Motivation

- A *semantic lexicon* contains semantic category assignments for words. For example:

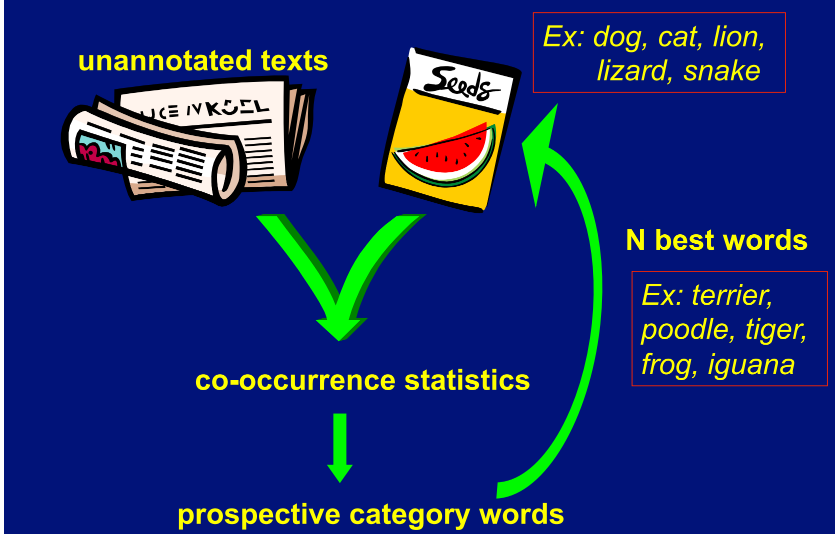
  blogger → *HUMAN*
  sedan → *VEHICLE*
  AK-47 → *WEAPON*

- General purpose resources, such as WordNet, are often insufficient for specific domains.

  ANIMAL: gshep, doxy, lab, labx, m/n, mix, patient
  HUMAN: o

- Automatic methods can be used to enhance existing resources or create domain-specific lexicons.

Bootstrapping Semantic Lexicons



Lexico-Syntactic Patterns

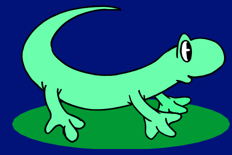
- Lexico-syntactic contexts often reveal the semantic class of a word.
- AutoSlog [Riloff 1993] is a pattern generator that was originally developed for event extraction tasks.
- Each pattern co-occurs with a NP in one of 3 syntactic positions: *subject, direct object, PP object*.

Example Location Patterns

<subject> was inhabited	the locality was inhabited…
patrolling <direct object>	…patrolling Zacamil neighborhood
lives in <PP object>	…lives in Argentina

Pattern Templates	Lexico-Syntactic Patterns
<subject> passive-vp	*<target> was bombed*
<subject> active-vp	*<perpetrator> bombed*
<subject> active-vp dobj	*<perpetrator> threw dynamite*
<subject> active-vp infinitive	*<perpetrator> tried to kill*
<subject> passive-vp infinitive	*<perpetrator> was hired to kill*
<subject> auxiliary dobj	*<victim> was fatality*
active-vp <dobj>	*bombed <target>*
infinitive <dobj>	*to kill <victim>*
active-vp infinitive <dobj>	*tried to kill <victim>*
passive-vp infinitive <dobj>	*was hired to kill <victim>*
subject auxiliary <dobj>	*fatality was <victim>*
passive-vp prep <np>	*was killed by <perpetrator>*
active-vp prep <np>	*exploded in <target>*
infinitive prep <np>	*to kill with <weapon>*
noun prep <np>	*assassination of <victim>*

