NLP Tools
LVG - Derivations
(New SD-Rule Set)

By: Dr. Chris Lu

The Lexical Systems Group

NLM. LHNCBC. CGSB

Oct., 2014

Table of Contents

Natural Language Processing

- Map terms to concepts (meaning)
- Difficulties: many to many mapping:

Terms	Concepts
• cold	• Cold Temperature|C0009264 • Common Cold|C0009443 • Cold Therapy|C0010412 • Cold Sensation|C0234192 • etc.
• cold • Cold Temperature • Cold Temperatures • Cold (Temperature) • Temperatures, Cold • Low temperature • Temperature; low • low temperatures • etc.	• Cold Temperature|C0009264

NLP – Concept Mapping

- Normalization:
  - A term might have many different variations, such as inflectional variants, spelling variants, synonyms, abbreviations (expansions), cases, ASCII conversion, etc.
  - Normalize different forms of a concept to a same form

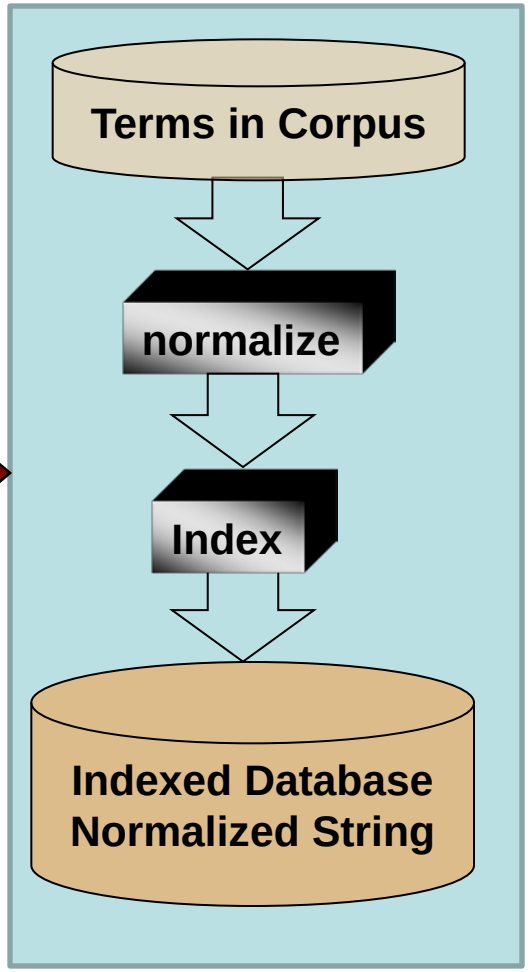
- Query Expansion:
  - Expand a term to its equal terms, such as subterm substitution of synonyms, derivational variants, spelling variants, abbreviations, etc.
  - Use to increase recall

NLP - Norm

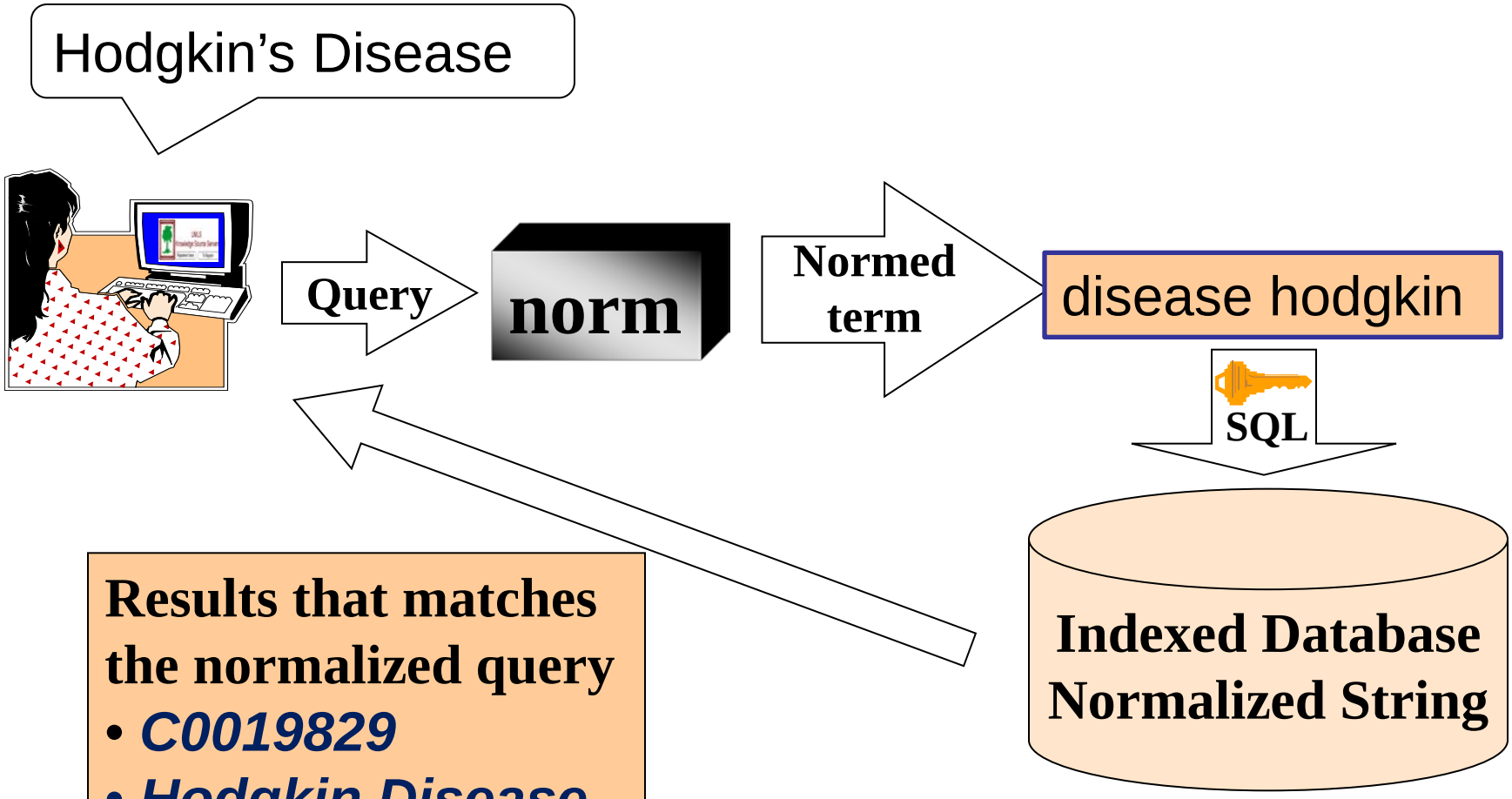
- Hodgkin Disease
- HODGKINS DISEASE
- Hodgkin's Disease
- Disease, Hodgkin's
- HODGKIN'S DISEASE
- Hodgkin's disease
- Hodgkins Disease
- Hodgkin's disease NOS
- Hodgkin's disease, NOS
- Disease, Hodgkins
- Diseases, Hodgkins
- Hodgkins Diseases
- Hodgkins disease
- hodgkin's disease
- Disease;Hodgkins
- Disease, Hodgkin
- …

