Semantic and Multimodal Annotation

CLARA
University of Copenhagen
15-26 August 2011
Susan Windisch Brown

Program: Monday

- Big picture
- Coffee break
- Lexical ambiguity and word sense annotation
- Lunch break
- Introduction to the annotation tool Stamp
- Coffee break
- Hands-on annotation

Semantic annotation

The big picture

Outline

1. Semantics: Definition and significance for NLP
2. Annotation: Its role in NLP
3. Design and implementation of annotation projects

Outline

1. Semantics: Definition and significance for NLP
2. Annotation: Its role in NLP
3. Design and implementation of annotation projects

What do we want computers to be able to do with human languages?

- Machine translation
- Text summarization (medical, business, military)
- Question-answering systems
- Tutoring systems

What we have now

- Search key terms
- Domain-specific interaction with computers (airline reservations, prescription ordering)
- Some question answering
  - Wolfram Alpha
  - IBM's Watson
- Rudimentary translation

Next step

- Need to move beyond surface level of text
  - Words in isolation
  - Part of speech
  - Syntax
- To deeper meaning of sentences, paragraphs, entire texts
- Need semantics

Example: Identifying relevant information

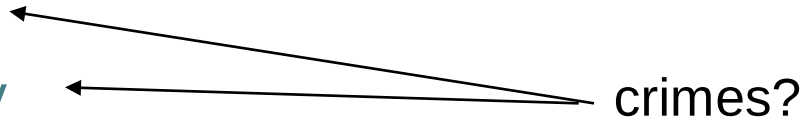
- Query
  - List crimes of John Lee

- Possible Answers
  - **Lee** smuggled heroin across the border
  - **Lee** participated in a bank robbery
  - The arson was committed by **Lee**
  - **Lee** blew up a bus
  - **Lee** stole second base
  - **Lee** stole a kiss from Mary
  - A drug dealer shot **Lee**
  - **Lee** blew up

Example: Identifying relevant information

- Query
  - List crimes of John Lee

- Possible Answers
  - Lee smuggled heroin across the border
  - Lee participated in a bank robbery
  - The arson was committed by Lee
  - Lee blew up a bus
  - Lee stole second base
  - Lee stole a kiss from Mary
  - A drug dealer shot Lee
  - Lee blew up



Senses for *steal*

- **Sense 1: take possession without consent or right**
  Someone stole my wallet on the train.
- **Sense 2: do or achieve something surreptitiously or stealthily**
  I leaned close and stole a kiss.
  She stole a glance at her watch.
- **Sense 3: draw attention or have successful performance**
  He stole the show with his offbeat puns and no-fear analogies.
- **Sense 4: BASEBALL-gain base without hit**
  He might be the last person in baseball you'd expect to steal a base.

Example: Identifying relevant information

- Query
  - List crimes of John Lee

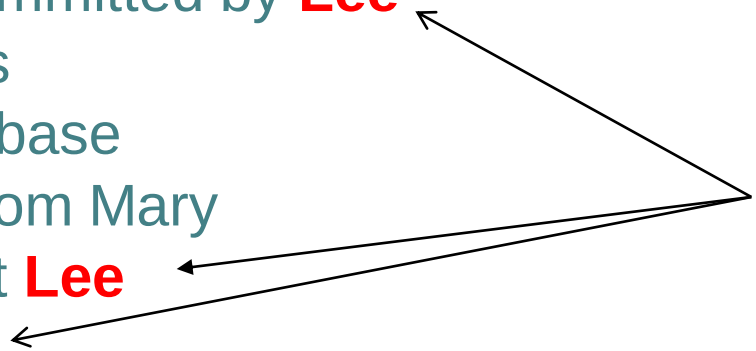
- Possible Answers
  - **Lee** smuggled heroin across the border
  - **Lee** participated in a bank robbery
  - The arson was committed by **Lee**
  - **Lee** blew up a bus
  - **Lee** stole second base
  - **Lee** stole a kiss from Mary
  - A drug dealer shot **Lee**
  - **Lee** blew up

Who actually did these things?

Example: Identifying relevant information

- Query
  - List crimes of John Lee

- Possible Answers
  - **Lee** smuggled heroin across the border
  - **Lee** participated in a bank robbery
  - The arson was committed by **Lee**
  - **Lee** blew up a bus
  - **Lee** stole second base
  - **Lee** stole a kiss from Mary
  - A drug dealer shot **Lee**
  - **Lee** blew up



Semantic roles

- Syntax only gets you so far in answering "Who did what to whom?"

Lee blew up the bus.

Syntax: NP_{SUB} V NP_{OBJ}

The bus blew up.