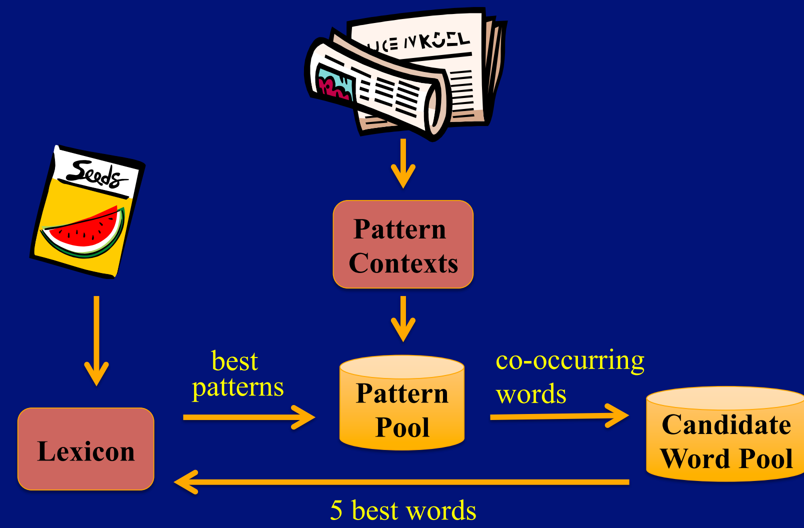
BASILISK = Bootstrapping Approach to SemantIc Lexicon Induction using Semantic Knowledge



Key Ideas behind Basilisk

- Collective evidence over extraction patterns.
- Learning multiple categories simultaneously.

Basilisk Bootstrapping Algorithm



Scoring Patterns

Every extraction pattern is scored and the best patterns are put into a *Pattern Pool*.

The scoring function is:

$$\text{RlogF}(\text{pattern}_i) = \frac{F_i}{N_i} * \log_2(F_i)$$

where:

F_i is the number of category members extracted by pattern_i
N_i is the total number of nouns extracted by pattern_i

The Pattern Pool

- Initially, we used a Pattern Pool of size 20, but the pool became stagnant over time.

  Solution: begin with a pattern pool of size 20, but increase the pool size by 1 after each iteration to infuse the pool with new candidates.

- All head nouns that co-occur with patterns in the Pattern Pool are put into the Candidate Word Pool.

Scoring Words based on Collective Evidence

1. Given a word, collect all of its pattern contexts.

2. Compute the average # of distinct class members per pattern. (Actually, average over logarithms.)

INTUITION: a word receives a high score if it occurs in contexts that also consistently co-occur with known semantic class members.

Selecting Words for the Lexicon

$score(word_i)$ = the average number of category members that co-occur with the pattern contexts containing the candidate word.

$$\textbf{score}(\textbf{word}_i) = \frac{\sum_{j=1}^{N_i} F_j}{N_i}$$

$$\textbf{AvgLog}(\textbf{word}_i) = \frac{\sum_{j=1}^{N_i} \log_2 (F_j + 1)}{N_i}$$

where:

F_j is the # of distinct category members that co-occur with $pattern_j$

N_i is the total number of patterns that co-occur with $word_i$

Experimental Design

- Used the MUC-4 corpus: 1700 texts related to terrorism.
- Experiments on 6 semantic categories: *building*, *event*, *human*, *location*, *time*, *weapon*.
- 10 seed words for each category.
- 1000 words automatically generated for each category.
- Basilisk was compared with our previous algorithm (*meta-bootstrapping*).

Baseline Results

Head Nouns (8460 words)		
building	188	(2.2%)
event	501	(5.9%)
human	1856	(21.9%)
location	1018	(12.0%)
time	112	(1.3%)
weapon	147	(1.7%)
(other)	4638	(54.8%)

Seed Words

We used the 10 most frequent words for each category.