disease hodgkin

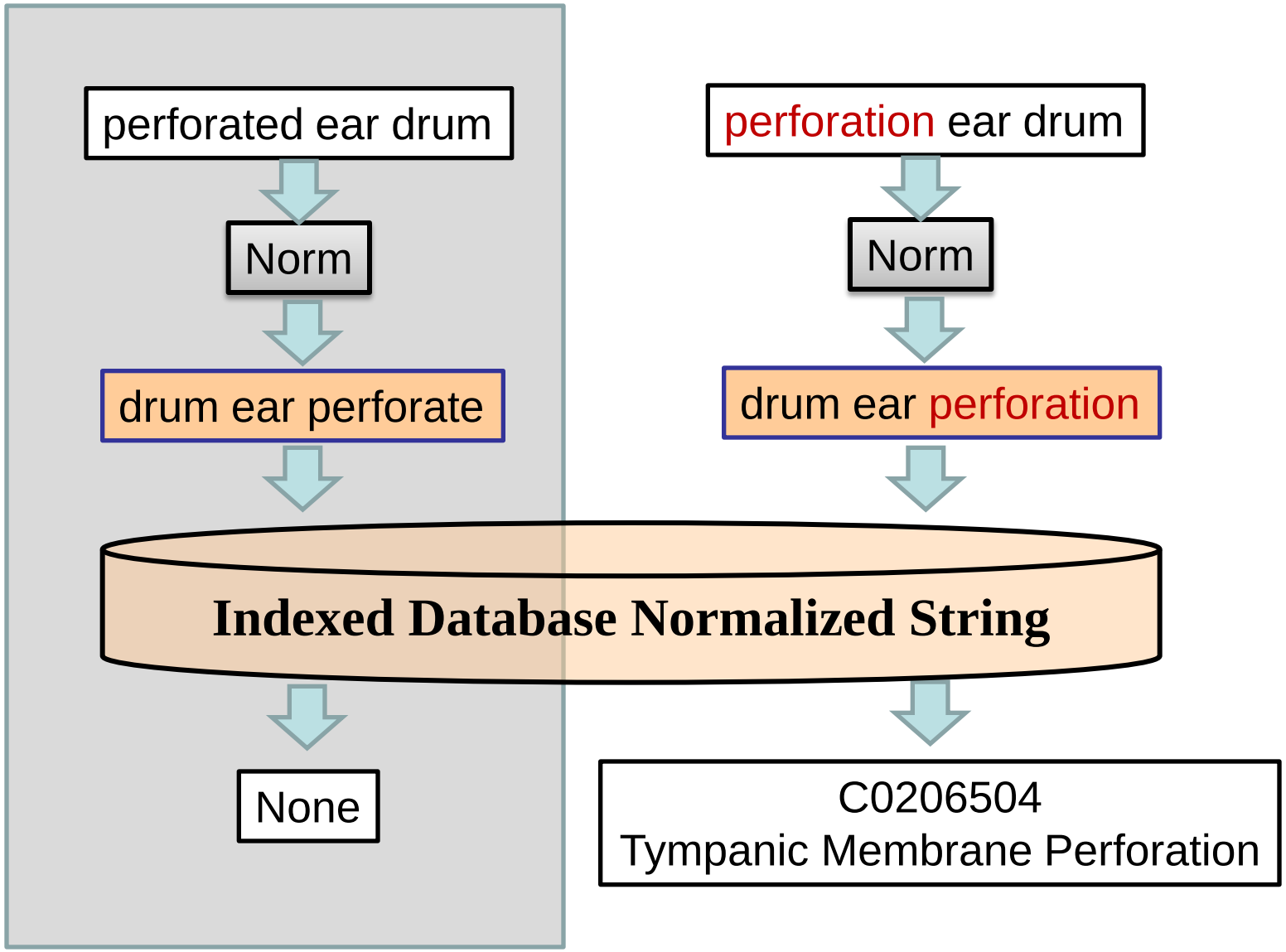
- C0019829
- Hodgkin Disease



NLP – Norm (Cont.)



NLP – Query Expansion



LVG - Lexical Variants Generation

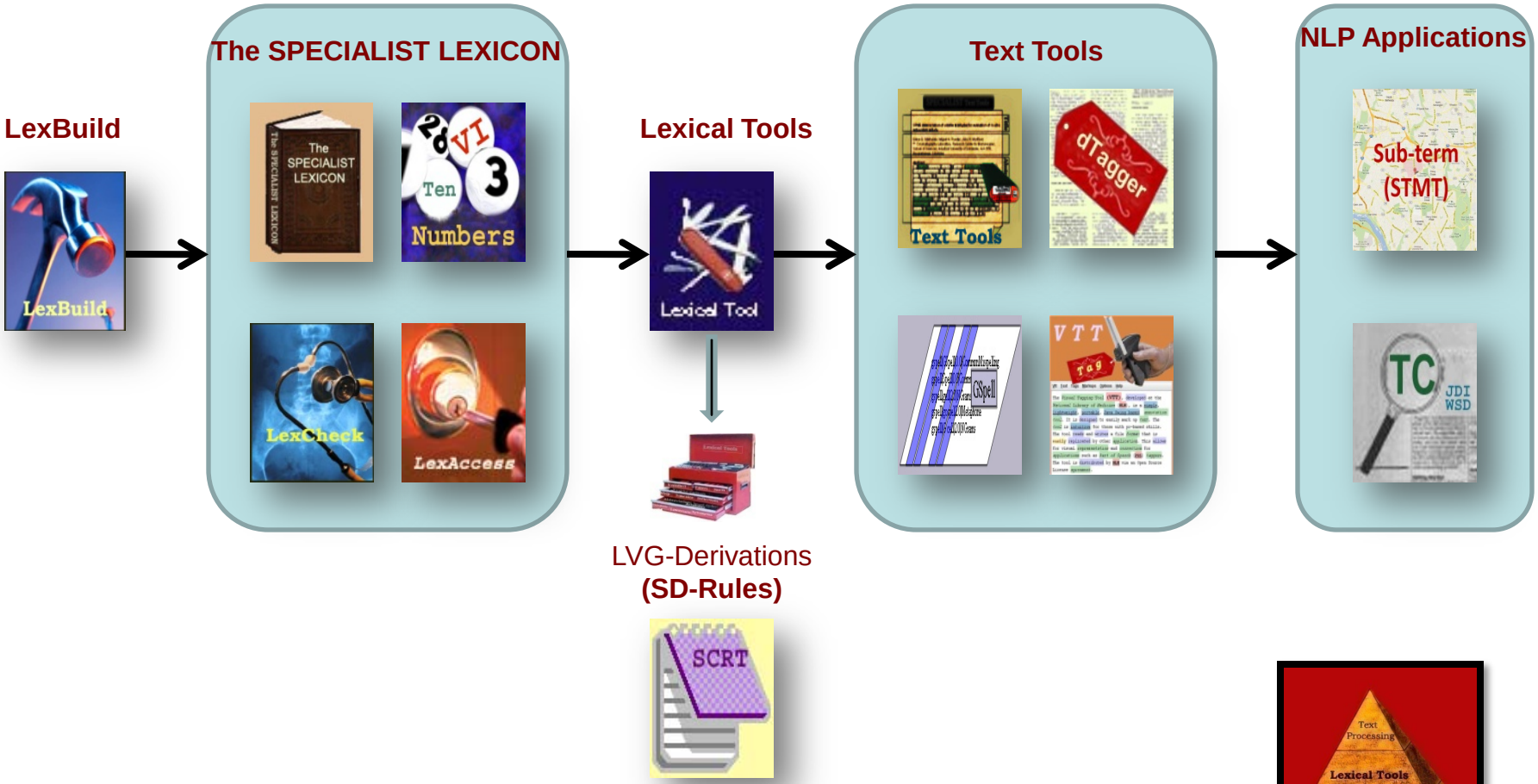
- To increase recall & precision

	Query Expansion (Recall)	POS Tagging (Precision)
Inputs	perforated ear drum	saw
UMLS-CUI	None	• C1947903|verb|see • C0183089|noun|saw (device)
Process	perforation ear drum	noun
UMLS-CUI	C0206504	• C0183089
Preferred Term	Tympanic Membrane Perforation	saw (device)

NLP Tools by LSG

- Lexical Systems Group: http://umlslex.nlm.nih.gov
- The SPECIALIST NLP Tools: http://specialist.nlm.nih.gov