Syntax: NP_{SUB} V

Semantic roles

- Syntax only gets you so far in answering "Who did what to whom?"

Lee blew up the bus.

Syntax:	NP_{SUB}	V	NP_{OBJ}
Semantics:	Exploder	**REL**	thing exploded

The bus blew up.

Syntax:	NP_{SUB}	V
Semantics:	thing exploded	**REL**

Example: Identifying relevant information

- Query
  - List crimes of John Lee

- Possible Answers
  - ☑ Lee smuggled heroin across the border
  - ☑ Lee participated in a bank robbery
  - ☑ The arson was committed by Lee
  - ☒ Lee stole second base
  - ☒ Lee stole a kiss from Mary
  - ☒ A drug dealer shot Lee
  - ☒ Lee blew up.

Different aspects of semantics

- Word sense
- Thematic (semantic) roles
- Event structure
- Discourse structure
- Sentiment analysis

Event and temporal structure

- “Carlos Rivera, president of the drama club, said last week that all the after-school clubs would host a fundraiser because the school lost its funding for the arts.”
- Temporal aspects of entities
- Anchoring events in time
- Ordering events with respect to one another
- Aspectual and modal predication
- Time ML (Pustejovsky et al., 2003)

Discourse structure

- Rhetorical relations between clauses and sentences
  - Lexically signalled (e.g., *but, because, as a result*)
  - Inferred from proximity (I needed a lot of money for books. My dad offered to lend me $100.)
- RST Corpus (Carlson, Marcu & Okurowski, 2001)
- The Discourse GraphBank (Wolf & Gibson, 2005)
- Penn Discourse TreeBank (Prasad et al., 2008)

Sentiment analysis

- Attitude of speaker/write toward a topic
- Sentiment polarity of document, sentence, feature
- JDPA Sentiment Corpus (Kessler et al., 2010)
  - Negators
  - Neutralizers
  - Committers
  - Intensifiers
  - Entities are annotated for expressed attitude

Semantics in text

- Deeper meaning than
  - A bag of words
  - syntax
- Means of discovering is growing
  - Word sense
  - Semantic roles
  - Event analysis
  - Discourse analysis
  - Sentiment analysis . . .

