29.31**	**34.83**

Results : Examples

Example

Japanese	ジムは鍵を回した。
Reference	Jim turned the key in the lock.
Moses	Jim pass me the key.
H&B	Jim turned the key.
LPA-S2	Jim passed on the key to oneself.
LPA-M2	Jim turned the key.
LPA-P2	Jim passed on the key to myself.
LPA-O2	Jim passed on the key to yourselves.

Results : Examples

Example

Japanese	彼はテレビのニュースキャスターとして働いている。
Reference	He works as a newscaster in television.
Moses	He worked as a tv ニュースキャスター.
H&B	He is working as a television anchorman.
LPA-O2	He is working as anchormans on a television.
SGA-O5	He is working as a television anchorman.

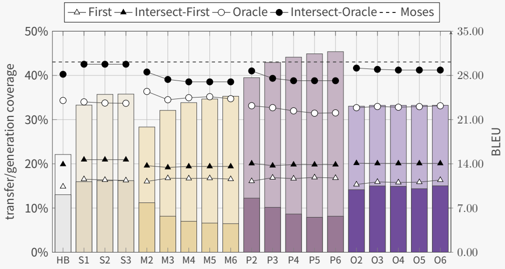
Results : Examples

Example

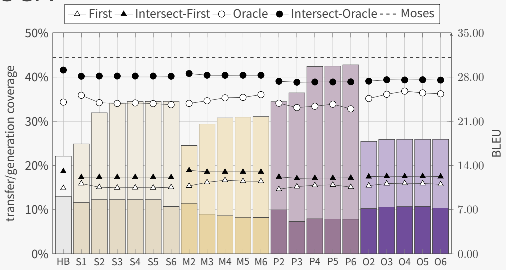
Japanese	彼らは音楽を聞いていませんでした。
Reference	They were not listening to music.
Moses	They were listening to music.
H&B	They were not hearing the music.
SGA-S2	They were not hearing the pieces of music.
SGA-M2	They were not hearing the music.
SGA-P5	They were not hearing the pieces of music.
SGA-O2	They were not hearing the pieces of music.

Analysis : Coverage and Quality

LPA



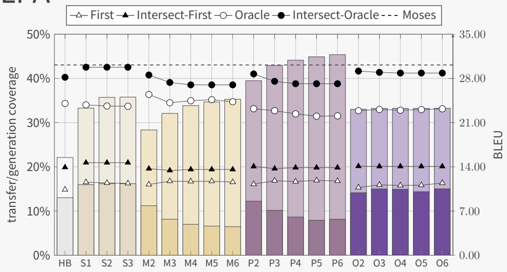
SGA



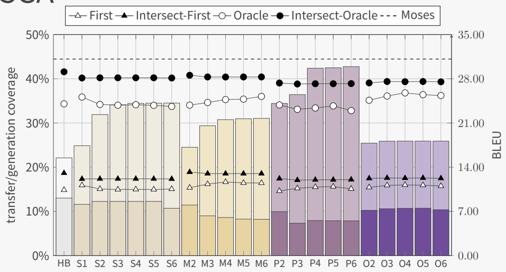
- Increasing subgraph order
  - Increases transfer coverage (S, M, P)
  - No effect on generation coverage (S and O)
  - Decreases generation coverage (M and P)
- BLEU scores change little across configurations

Analysis : Other Factors

LPA



SGA



- Timeouts (LPA > SGA)
- Memouts (H&B, M > S, P, O)
- Semantic errors (M, P > H&B, S, O)
- Incomplete transfer (SGA > LPA, M > S, P, O)
- (Quasi-)Lexical gaps on generation (S, P, O > H&B, M; `thing`, `manner`, etc.)

Analysis : Subgraph Topologies

Order	Number	Examples
1	$C_0 = 1$	()
2	$C_1 = 1$	(())
3	$C_2 = 2$	(()()), ((()))
4	$C_3 = 5$	(()()()), (()(())), ((())()), ((()())), (((())))
5	$C_4 = 14$	...
6	$C_5 = 42$	...