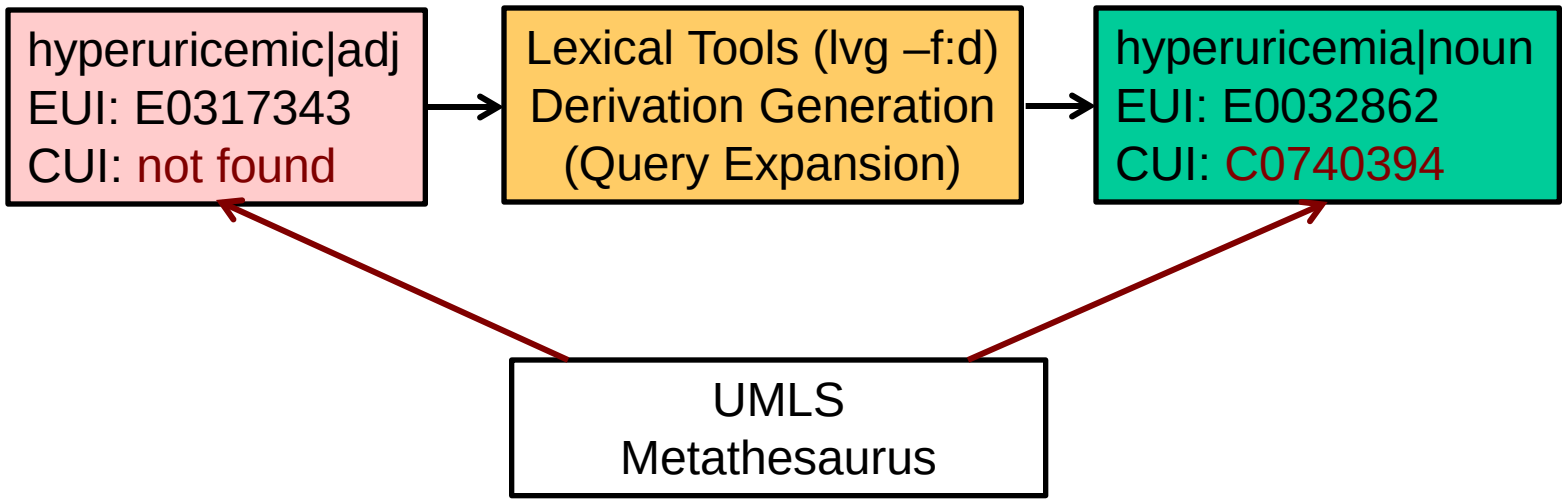
The SPECIALIST NLP Tools



- Lexical Systems Group: http://umlslex.nlm.nih.gov
- The SPECIALIST NLP Tools: http://specialist.nlm.nih.gov

Derivations in NLP Application

- hyperuricemic|adj, E0317343, no CUI
- hyperuricemia|noun, E0032862,
  is a UMLS Metathesaurus term (C0740394)



What are Derivational Variants?

- Words are related by a derivational process
  - Used to create new words based on existing words
  - Derivational process: suffix, prefix, and conversion
  - Meaning change (related)
  - Category may change

- Focus on relatedness (no direction)

Derivation Types (-kdt)

- Example (kind|adj):
  - zeroD: kind|adj|kind|noun
  - prefixD: kind|adj|unkind|adj
  - suffixD: kind|adj|kindly|adv

Derivational Pair

- Each link and the associated two nodes in derivational network define a derivational pair
- Includes base forms and syntactic category information
- Bi-directional
- **Only involves one or none derivational affix**
- Lvg format: base 1|category 1|base 2|category 2
- Examples:
  - kind|adj|kindness|noun
  - kind|adj|kindly|adv
  - kind|adj|unkind|adj
  - kind|adj|kind|noun
  - **kind|adj|unkindly|adv (not a dPair)**

Derivations in LVG

- 7 flow components (62):
  - -f:d
  - -f:dc
  - -f:R
  - -f:G
  - -f:Ge
  - -f:Gn
  - -f:v
- 3 flow specific options (39):
  - -kd: 1|2|3 (default: 1)
  - -kdn: B|N|O (default: O)
  - -kdt: Z|S|P (default: ZSP)

LVG - Derivation Examples

- **shell> lvg –f:d –p –SC –SI**

```
 - Please input a term (type "Ctl-d" to quit) >
hyperuricemic

hyperuricemic|hyperuricemic|<noun>|<base>|d|1|
hyperuricemic|hyperuricemia|<noun>|<base>|d|1|
hyperuricemic|hyperuricemic|<adj>|<base>|d|1|
```

Derivations Generation

- Before 2011-, issues of precision and recall
- A new systematic approach to automatically generate derivational variants using LVG conjunction with the SPECIALIST Lexicon

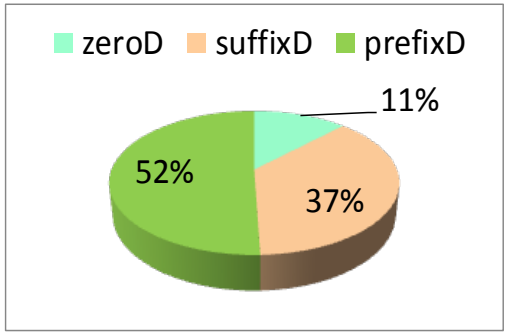
References:

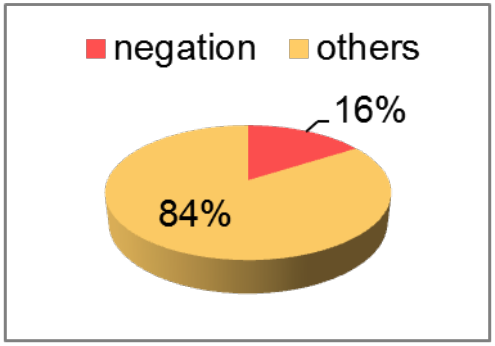
- “A Systematic Approach for Automatically Generating Derivational Variants in Lexical Tools Based on the SPECIALIST Lexicon”, Chris J Lu, Lynn McCreedy, Destinee Tormey, and Allen Browne, IEEE IT Professional Magazine, May/June, 2012, p. 36-42
- “Implementing Comprehensive Derivational Features in Lexical Tools Using a Systematical Approach”, Chris J Lu, Lynn McCreedy, Destinee Tormey, and Allen Browne, AMIA 2013 Annual Symposium, Nov. 16-20, Washington, DC, p. 904

Derivations Growth

- Better coverage:
  - Facts: cover all dPairs known to Lexicon (grow proportionally with Lexicon annually)

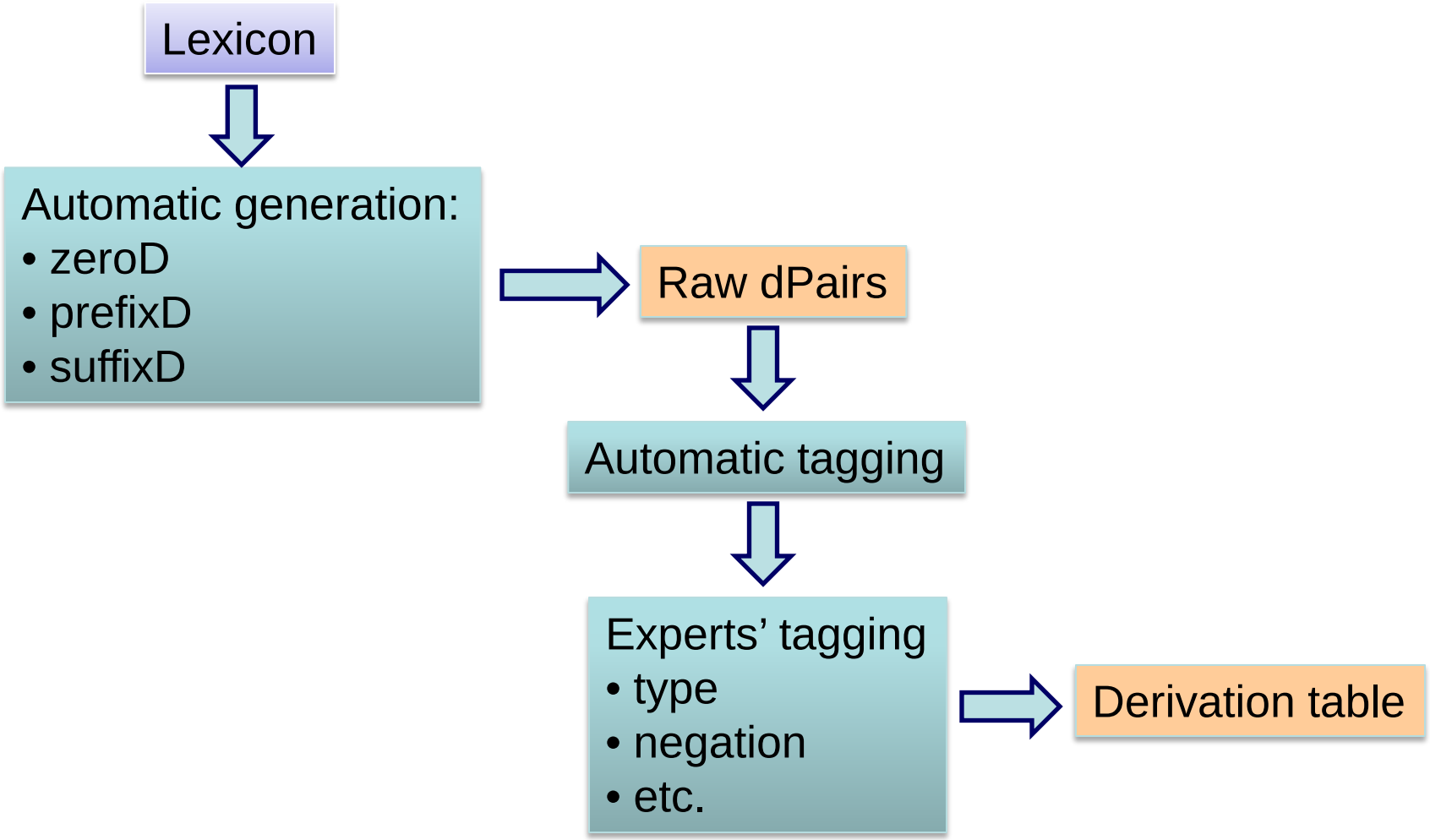
2011-	2012	2013	2014	2015
4,559	89,950	121,078	140,203	141,623