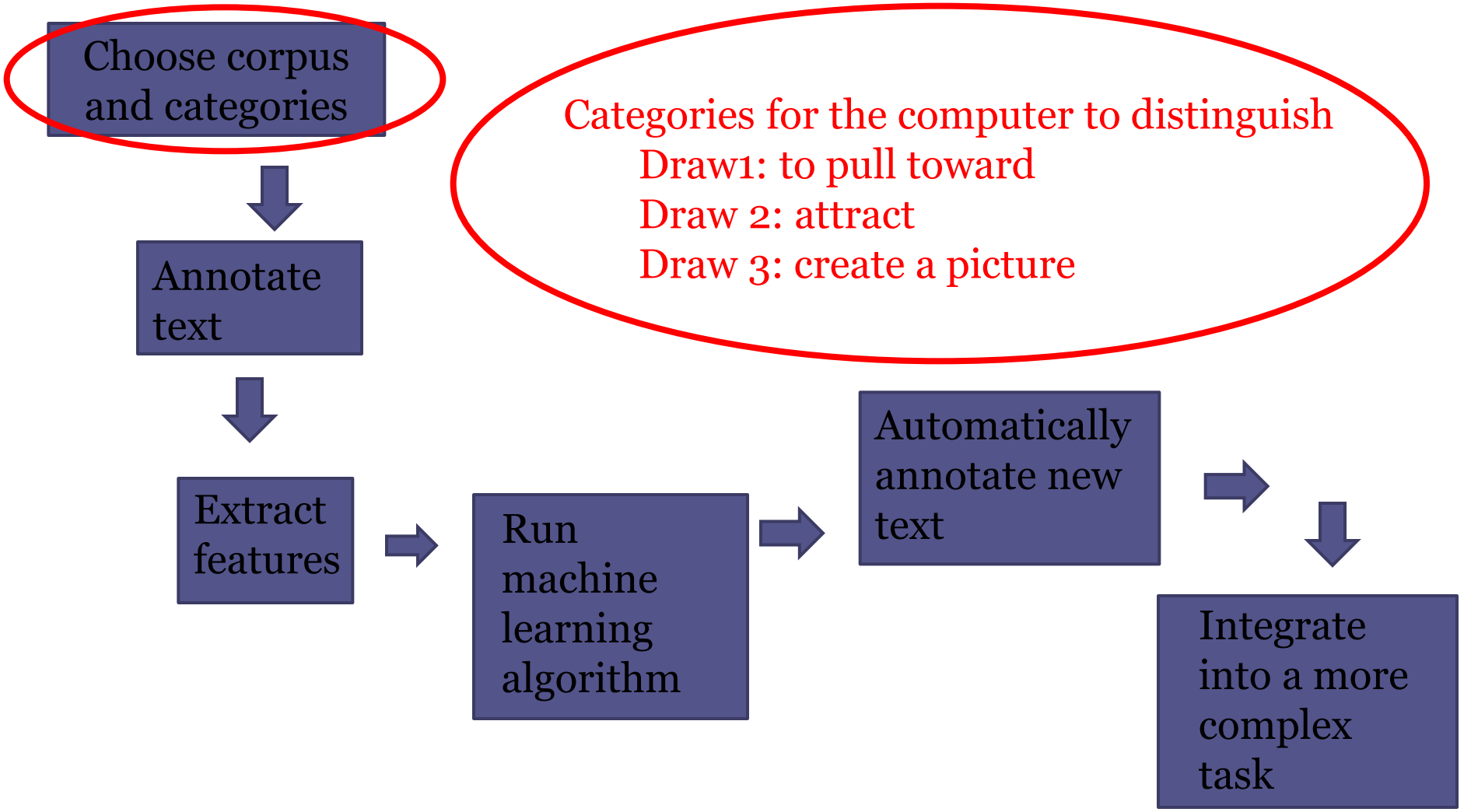
Outline

1. Semantics: Definition and significance for NLP
2. Annotation: Its role in NLP
3. Design and implementation of annotation projects

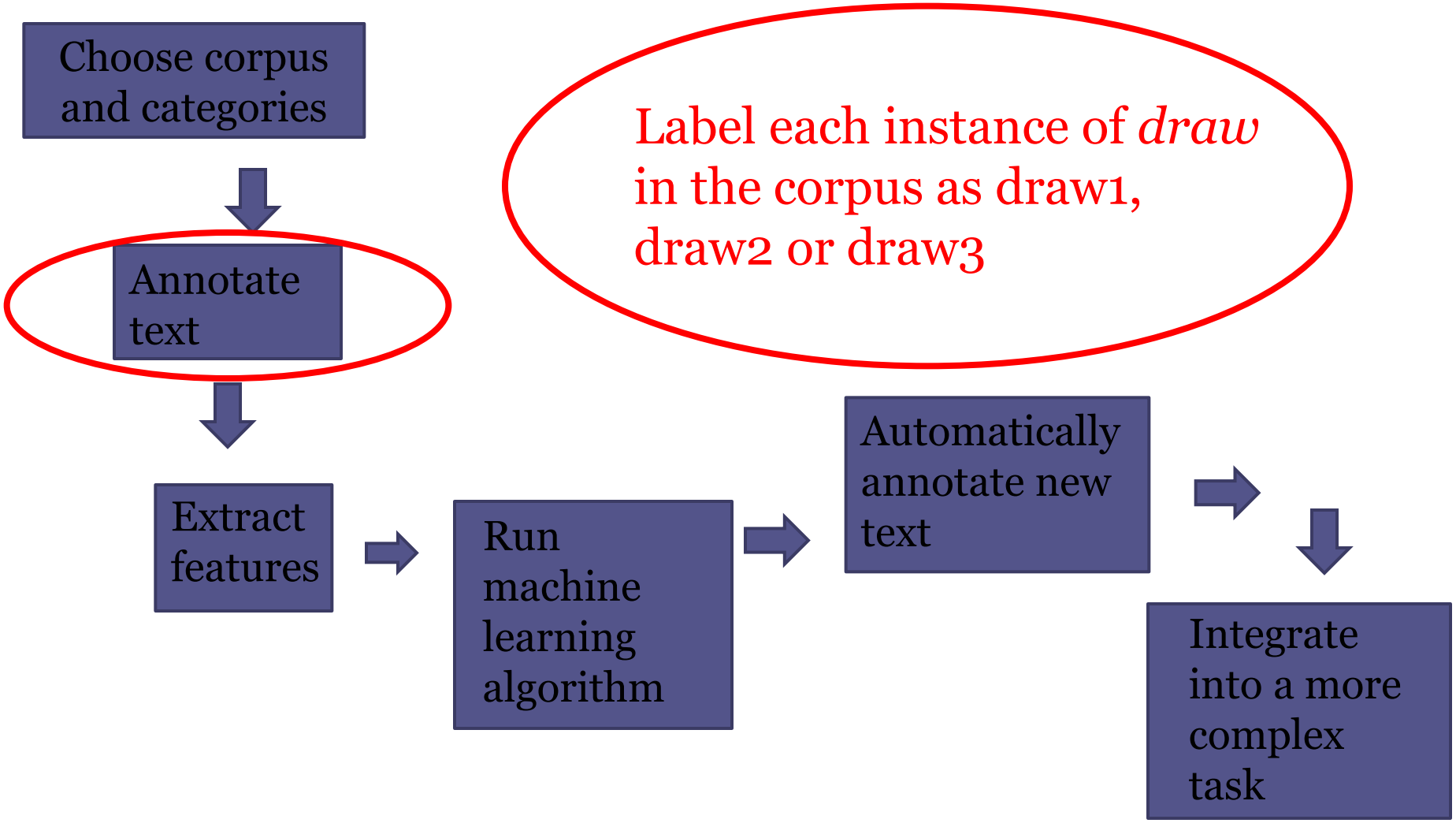
Why annotation?