Building: embassy, office, headquarters, church, offices, house, home, residence, hospital, airport

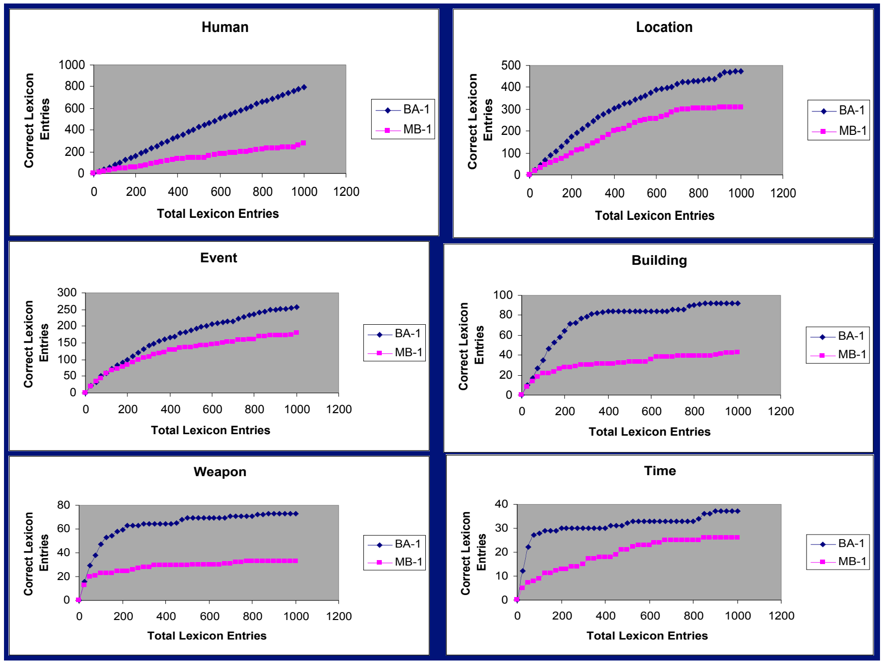
Event: attack, actions, war, meeting, elections, murder, attacks, action, struggle, agreement

Human: people, guerrillas, members, troops, Cristiani, rebels, president, terrorists, soldiers, leaders

Location: country, El Salvador, Salvador, United States, area, Colombia, city, countries, department, Nicaragua

Time: time, years, days, November, hours, night, morning, week, year, day

Weapon: weapons, bomb, bombs, explosives, arms, missles, dynamite, rifles, materiel, bullets



Semantic Learning Case Study

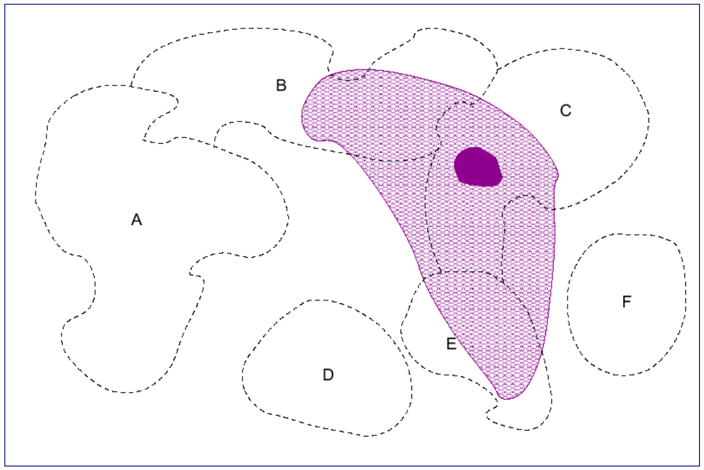
- Seed Words: 10 common disease names
- Of the top 200 words hypothesized to be diseases: 89 were already in the UMLS metathesaurus (32,000 names of diseases and organisms), but 111 were not! Including:

adenomatosis	flu	h5n1
tularaemia	kawasaki	h7n3
tularamia	mad-cow-disease	ev71
diarrhoea	smut	yf
diphtheriae	pertussis	jyf
enterovirus-71	pleuro-pneumonia	nvcjd
fibropapillomas	polioencephalomyelitis	pepmv
gastroeneteritis	poliovirus	wsmv