- Better precision:
  - Mainly relies on facts: virtually 100% accurate
- Derivations not in Lexicon => SD-Rules

Facts Generation

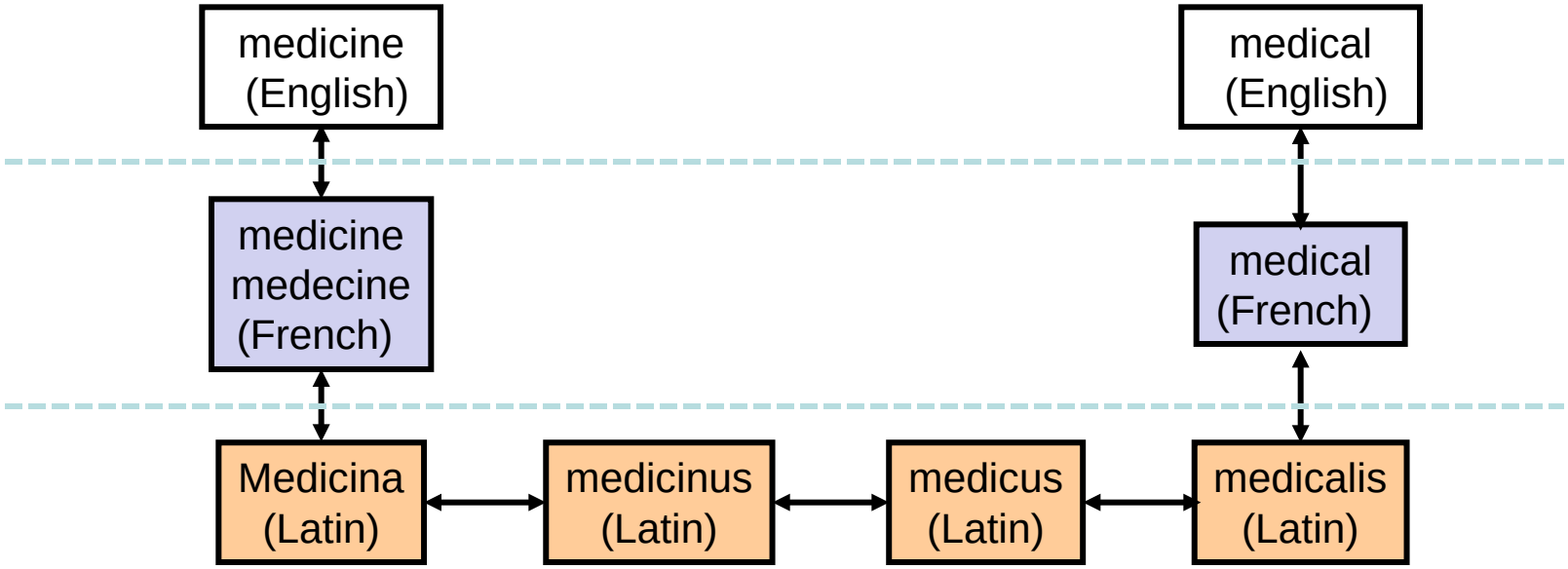


References:

•"A Systematic Approach for Automatically Generating Derivational Variants in Lexical Tools Based on the SPECIALIST Lexicon", Chris J Lu, Lynn McCreedy, Destinee Tormey, and Allen Browne, IEEE IT Professional Magazine, May/June, 2012, p. 36-42

dPair Tagging Process

- A valid dPair:
  derivational related if and only if exactly 1 derivational step between two forms
- The recursive derivational flow component is used to retrieve related words that are more than 1 derivational step
- Example (medicine | medical):
  medical|adj|E0039257|medicine|noun|E0039272 ?



SD- Facts & Rules

- Facts
  - Derivational pairs database table

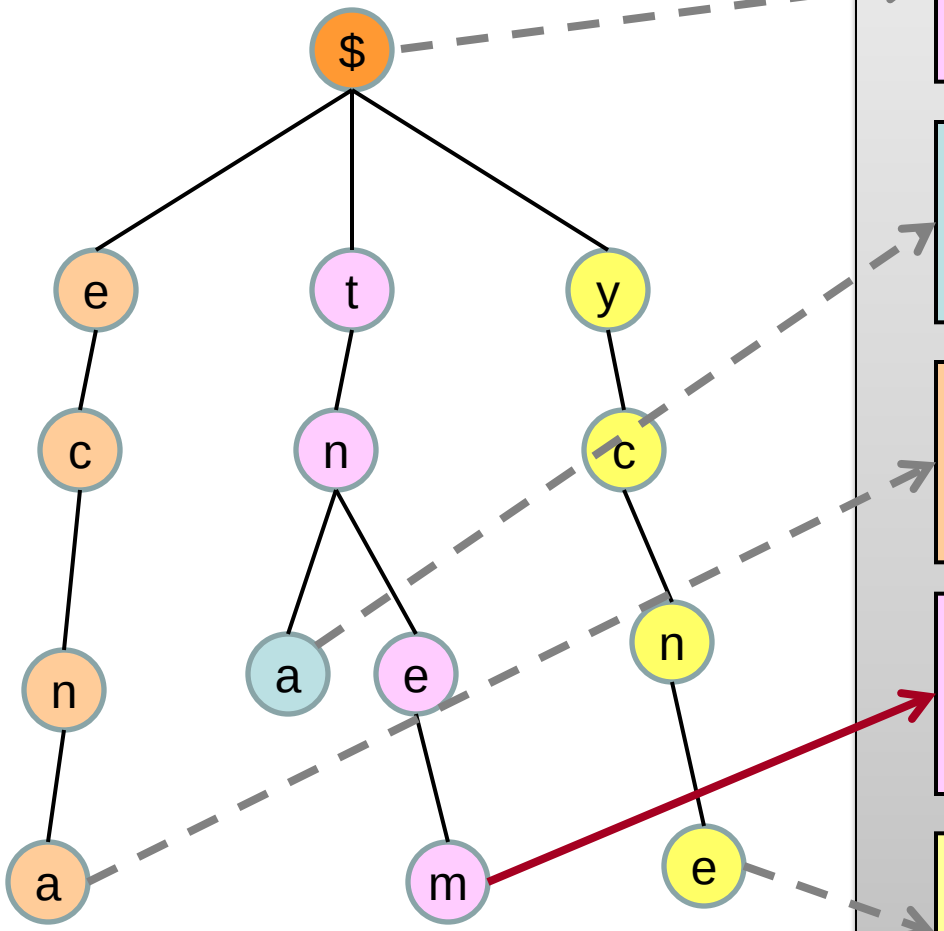
Base-1	Cat-1	EUI-1	Base-2	Cat-2	EUI-2	Negation	Type	prefix
…	…	…	…	…		…	…	…
care	noun	E0015334	**pre**care	noun	E0611704	O	P	**pre**
care	noun	E0015334	care**less**	adj	E0015344	N	S	None
care	noun	E0015334	care	verb	E0015335	O	Z	None
…	…	…	…	…	…	…	…	…