- Real-world data for linguistics research (corpus linguistics)
- Training data for supervised machine learning

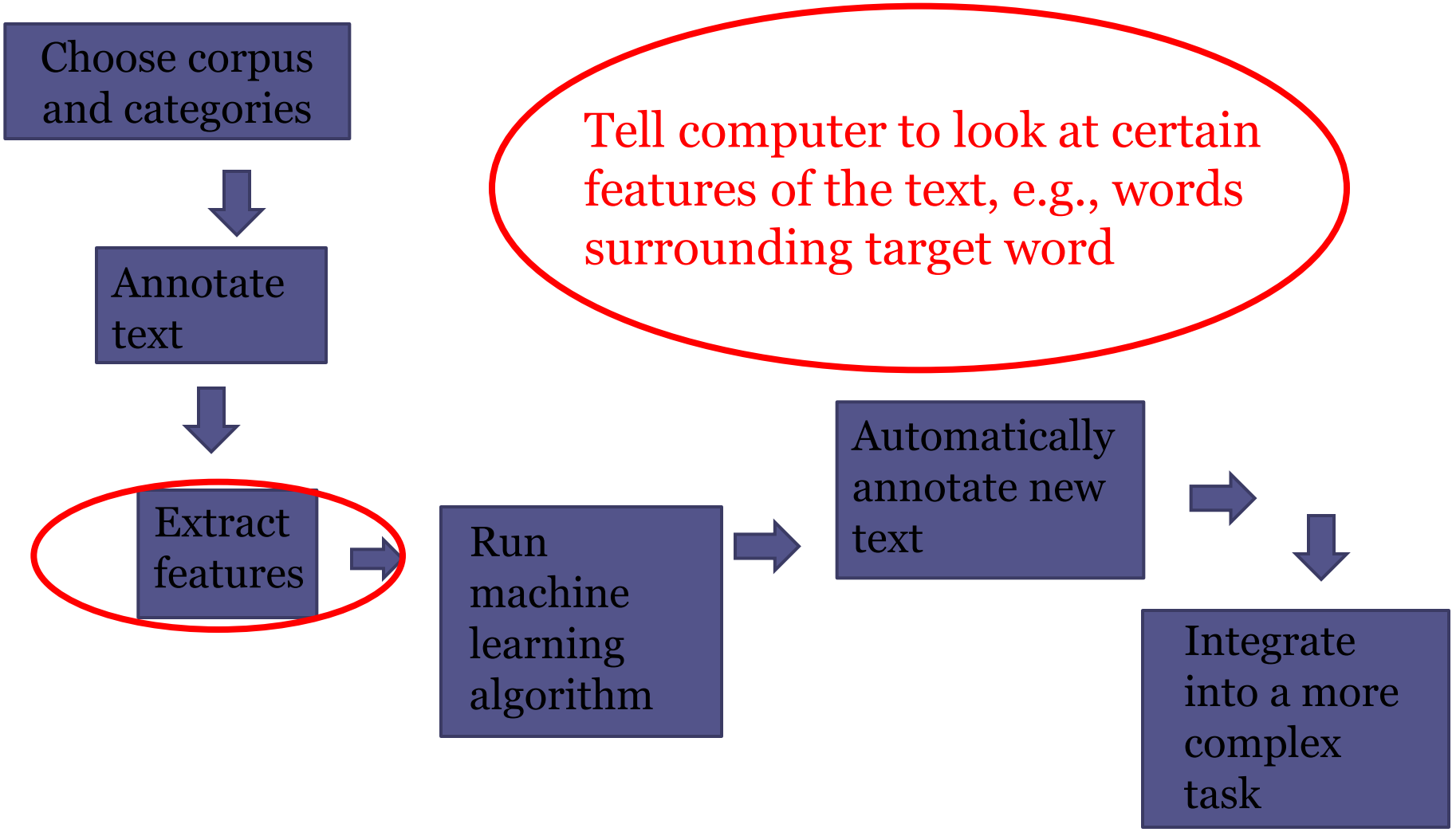
Supervised machine learning for NLP



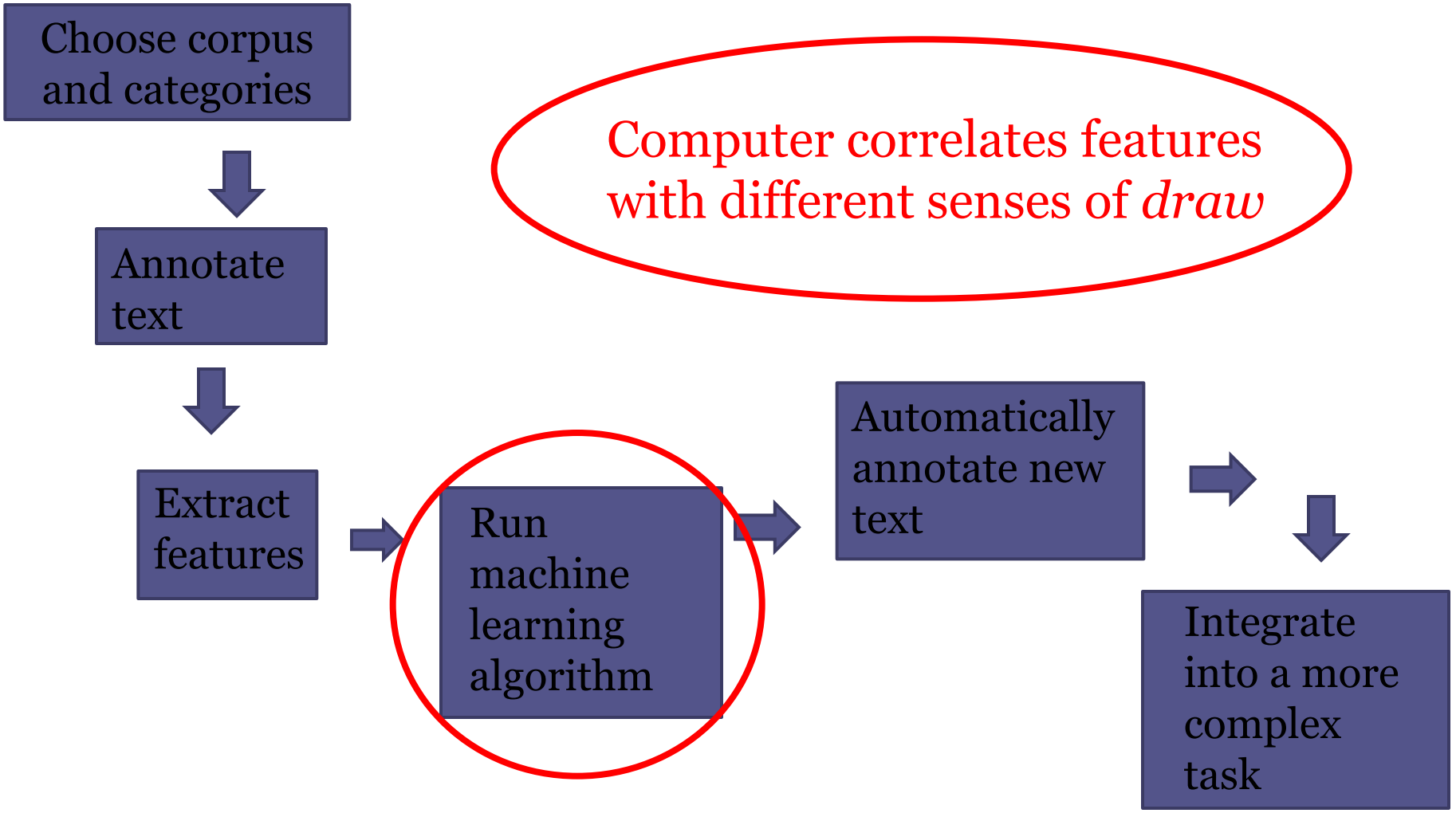
Supervised machine learning for NLP



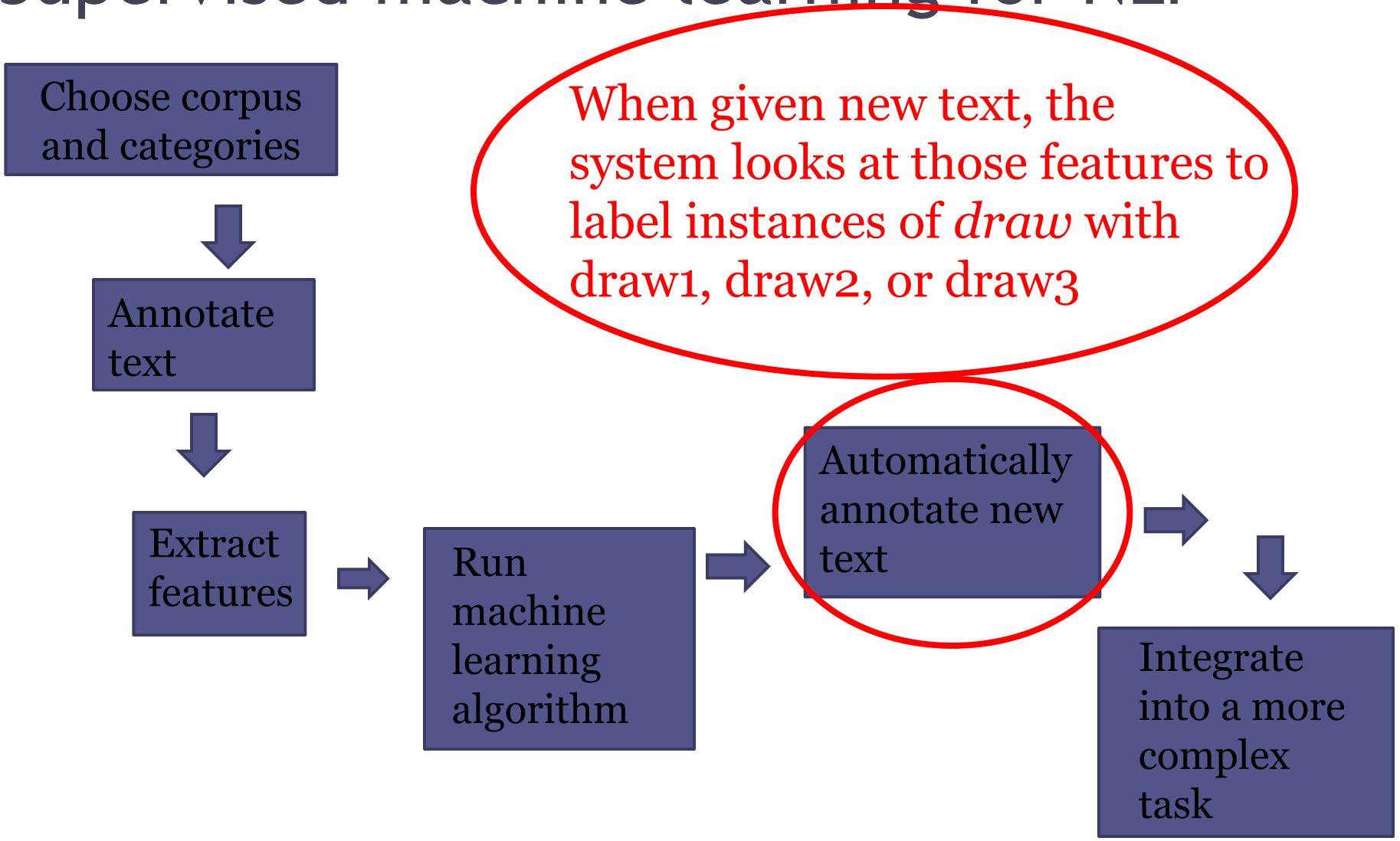
Supervised machine learning for NLP



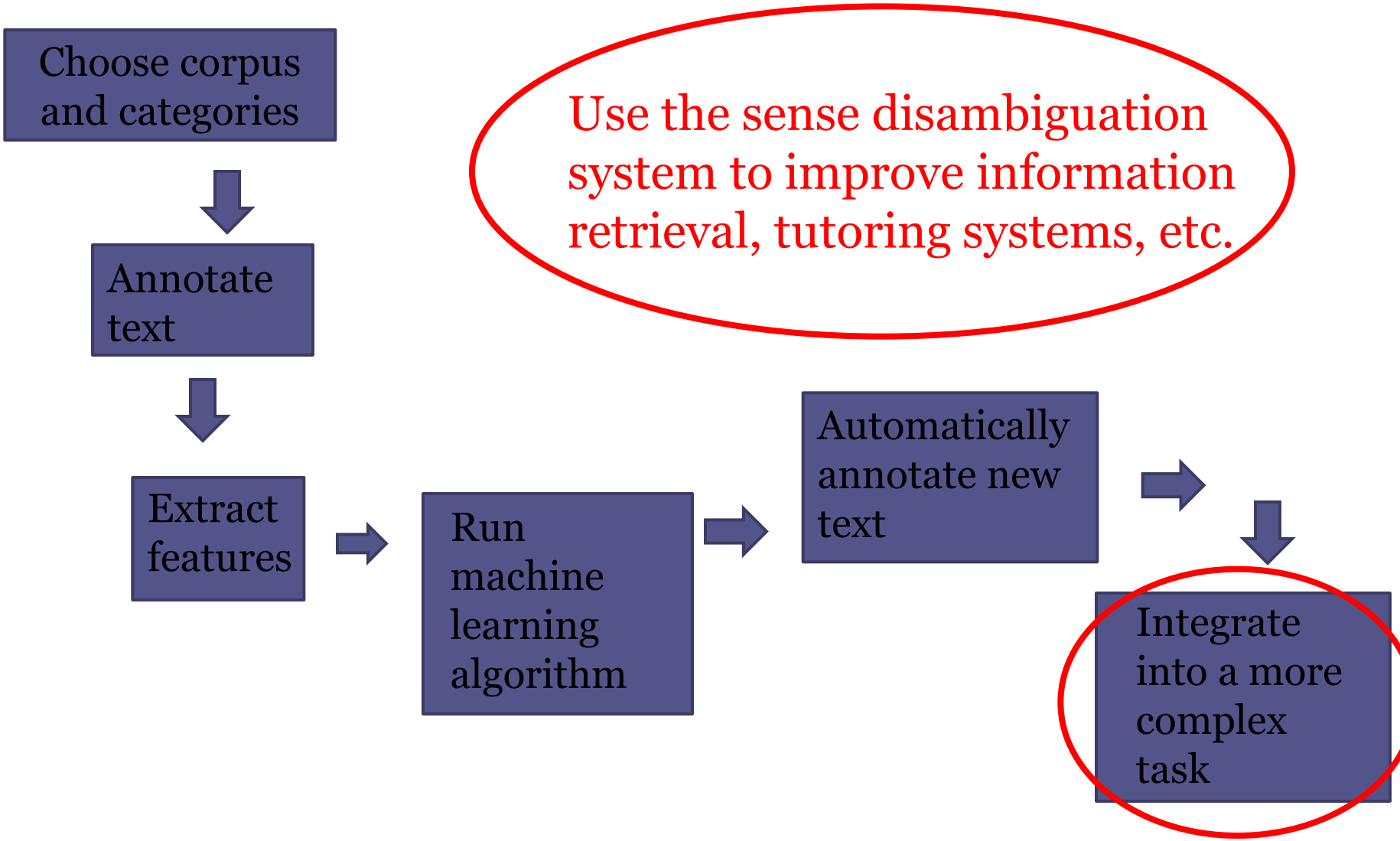
Supervised machine learning for NLP



Supervised machine learning for NLP



Supervised machine learning for NLP



Outline

