Knowledge Acquisition From Text

- Statistical NLP methods need data in large quantities
- Explicit Knowledge
  - relations among words:
    - synonymy (buy ≈ acquire), cohypohymy (cat ≈ dog), hyponymy, (cat → animal, buy → own), part of relation (wheel partof car), etc.
  - relations among patterns (sentence prototypes)
    - X bought Y ≈ X acquired Z% of the Y's shares
- Implicit knowledge
  - pairs of sentences with entailment relationship

Principles of Explicit knowledge Acquisition

- **Harris' Distributional Hypothesis (DH)** (Harris, 1964) "Words that tend to occur in the same contexts tend to have similar meanings".
  - E.g.
  - "Dickens wrote David Copperfield"
  - "Dickens penned David Copperfield"
  - ➔ "write" = "pen"

  - E.g. (pattern)
  - "countries such as Italy"
  - "the country of Italy"
  - ➔ "X such as Y" = "the X of Y".

- **Robison's Point-wise Assertion Patterns (PAP)** (Robison, 1970) "w1 is in a relation r with w2 if context pattern r(w1, w2) is observed"
  - E.g.
  - "countries such as Italy"
  - Pattern "*Y such as X*"
  - ➔ "Italy" is-a "country"

A Taxonomy of Explicit Knowledge

- Symmetric vs. Directional
  - Trend is from symmetric to directional
- NP/concept vs. verb/pattern
  - The development of verb/pattern relation is younger

	NP/concept	verb/pattern
Symmetric (semantic class)	(Lin and Pantel 2001a) (Ravichandran and Hovy 2002)	(Lin and Pantel, 2001b) (Szepktor et al., 2004)
Directional (Relation)	(Girju et al. 2006) part-of relation, (Etzioni et al. 2005), (Pantel&Pennacchiotti, 2006), (Banko et al. 2007) all relations	(Zanzotto et al 2006) presupposition relation (Chklovski and Pantel, 2004) strength, antonymy, enablement, happens-before

Systems--Patterns & Instances

- Concepts (NP) and patterns (verb) can play as each other's context
- They enhance each other in iterative approaches

System	Context	Instance
UNICON (Lin and Pantel 2001a)	links of dependency parse tree	NP NP
DIRT (Lin and Pantel 2001b)	NP NP	Paths of dependency trees
# VerbOCEAN (Chklovski & Pantel, 2004)	manually created patterns	verb verb
+ TEASE (Szepktor et al., 2004)	fragment of dependency parse tree	verb verb
+ (Zanzotto et al 2006)	"Agentified-verb verb"	verb verb
* Espresso (Pantel & Pennacchiotti 2006)	Word sequence with Term generalization	NP NP
TextRunner (Banko et al. 2007)	Shallow parsing features	NP NP

+: iterative *: supervised, #: fixed pattern
MI: mutual information, PMI: point wise MI, TE: textual entailment

Systems--Context & Instances Selection

- Complexity of selection range from frequency to mutual information, to classifiers

System	Context Selection	Instance selection
UNICON (Lin and Pantel 2001a)	frequency	feature overlap & clustering
DIRT (Lin and Pantel 2001b)	MI	feature overlap & heuristic constraints
# VerbOCEAN (Chklovski & Pantel, 2004)		MI
+ TEASE (Szepktor et al., 2004)	frequency	Term web frequency & conditional probability
+ (Zanzotto et al 2006)		PMI
* Espresso (Pantel & Pennacchiotti 2006)	modified MI	modified MI
TextRunner (Banko et al. 2007)	Naïve Bayes Classifier	Redundancy (frequency)

+: iterative *: supervised, #: fixed pattern
MI: mutual information, PMI: point wise MI, TE: textual entailment

Systems--Evaluation

- Precision
  - Human validation
- Recall
  - Compare to reference corpus
- Effectiveness
  - NLP tasks: TE, QA, and etc.

System	evaluation
UNICON (Lin and Pantel 2001a)	Human validation
DIRT (Lin and Pantel 2001b)	Human generated paraphrases
VerbOCEAN# (Chklovski & Pantel, 2004)	Human validation
TEASE+ (Szepktor et al., 2004)	Human validation
+ (Zanzotto et al 2006)	WordNet, TE
Espresso* (Pantel & Pennacchiotti 2006)	Human validation
TextRunner (Banko et al. 2007)	Human validation

Principles of Implicit Knowledge Acquisition

- Similarity
  - "Sentences are similar if they share enough content"
  - (Lee & Barzilay 2003) (Dolan&Quirk, 2004)

- Patterns
  - "Some patterns of sentences reveal relations among sentences"
  - (Burger&Ferro, 2005) (Hickl et al., 2006).

Past Works 1

- Lee & Barzilay (2003)
  - Terrorist news
  - Replace dates, numbers, and proper names with generic tokens
  - Similarity metric based on word n-gram overlap.
  - Hierarchical complete-link clustering to the sentences

- Dolan & Quirk (2004)
  - Extracts news stories with HMM from new sites
  - Cluster the articles based on words and publication time
  - (1) edit distance similarity
  - (2) use the first two sentences across documents (with some heuristic constraints)
  - Evaluated by paraphrase generation
  - Found that pattern based extraction give more interesting paraphrase relationships.

Past Works 2

- Burger & Ferro (2005)
  - Paraphrase between leading sentence and title of a new articles
  - SVM document classifier to help reduce the noisiness of data.
  - About three-forth of the generated corpus are genuine entailment pairs.
  - Sometimes judgers are hard to achieve agreement.

- Hickl et al. (2006)
  - Pattern 1: in a text, sentences with a same name entity generally do not entail each other.
  - Pattern 2: sentences linked by discourse connectives (e.g. “even though”, “although”, “otherwise”, and “in contrast”) generally do not entail each other.

Conclusion

- Finding sentence pairs by similarity is against the purpose of textual entailment
  - which is all about learning textual variability
- Using patterns of sentences is more favorable
  - other contexts still need to be explored