- SD-Rules
  - Use exceptions to increase precision

```
EXAMPLE: retirement|noun|retire|verb
RULE: ment$|noun|$|verb
EXCEPTION: apartment|apart;
EXCEPTION: basement|base;
EXCEPTION: department|depart;
…
```

SD-Rules (Trie)

- retirement|noun => retire|verb



EXAMPLE: retire|verb|retirement|noun
RULE: $|verb|ment$|noun
EXCEPTION: apart|apartment;
…

EXAMPLE: relaxant|adj|relax|verb
RULE: ant$|adj|$|verb
EXCEPTION: important|import;
…

EXAMPLE: conformant|adj|conformance|noun
RULE: ance$|noun|ant$|adj
EXCEPTION: ambulant|ambulance;
…

EXAMPLE: retirement|noun|retire|verb
RULE: ment$|noun|$|verb
EXCEPTION: apartment|apart;
…

EXAMPLE: fluent|adj|fluency|noun
RULE: ency$|noun|ent$|adj
EXCEPTION: emergency|emergent;
…

SD-Rules - 2014

- SD-Rules to generate SD-Facts:

Rank	Rules to generate Raw SD-Pairs	Retrieved	Valid	Precision
1	$\|adj\|ness$|noun	2723	2723	100.00%
2	ability$\|noun\|able$|adj	1278	1278	100.00%
3	le$\|adj\|ly$|adv	326	326	100.00%
4	de$\|verb\|sion$|noun	57	57	100.00%
5	ence$\|noun\|ential$|adj	42	42	100.00%
	…	…	…	…
71	ant$\|adj\|ate$|verb	109	70	64.22%
72	$\|noun\|ist$|noun	332	208	62.65%
73	ar$\|adj\|e$|noun	183	111	60.66%
	…	…	…	…
92	ism$\|noun\|ist$|noun	334	4	1.20%
93	a$\|noun\|an$|noun	273	1	0.37%
94	gram$\|noun\|graphy$|noun	358	0	0.00%
95	gram$\|noun\|graphic$|adj	228	0	0.00%
96	$\|verb\|ably$|adv	57	0	0.00%

Good Rules

- High precision
- High frequency
- High system performance (precision + recall)
  - Objective:

  To find an optimized set of SD-Rules to reach best performance (system precision and recall)
    - o to have high precision (**95%**)
    - o to cover more derivations (recall) that are not in Lexicon
  - Assumption:

  Use Lexicon as the testing corpus by assuming Lexicon is a representable subset of general English

System Performance

- Sort all SD-Rules by:
  - precision (= valid No. / raw No.)
  - raw No. (frequency).
  - alphabetic order of SD-Rules

- System performance:
  - System precision (cumulative):
    P = relevant, retrieved / retrieved
  - System recall:
    R = relevant, retrieved / relevant

  - More SD-Rules (for tie-breaker)

2015 New SD-Rules

- From Linguists' Suggestions (7):

Suggested Rules	Examples
al$\|adj\|us$|noun	viral|adj|virus|noun
$\|noun\|ize$|verb	terror|noun|terrorize|verb
e$\|verb\|ing$|noun	rave|verb|raving|noun
ian$\|adj\|ia$|noun	australian|adj|australia|noun
ian$\|noun\|ia$|noun	australian|noun|australia|noun
es$\|noun\|ic$|adj	diabetes|noun|diabetic|adj
es$\|noun\|ic$|noun	diabetes|noun|diabetic|noun

2015 New SD-Rules

- From Computational Linguistics (8):

Derived Rules from NomD	Examples
e$\|verb\|ion$|noun	evocate|verb|E0538633|evocation|noun|E0417865
ility$\|noun\|le$|adj	appliability|noun|E0541203|appliable|adj|E0541202
ce$\|noun\|t$|adj	equivalence|noun|E0025964|equivalent|adj|E0025966
cy$\|noun\|t$|adj	reluctancy|noun|E0514595|reluctant|adj|E0052653
se$\|verb\|zation$|noun	sanitise|verb|E0054320|sanitization|noun|E0054319
sation$\|noun\|ze$|verb	manualisation|noun|E0579348|manualize|verb|E0579347

Derived Rules from OrgD	Examples
$\|adj\|ally$|adv	basic|adj|E0012047|basically|adv|E0218453
c$\|adj\|s$|noun	gastritic|adj|E0029371|gastritis|noun|E0029372

Retrieve SD-Rules from Facts

- Facts – nominalization:

```
{base=celebrate
entry=E0015729
        cat=verb
        variants=reg
        intran
        tran=np
        nominalization=celebration|noun|E0015730
}
```

- Derived SD-Rules:

  celebrate|verb|celebration|noun => e$|verb|ion$|noun

SD-Rules from NomD (6)

- Derived 1,017 SD-Rules from 23,384 from NomD in 2015 release
- 6 (17) rules with high frequency (>= 200, 1%, Accu. 80%) are evaluated:

SD-Rules	Instances No.	Accu. No.	Notes
$\|adj\|ness$|noun	2733 (11.69%)	2733 (11.69%)	Exists
ation$\|noun\|e$|verb	2472 (10.57%)	5205 (22.26%)	Exists
e$\|verb\|ion$|noun	2298 (9.83%)	7503 (32.09%)	New, has child rules exist
$\|adj\|ity$|noun	2036 (8.71%)	9539 (40.79%)	Exists
ility$\|noun\|le$|adj	1609 (6.88%)	11148 (47.67%)	New, has child rules exist
se$\|verb\|zation$|noun	1100 (4.70%)	12248 (52.38%)	New, has no child rules exist
sation$\|noun\|ze$|verb	1064 (4.55%)	13312 (56.93%)	New, has no child rules exist
ce$\|noun\|t$|adj	836 (3.58%)	14148 (60.50%)	New, has child rules exist
e$\|adj\|ity$|noun	830 (3.55%)	14978 (64.05%)	Exists
ed$\|adj\|ion$|noun	675 (2.89%)	15653 (66.94%)	Exists
$\|verb\|ment$|noun	574 (2.45%)	16227 (69.39%)	Exists
iness$\|noun\|y$|adj	544 (2.33%)	16771 (71.72%)	Exists
$\|verb\|ion$|noun	535 (2.29%)	17306 (74.01%)	Exists
$\|verb\|ing$|noun	478 (2.04%)	17784 (76.05%)	Exists
cy$\|noun\|t$|adj	400 (1.71%)	18184 (77.76%)	New, has child rules exist
$\|verb\|ation$|noun	307 (1.31%)	18491 (79.08%)	Exists
ication$\|noun\|y$|verb	295 (1.26%)	18786 (80.34%)	Exists

SD-Rules from OrgD (2)

- Derived 1,421 SD-Rules from 4,110 from OrgD in 2015 release
- 2 (6) rules with high frequency (>= 40, 1.0%, Accu. 11.5%) are evaluated:

SD-Rules	Instances No.	Accu. No.	Notes
less$\|adj\|$|noun	131 (3.19%)	131 (3.19%)	Exists
$\|verb\|ion$|noun	111 (2.70%)	242 (5.89%)	Exists
ist$\|noun\|y$|noun	63 (1.53%)	305 (7.42%)	Exists
ally$\|adv\|$|adj	58 (1.41%)	363 (8.83%)	New, with existing child rules
ful$\|adj\|$|noun	58 (1.41%)	421 (10.24%)	Exists
c$\|adj\|s$|noun	54 (1.31%)	475 (11.56%)	New, with existing child rules