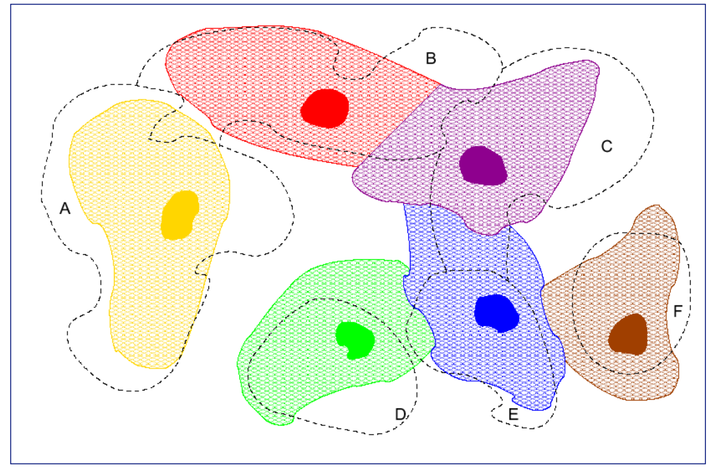
Learning Multiple Categories Simultaneously

- We hypothesized that confusion errors can be reduced by learning multiple semantic categories simultaneously.
- “One Sense per Domain” assumption.
- Knowledge about competing categories can constrain and steer the bootstrapping process.