1. Semantics: Definition and significance for NLP
2. Annotation: Its role in NLP
3. Design and implementation of annotation projects
    - Corpus choice
    - Categories/classes choice
    - Task design
    - Annotation reliability
    - Annotation efficiency

Outline

1. Semantics: Definition and significance for NLP
2. Annotation: Its role in NLP
3. Design and implementation of annotation projects
    - Corpus choice
    - Categories/classes choice
    - Task design
    - Annotation reliability
    - Annotation efficiency

The more data the better

- To see patterns, ML algorithm needs many instances of those patterns
- Need examples of all the categories in order to learn how to distinguish them
  - Skewed data
  - Rare categories
- Active learning can reduce the amount needed

The more varied the data, the better

- Must have examples in order to generalize
  - Of categories
  - Of contexts
- Different genres
  - Fiction, non-fiction
  - Conversation, blogs, broadcast news
- Different domains
  - Financial
  - Current events
  - Medical

Desired features can influence corpus choice

- Is the corpus tagged with parts of speech?
- How it can help
  - The sail tore./We will sail at dawn.
    noun verb
  - She left at 8:00.
    prep noun
  - She left her diamonds to her daughter.
    det noun

Syntactic structure

- Transitive vs. intransitive
  - She left at 8:00.
  - She left her diamonds to her daughter.
- Is there a PP in the sentence?
  - The paint ran.
  - The boy ran to the store.
- If so, which preposition?
  - The paint ran down her face.
  - The boy ran around the park.