Optimization Procedures

- Baseline
  - Add 15 new SD-Rules to 2014 SD-Rule set
  - 14 parents and 19 child rules are found
  - Normalize SD-Rule set by removing child rules
    Unify bi-directional SD-Rules (alphabetic order sorting)
- Optimization: Parent-Child Evaluation
  - Use program to decompose parent rule and retrieve child rules:
    - Coverage rate for decomposing: 40%
    - Coverage rate for child rules: 25%
  - Find a set with better system performance

References:

- "Implementing Comprehensive Derivational Features in Lexical Tools Using a Systematical Approach", Chris J Lu, Lynn McCreedy, Destinee Tormey, and Allen Browne, AMIA 2013 Annual Symposium, Nov. 16-20, Washington, DC, p. 904

Parent-Child Rules

- Examples –
  - ❖ confidence|noun|E0018410|confident|adj|E0018411
  - ❖ relevance|noun|E0052632|relevant|adj|E0052634

  - 0|858|837|21| ce$|noun|t$|adj |97.55%|100.00% (root parent rule)
  - 1|847|837|10| nce$|noun|nt$|adj |98.82%|98.72%

  - 2|319|316|3| ance$|noun|ant$|adj |99.06%|37.18%
  - 2|528|521|7| ence$|noun|ent$|adj |98.67%|61.54%

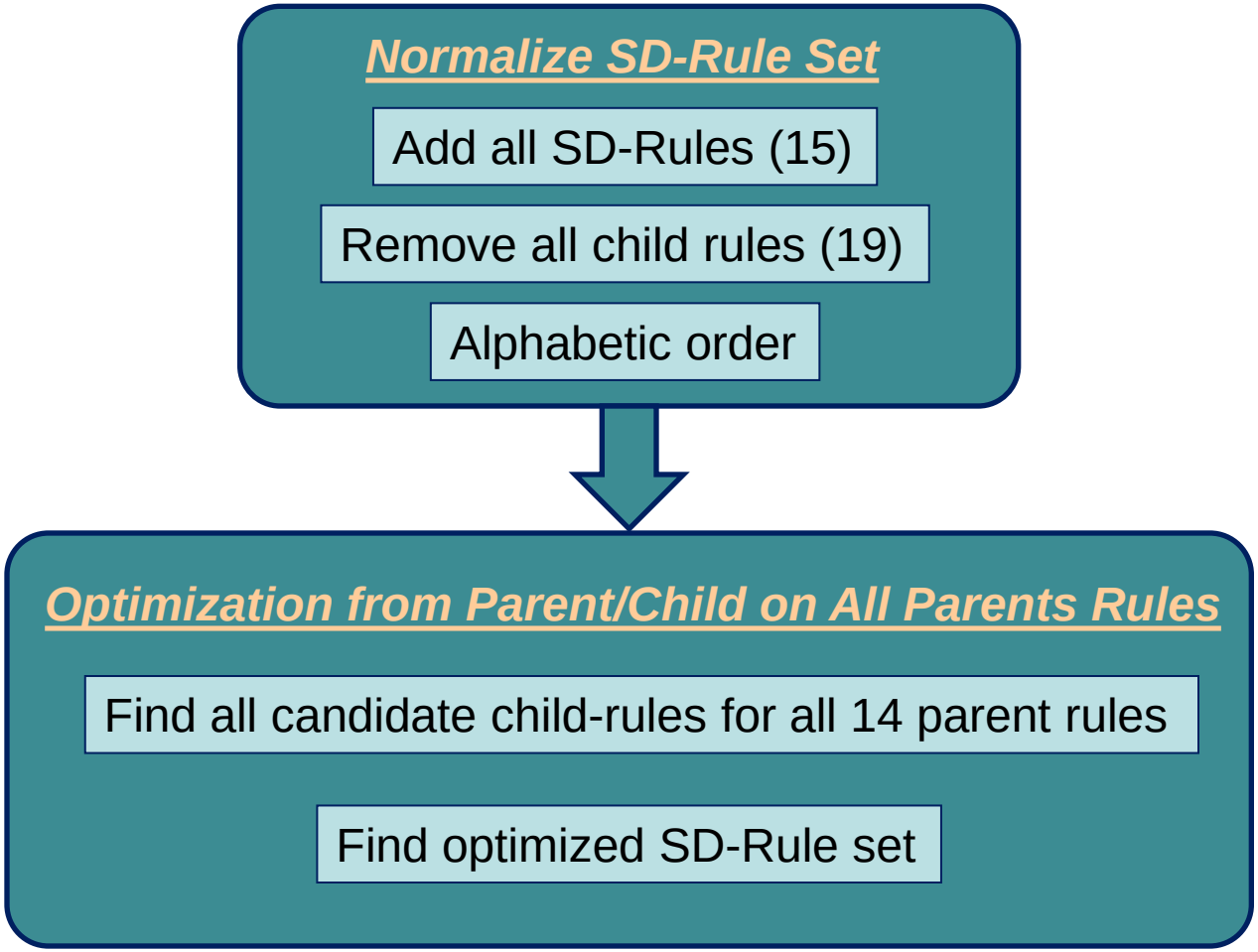
- Heuristic algorithm for candidate child rules:
  - Coverage rate for decomposing: 40%
  - Coverage rate for child rules: 25%
  - Child precision >= parent precision

- Root parent rule:
  A rule does not have any possible parent rule.

14 Parent Rules (2015)

No.	Parent SD-Rules
1	$\|adj\|ally$|adv|2015|ORG_FACT|PARENT
2	$\|adj\|ity$|noun|2013|ORG_RULE|PARENT
3	$\|noun\|al$|adj|2013|ORG_RULE|PARENT
4	$\|verb\|ion$|noun|2013|NOM_D|PARENT
5	a$\|noun\|an$|adj|2013|ORG_RULE|PARENT
6	a$\|noun\|an$|noun|2013|ORG_RULE|PARENT
7	a$\|noun\|ar$|adj|2013|ORG_RULE|PARENT
8	ation$\|noun\|e$|verb|2013|ORG_RULE|PARENT
9	c$\|adj\|s$|noun|2015|ORG_FACT|PARENT
10	ce$\|noun\|t$|adj|2015|NOM_D|PARENT
11	cy$\|noun\|t$|adj|2015|NOM_D|PARENT
12	e$\|verb\|ion$|noun|2015|NOM_D|PARENT
13	ility$\|noun\|le$|adj|2015|NOM_D|PARENT
14	sis$\|noun\|tic$|adj|2013|ORG_RULE|PARENT

Adding 15 New Rules, 2015



Optimization Set (2015)