Bootstrapping a Single Category

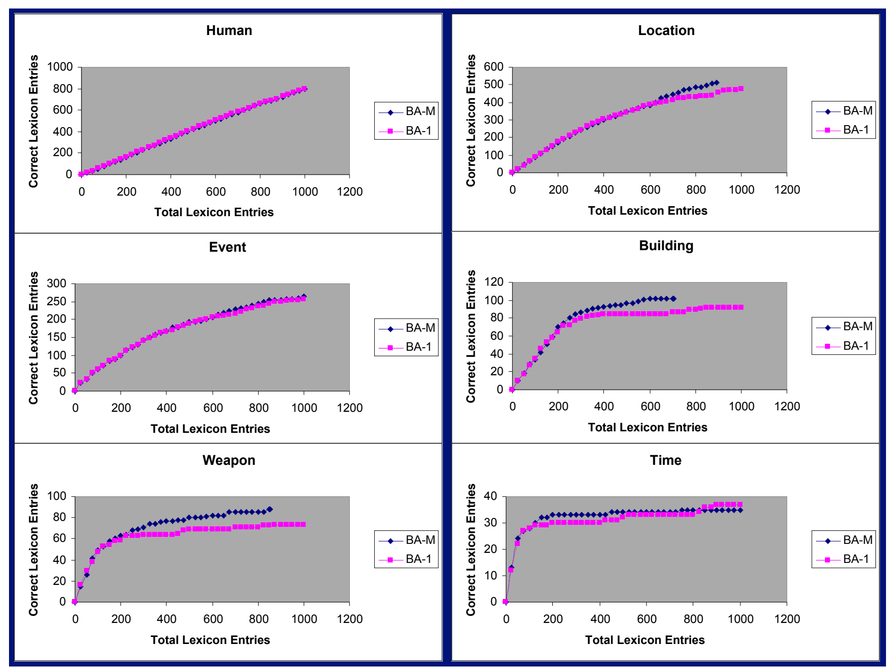


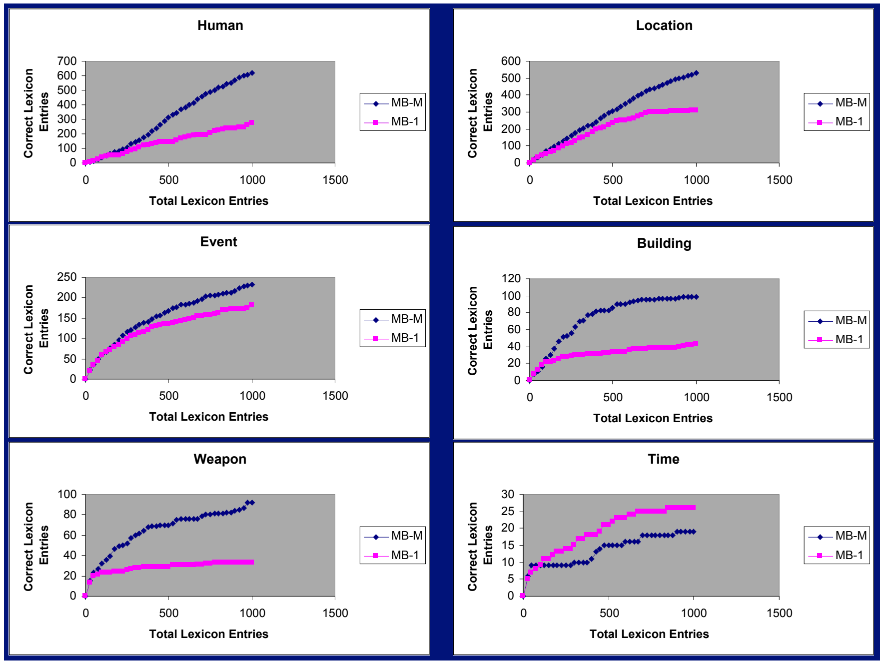
Bootstrapping Multiple Categories



Simple Conflict Resolution

- A word cannot be assigned to category X if it has already been assigned to category Y.
- If a word is hypothesized for both category X and category Y at the same time, choose the category that receives the highest score.



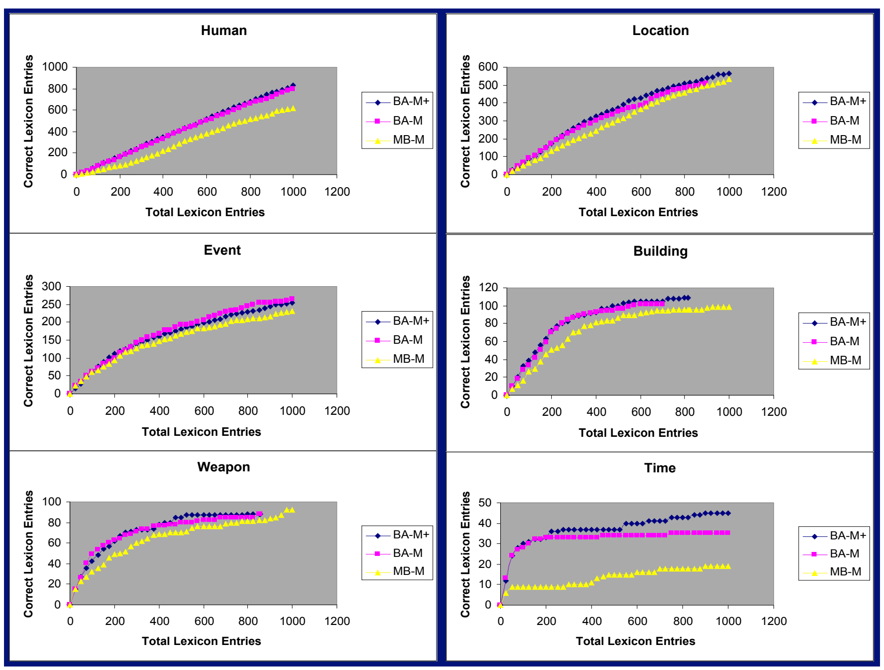


A Smarter Scoring Function

A more proactive approach: incorporate knowledge about other categories directly into the scoring function.

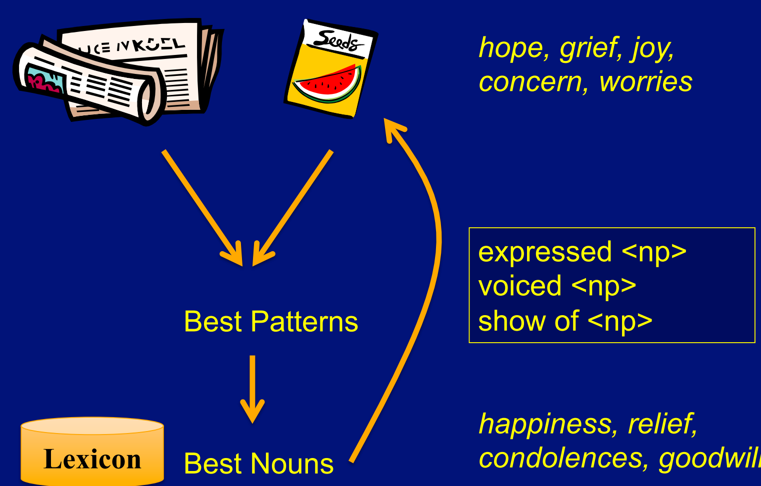
New scoring function:

$$\mathrm{diff}\,(w_i,c_a) = \mathrm{AvgLog}\,(w_i,c_a) - \max_{b \neq a}\,(\mathrm{AvgLog}(w_i,c_b))$$



Subjective Noun Bootstrapping

[Riloff, Wiebe, and Wilson, 2003]

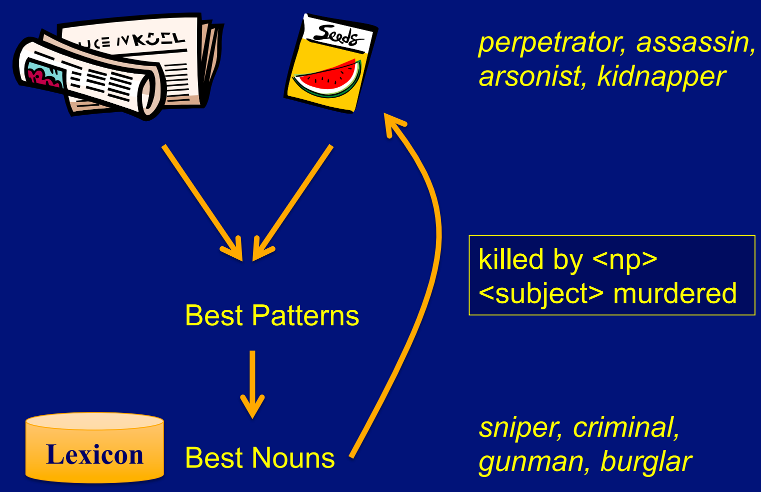


Examples of Learned Subjective Nouns

tyranny smokescreen apologist barbarian belligerence condemnation sanctimonious exaggeration repudiation insinuation antagonism atrocities denunciation exploitation humiliation ill-treatment sympathy scum bully devil liar pariah venom diatribe mockery anguish fallacies evil genius goodwill injustice innuendo revenge rogue

Role-Identifying Noun Bootstrapping

[Phillips and Riloff, 2007]



Learned Role-Identifying Nouns

<u>Terrorism Perpetrators:</u>
assailants, attackers, cell, culprits, extremists, hitmen, kidnappers, militiamen, MRTA, narco-terrorists, sniper

<u>Outbreak Victims:</u>
bovines, crow, dead, eagles, fatality, pigs, swine, teenagers, toddlers, victims

Patient Polarity Verbs

- Many everyday actions are good or bad for the entity that is acted upon (the *patient*).

  Bad: *eaten, arrested, captured, hospitalized*
  Good: *fed, adopted, paid, rescued*

- *Hypothesis:* conjoined verbs often share the same polarity.
  − *abducted and killed; indicted and arrested*
  + *rescued and rehabilitated; promoted and tenured*

Patient Polarity Verb Bootstrapping

[Goyal, Riloff, and Daume III, 2010]

Examples of Learned PPVs

Some examples of patient polarity verbs learned by Basilisk using conjunction pattern contexts:

- censor, chase, fire, orphan, paralyze, scare, sue

\+ accommodate, harbor, nurse, obey, respect, value

Conclusions

- Using collective evidence from a set of extraction patterns improves the accuracy of semantic lexicon induction.
- Learning multiple semantic categories at the same time can constrain bootstrapping and improve performance.
- Manual review is still necessary to use the learned dictionaries.
- Performance for some categories is beginning to approach levels for which manual review may not be necessary.