Corpus choice

- Part of speech tagged?
- Syntactically parsed?
- Varied (balanced)?
  - Different domains of non-fiction
  - Fiction
  - Transcribed speech

Outline

1. Semantics: Definition and significance for NLP
2. Annotation: Its role in NLP
3. Design and implementation of annotation projects
    - Corpus choice
    - Categories/classes choice
    - Task design
    - Annotation reliability
    - Annotation efficiency

Category generality

- How broad or narrow will the labels be?
- Word sense
  - WordNet: 36 senses for the verb *draw*
  - OntoNotes: 11 senses for the verb *draw*
  - PropBank: 3 senses for the verb *draw*
- Semantic roles
  - Scott ate with a fork.

    agent instrument

    eater utensil

Other considerations

- Theoretical
  - Demonstrate a particular theory
  - Follow an established model
- Practical
  - Can annotators distinguish between the labels?
  - How much time does it take to annotate with the labels?

Outline

1. Semantics: Definition and significance for NLP
2. Annotation: Its role in NLP
3. Design and implementation of annotation projects
    - Corpus choice
    - Categories/classes choice
    - Task design
    - Annotation reliability
    - Annotation efficiency

Organization of annotation tasks

- Goals
  - Speed
  - Consistency
- Limit choices
- Limit switching between sets of choices

Divide into tasks based on specific words

- All and only instances of *draw* in the same task
- Annotator becomes familiar with the choices
- And familiar with the contexts
- Excellent for word sense and semantic role labeling
- Not feasible for certain types of annotation
  - Event course annotation
  - Discourse annotation
  - Ontology annotation with many rare terms

Divide into categories (two-stage annotation)

- Annotate for broad categories first
- Collect all instances of one broad category into a task
- Then annotate with finer grained categories
- Ontology annotation, such as medical ontologies

Outline

1. Semantics: Definition and significance for NLP
2. Annotation: Its role in NLP
3. Design and implementation of annotation projects
    - Corpus choice
    - Categories/classes choice
    - Task design
    - Annotation reliability
    - Annotation efficiency

Ensuring reliable annotation

- Machine learning requires clear patterns to work
- Consistent training, with written guidelines
- Multiple annotators tagging the same material
  - Less idiosyncratic
  - Can calculate the level of agreement between annotators
  - Indicates how reliable the annotators and/or the annotation scheme is
- Adjudication

Types of multiple annotation

- Concensus annotation
  - Reliable but slow
- Independent multiple annotation
  - Double annotation is common
  - Crowd annotation is growing
- If independent, need an adjudication method
  - Specially trained adjudicator
  - Algorithm using most common choice, annotator statistics, etc.

Outline

1. Semantics: Definition and significance for NLP
2. Annotation: Its role in NLP
3. Design and implementation of annotation projects
    - Corpus choice
    - Categories/classes choice
    - Task design
    - Annotation reliability
    - Annotation efficiency

Improving annotation efficiency

- Making annotation faster and cheaper
- Crowdsourcing
- Active learning
- Other semi-supervised annotation methods

Crowdsourcing: Games with a Purpose

- Web-based games
- Free labor
- High up-front costs
- Only feasible for very large projects and simple tasks
- NLP and semantic annotation examples
  - Phrase Detectives (coreference annotation; http://anawiki.essex.ac.uk/phrasedetectives/index.php)
  - ESP Game (image recognition; http://www.gwap.com/gwap/gamesPreview/espgame/)

Crowdsourcing: Wisdom of the Crowd

- Use volunteers
  - Altruism and/or interest in using the resulting resource
- Difficult to find and maintain a corps of volunteers; best success with:
  - Open-source resources
  - Domain-specific resources
- Feasible for non-time-sensitive projects
- Language-related examples
  - Oxford English Dictionary
  - Open Mind Initiative (word relations, word sense)

Crowdsourcing: Mechanical Turk

- Clearinghouse for web-based labor
- Pros
  - Very low cost
  - Large pool of laborers
  - Infrastructure for task creation and management
  - Screen annotators
- Cons
  - Laborers have little expertise
  - Money incentive promotes cheating
- Semantic annotation examples
  - Word similarity
  - Event ordering
  - Word sense annotation

Improving annotation efficiency: Active learning

- Find most informative examples to manually annotate
- Manually annotate a small “seed” set of instances
- Train a classifier on them
- Have the classifier choose next instances to annotate
- Dramatically reduces the amount of annotation

Active learning: missed cluster effect

- Skewed data can be a problem
  - E.g., word senses
- Hand select seed set (Tomanek et al. 2009)
- Automatic method using language modeling (Dligach and Palmer, 2011)

Improving annotation efficiency: Other types of semi-automatic annotation

- Building off existing resources
- Semlink
  - Used a corpus annotated with PropBank thematic roles
  - Created a mapping from PB roles and rolesets to VerbNet roles and classes
  - Applied to corpus
  - Hand corrected