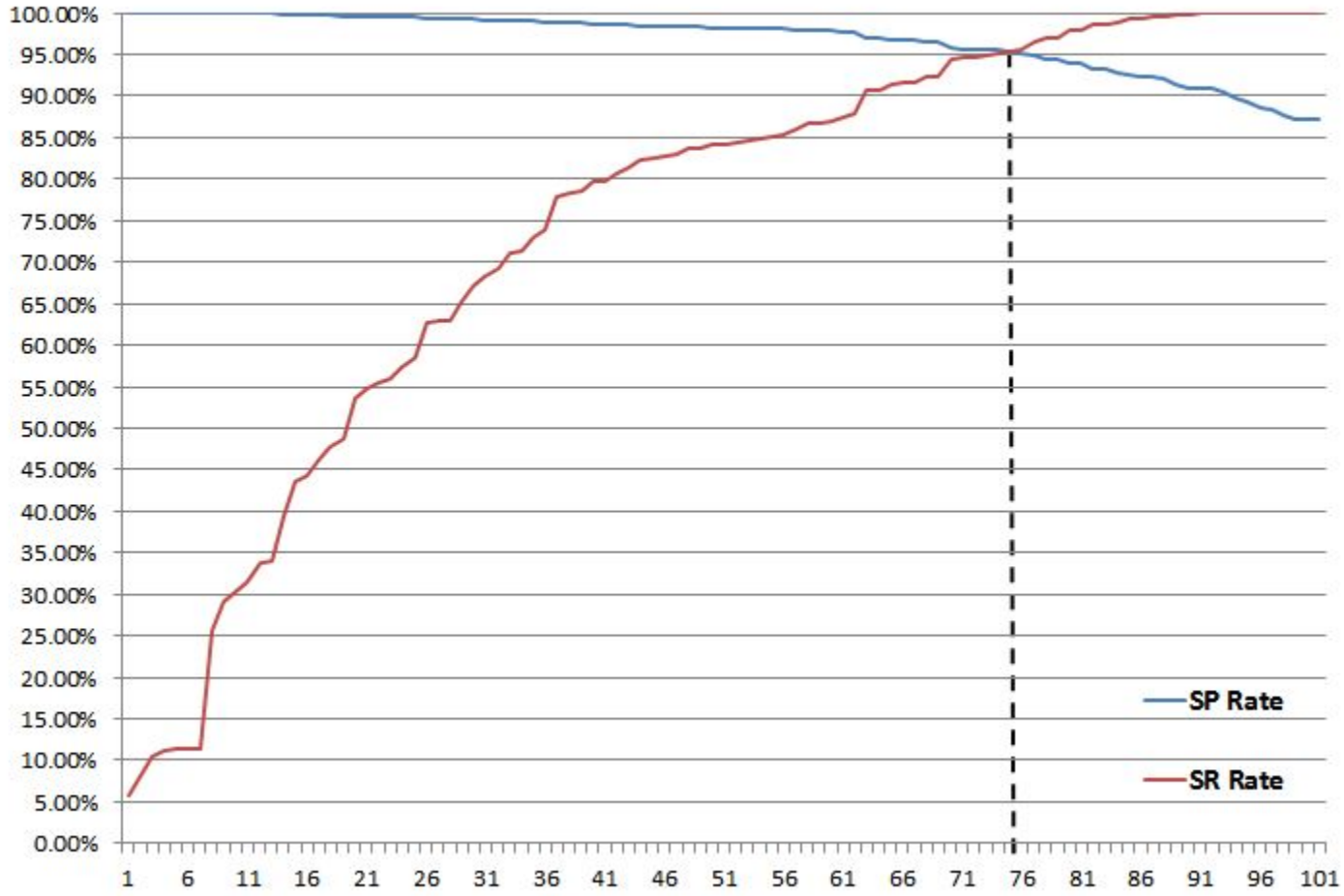
- Optimized Set (76/101) with 95.22% precision, 95.70% recall
- Total relevant, retrieved by all parents rules: 46,950

Rank	Rule Precision	Raw Retrieve	Valid R.R.	SD-Rule	Accum. Retrieve	Accum. R.R.	System Precision	System Recall	System Performance
1	100.00%	2734	2734	$\|adj\|ness$|noun	2734	2734	100.00%	5.82%	1.0582
2	100.00%	1108	1108	se$\|verb\|zation$|noun	3842	3842	100.00%	8.18%	1.0818
3	100.00%	1071	1071	sation$\|noun\|ze$|verb	4913	4913	100.00%	10.45%	1.1046
	…	…	…	…	…	…	…	…	…
75	63.14%	331	209	$\|noun\|ist$|noun	46999	44817	95.36%	95.46%	1.9081
76	**61.70%**	**188**	**116**	**ar$\|adj\|e$|noun**	**47187**	**44933**	**95.22%**	**95.70%**	**1.9093**
77	60.23%	596	359	al$\|adj\|e$|noun	47783	45292	94.79%	96.47%	1.9126
	…	…	…	…	…	…	…	…	…
99	0.36%	274	1	a$\|noun\|an$|noun	53809	46948	86.60%	98.75%	1.8725
100	0.00%	57	0	$\|verb\|ably$|adv	53866	46948	86.13%	98.75%	1.8715
101	0.00%	19	0	es$\|noun\|ic$|noun	53885	46948	86.02%	98.75%	1.8712

Total R.R.: (46,950)

Optimization Results

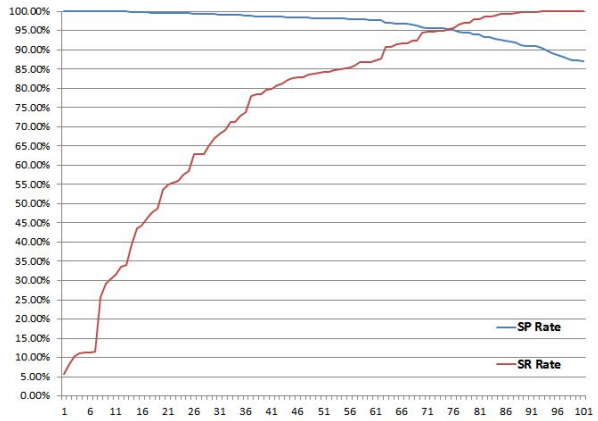
- Set 10.1 (optimized Sd-Rule Set, 76/101 rules)



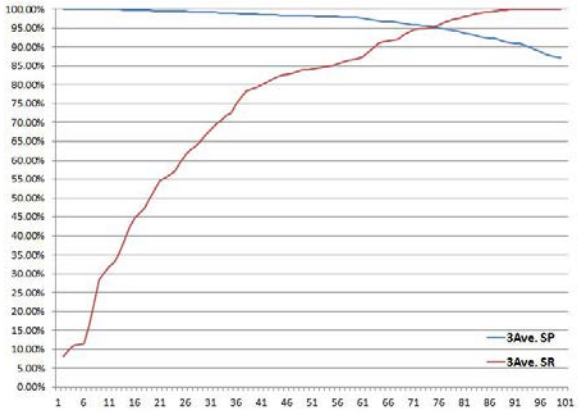
Results - Noise Reduction

- Smoothing algorithm – simple moving average of 3, 5, 7 window size
- The intersections are all around 95% for all cases
- Confirm our optimized goal of 95% S.A. is a good choice