- Including roles, there are more than 300 observed patterns at order 3
- And more yet with node identifiers

Data	% (()())	% ((()))
LPA-Jacy	21.2	78.8
LPA-ERG	24.8	75.2
SGA-Jacy	39.7	60.3
SGA-ERG	68.1	31.9

```
(e0 / _across_p_dir
    :ARG2-NEQ (x1 / _river_n_of
        :RSTR-H-of (u2 / _the_q)))
```

```
(e0 / _send_v_for
    :ARG1-NEQ (x1 / pron)
    :ARG2-NEQ (x2 / _doctor_n_1))
```

Conclusion

Contributions : Scientific / Methodological

- Two new methods for automatic transfer rule creation
  - Easier to keep up-to-date with source/target grammars
- Rule creation without templates
  - Improved language independence
  - Quality comparable with state of the art
- New test for semantics debugging: headed-traversal
  - Identifies unlikely analyses automatically
  - Enables singly-rooted graphs, e.g., for PENMAN serialization

Contributions : Engineering / Artifactual

- Introductions:
  - **PyDelphin** semantics library (+α) (https://github.com/delph-in/pydelphin)
  - **Penman** graph library (https://github.com/goodmami/penman)
  - **gTest** grammar testing utility (https://github.com/goodmami/gtest)
  - **Bottlenose** web API and server (https://github.com/delph-in/bottlenose)
  - **Demophin** web demo (https://github.com/goodmami/demophin)
  - **Delphin-Viz** visualizations (https://github.com/delph-in/delphin-viz)
- Improvements:
  - **ACE** processor (http://sweaglesw.com/linguistics/ace/)
  - **Jacy** Japanese grammar (https://github.com/delph-in/jacy)
  - **ERG** English grammar (http://www.delph-in.net/erg/)
  - **JaEn** transfer grammar (https://github.com/delph-in/JaEn)
  - **NLTK** library (http://www.nltk.org/)
  - **DELPH-IN wiki** documentation (http://moin.delph-in.net/)

References

Francis Bond, Stephan Oepen, Eric Nichols, Dan Flickinger, Erik Velldal, and Petter Haugereid. Deep open-source machine translation. Machine Translation, 25(2):87–105, 2011.

Kenneth W Church and William A Gale. Concordances for parallel text. In Proceedings of the Seventh Annual Conference of the UW Centre for the New OED and Text Research, pages 40–62, 1991.

Ann Copestake. Slacker semantics: why superficiality, dependency and avoidance of commitment can be the right way to go. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, pages 1–9. Association for Computational Linguistics, 2009.

Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A. Sag. Minimal Recursion Semantics. An introduction. Research on Language and Computation, 3(4):281–332, 2005.

Ann A Copestake, Guy Emerson, Michael Wayne Goodman, Matic Horvat, Alexander Kuhnle, and Ewa Muszynska. Resources for building applications with Dependency Minimal Recursion Semantics. In LREC, 2016.

Yvette Graham, Josef van Genabith, Anton Bryl, Miriam Butt, and Tracy Holloway King. F-structure transfer-based statistical machine translation. Proceedings of LFG09, pages 317–337, 2009.

Petter Haugereid and Francis Bond. Extracting transfer rules for multiword expressions from parallel corpora. In Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World, pages 92–100, Portland, Oregon, USA, June 2011. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/W11-0814`.

Petter Haugereid and Francis Bond. Extracting semantic transfer rules from parallel corpora with smt phrase aligners. In Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation, pages 67–75, Jeju, Republic of Korea, July 2012. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W12-4208.

Mary Hearne and Andy Way. Seeing the wood for the trees: Data-Oriented Translation. In In Machine Translation Summit IX. Citeseer, 2003.

Angelina Ivanova, Stephan Oepen, Lilja Øvrelid, and Dan Flickinger. Who did what to whom? a contrastive study of syntacto-semantic dependencies. In Proceedings of the Sixth Linguistic Annotation Workshop, pages 2–11, Jeju, Republic of Korea, July 2012. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W12-3602.

Michael Jellinghaus. Automatic acquisition of semantic transfer rules for machine translation. Master's thesis, Universität des Saarlandes, 2007.

Adrien Lardilleux, François Yvon, and Yves Lepage. Hierarchical sub-sentential alignment with Anymalign. In 16th annual conference of the European Association for Machine Translation (EAMT 2012), pages 279–286, 2012.

Jan Tore Lønning, Stephan Oepen, Dorothee Beermann, Lars Hellan, John Carroll, Helge Dyvik, Dan Flickinger, Janne Bondi Johannessen, Paul Meurer, Torbjørn Nordgård, et al. LOGON. a norwegian MT effort. In Proceedings of the Workshop in Recent Advances in Scandinavian Machine Translation, volume 6, 2004.

Stephan Oepen, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Silvie Cinková, Dan Flickinger, Jan Hajic, and Zdenka Uresova. Semeval 2015 task 18: Broad-coverage semantic dependency parsing. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pages 915–926, 2015.

Thank you