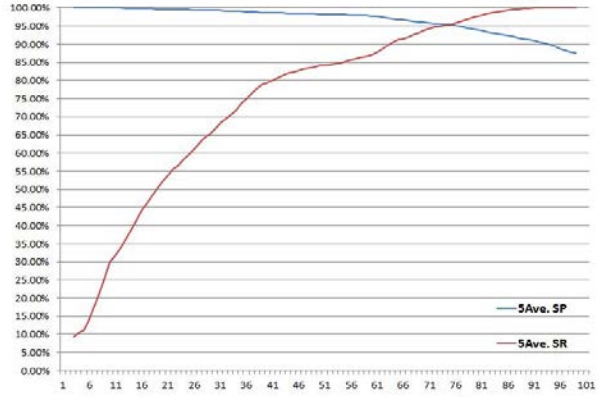
Final Set, 2015: Original data



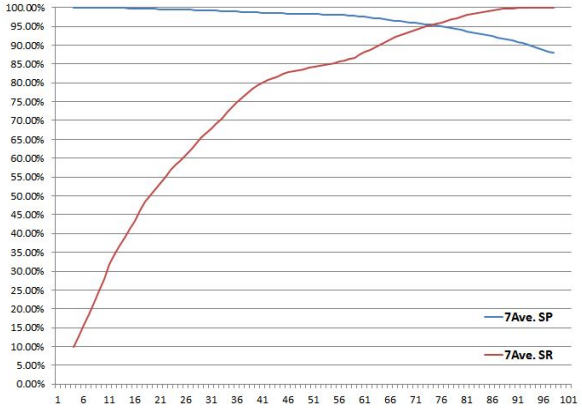
Final Set, 2015: Avg. with 3 points



Final Set, 2015: Avg. with 5 points



Final Set, 2015: Avg. with 7 points



Summary on New Rules

SD-Rule	Rank	Precision	Instances	Source	Results
11 Good Rules					
se$\|verb\|zation$|noun	2	100.00%	1108	NOM_D	Good SD-Rule
sation$\|noun\|ze$|verb	3	100.00%	1071	NOM_D	Good SD-Rule
ility$\|noun\|le$|adj	9	99.94%	1626	NOM_D	Good SD-Rule
$\|adj\|ally$|adv	15	99.08%	2072	ORG_D	Good SD-Rule
ce$\|noun\|t$|adj	18	98.82%	847	NOM_D	Child rule nce$\|noun\|nt$|adj is used
cy$\|noun\|t$|adj	19	98.77%	406	NOM_D	Good SD-Rule
e$\|verb\|ion$|noun	20	98.76%	2336	NOM_D	Good SD-Rule
c$\|adj\|s$|noun	43	91.46%	281	ORG_D	Child rule ic$\|adj\|is$|noun is used
e$\|verb\|ing$|noun	45	91.43%	210	Suggestions	Good SD-Rule
al$\|adj\|us$|noun	61	84.35%	262	Suggestions	Good SD-Rule
es$\|noun\|ic$|adj	67	73.91%	23	Suggestions	Good SD-Rule
4 Bad Rules					
ian$\|adj\|ia$|noun	57	86.31%	263	Suggestions	Duplicated, parent rule an$\|adj\|a$|noun is used
$\|noun\|ize$|verb	78	59.05%	442	Suggestions	Bad SD-Rule
ian$\|noun\|ia$|noun	99	0.36%	274	Suggestions	Duplicated, parent rule an$\|noun\|a$|noun is a bad SD-Rule
es$\|noun\|ic$|noun	101	0.00%	19	Suggestions	Bad SD-Rule

Summary - Comparison

- All computer generated SD-Rules from NOM_D and ORG_D are good rules with high precision and instances (8/8)

	Linguists' Suggestion	Computational Linguistics
Suggestion No.	7	8
Bad Rules	4 (2 duplicates)	0
New good rules	3 (43%)	8 (100%)

- Evaluate more SD-Rules with lower frequency from NomD and OrgD

Computational Linguistics - NomD

The top 17 SD-Rules generated from NOM_D are all good SD-Rules.

SD-Rule	Rank	Precision	Instances	status
Frequency > 200, Instance coverage > 1.00% , Accum. coverage > 80.0%				
$\|adj\|ness$|noun	1	100.00%	2734	Existing SD-Rule
se$\|verb\|zation$|noun	2	100.00%	1108	New SD-Rule
sation$\|noun\|ze$|verb	3	100.00%	1071	New SD-Rule
ility$\|noun\|le$|adj	9	99.94%	1626	New SD-Rule
iness$\|noun\|y$|adj	10	99.82%	545	Existing SD-Rule
ation$\|noun\|e$|verb	14	99.24%	2514	Existing SD-Rule
ce$\|noun\|t$|adj	18	98.82%	847	New child rule nce$\|noun\|nt$|adj is used
cy$\|noun\|t$|adj	19	98.77%	406	New SD-Rule
e$\|verb\|ion$|noun	20	98.76%	2336	New SD-Rule
$\|verb\|ment$|noun	21	98.35%	607	Existing SD-Rule
ication$\|noun\|y$|verb	22	98.33%	300	Existing SD-Rule
ed$\|adj\|ion$|noun	24	98.26%	689	Existing SD-Rule
$\|adj\|ity$|noun	26	97.88%	2080	Existing SD-Rule
e$\|adj\|ity$|noun	30	97.54%	894	Existing SD-Rule
$\|verb\|ion$|noun	40	93.71%	572	Existing SD-Rule
$\|verb\|ing$|noun	42	92.53%	522	Existing SD-Rule
$\|verb\|ation$|noun	48	90.29%	340	Existing SD-Rule

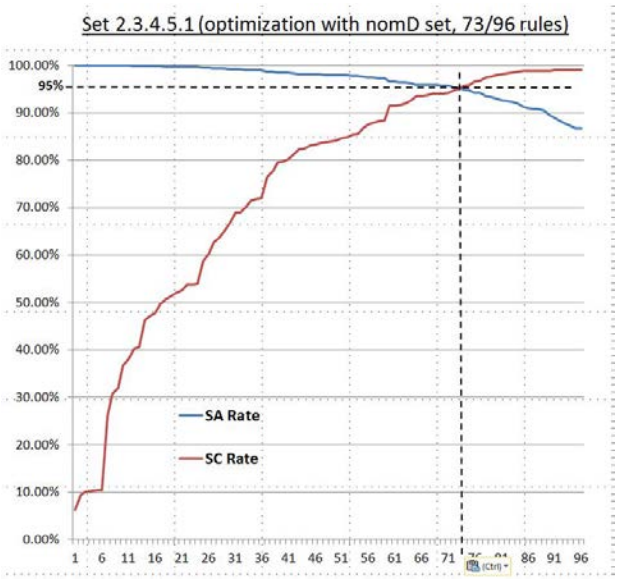
Computational Linguistics - OrgD

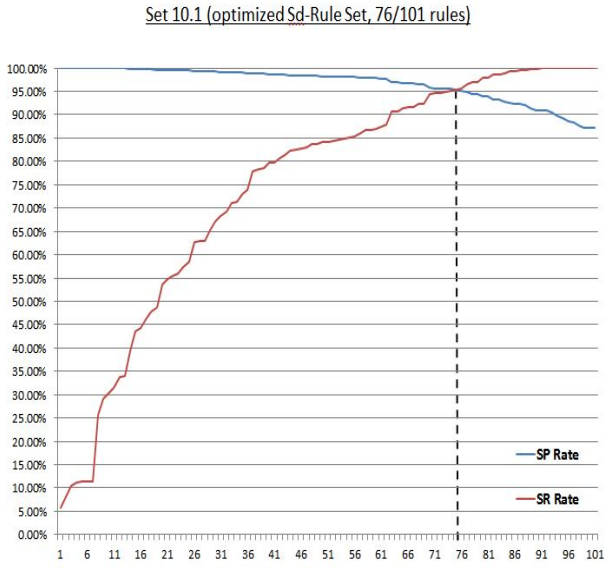
The top 6 SD-Rules generated from ORG_D are all good SD-Rules

SD-Rule	Rank	Precision	Instances	status
Frequency > 40, Instance coverage > 1.00% , Accum. coverage > 11.50%				
less$\|adj\|$|noun	11	99.64%	561	Existing SD-Rule
$\|adj\|ally$|adv	15	99.08%	2072	New SD-Rule
$\|verb\|ion$|noun	40	93.71%	572	Existing SD-Rule, also derived from NOM_D
ist$\|noun\|y$|noun	36	95.48%	509	Existing SD-Rule
c$\|adj\|s$|noun	43	91.46%	281	New child rule ic$\|adj\|is$|noun is used
ful$\|adj\|$|noun	50	89.21%	139	Existing SD-Rule

SD-Rule Set Comparison

Item	2014	2015
Total Unique Rules	96	101
Total Good Rules	73	76
Opti. System Precision	95.30%	95.22%
Opti. System Recall	95.01%	95.70%
Opti. System Performance	1.9031	1.9093
Cutoff for Good SD-Rule	ar$\|adj\|e$|noun	ar$\|adj\|e$|noun





Comparison (Cont.)

- All good rules in 2014 are also good in 2015 release.
- 2014 has 73 good rules translated to 76 food rules in 2015:
  - 2 pairs of two child rules (4) are replaced by 2 parent rules (-2)
  - 3 others of single child rules are replaced by 3 parents rules (0)
  - 5 new rules are added (+5)
  - Total rules count: 73 - 2 + 5 = 76 (2015)
  - Total new good rules count: 2 + 3 + 5 = 11
- The conclusion is the optimized set of SD-Rules are very steady, no odd behavior is observed:
  - Same cutoff rule
  - Similar optimization result: precision and recall intersect around 95%

Future Work

- Annual updates on SD-Rule set with the Lexicon release
- Evaluate more SD-Rules (lower frequency) from facts
  - NomD and OrgD to have better coverage
- Replace all rules by their root parents rules for optimization
  - currently, only parent-rules are evaluated

- Assumption (from Lexicon to English):
  - Is Lexicon a representable subset of general English?
  - Characteristics on derivations from Lexicon is very stable

Questions

- Lexical Systems Group: http://umlslex.nlm.nih.gov
- The SPECIALIST NLP Tools: http://specialist.nlm.nih.gov