

# Polysemy in Verbs: Systematic Relations between Senses and Their Effect on Annotation

Anna Rumshisky  
Olga Batiukova



Brandeis University

# *Annotation Task*

- Problem:
  - Word sense disambiguation
- In particular:
  - Disambiguation of polysemous verbs
- Subproblem:
  - Senses distinguished predominantly through semantics of the arguments

# *Word Sense Annotation*

- Semantically annotated corpora are routinely developed for the training and testing of automatic sense detection and induction systems
  - SemCor (Landes et al., 1998)
  - OntoNotes (Hovy et al., 2006)
  - PropBank (Palmer et al. 2005)
  - Framenet Corpus (Ruppenhofer et al., 2006)

# *Task Stages*

- Stage 1. Data set construction
  - sense inventory construction
  - data extraction/collection
  - data preprocessing
- Stage 2. Annotation
  - annotating examples
  - checking agreement

# *Task Stages*

- **Stage 1. Data set construction**
  - sense inventory construction
  - data extraction/collection
  - data preprocessing
- **Stage 2. Annotation**
  - annotating examples
  - checking agreement

# *Task Stages*

- Stage 1. Data set construction
  - sense inventory construction
  - data extraction/collection
  - data preprocessing
- Stage 2. Annotation
  - annotating examples
  - checking agreement

# *Task Stages*

- Stage 1. Data set construction
  - sense inventory construction
  - data extraction/collection
  - data preprocessing
- Stage 2. Annotation
  - annotating examples
  - checking agreement

# *Task Stages*

- Stage 1. Data set construction
  - sense inventory construction
  - data extraction/collection
  - data preprocessing
- Stage 2. Annotation
  - annotating examples
  - checking agreement



# *Task Stages*

- Stage 1. Data set construction
  - sense inventory construction
  - data extraction/collection
  - data preprocessing
- **Stage 2. Annotation**
  - annotating examples
  - checking agreement

# *Task Stages*

- Stage 1. Data set construction
  - sense inventory construction
  - data extraction/collection
  - data preprocessing
- Stage 2. Annotation
  - **annotating examples**
  - checking agreement

# *Task Stages*

- Stage 1. Data set construction
  - sense inventory construction
  - data extraction/collection
  - data preprocessing
- Stage 2. Annotation
  - annotating examples
  - **checking agreement**

# *Task Stages*

- Stage 1. Data set construction

- sense inventory construction
- data extraction/collection
- data preprocessing

- Stage 2. Annotation

- annotating examples
- checking agreement



# *Focus of Our Study*

- What happens under the magnifying glass?
  - What are the sources of disagreement in annotation?
  - There are plenty of difficult cases. What do the annotators do in those cases?
- We would like to look at
  - The effects relations between senses have on decisions made by the annotators and annotation error
  - Some common traps and pitfalls in design of sense inventories

# *Talk Outline*

- Introduction
- Motivation
- Annotation task
  - Data set
  - Annotation guidelines
- Design of sense inventories
- Relations between senses
- Annotation decisions
- Conclusions

# *Sense Differentiation for Verbs*

Establishing a set of senses for polysemous verbs is notoriously hard, since the meaning

- is very often determined in composition
- depends to the same extent on semantics of the arguments as on the base meaning of the verb itself
- often, constellations of related meanings exist and the meanings are extended “on the fly”

# *Constellations of related meanings*

to drive = ?



# *Constellations of related meanings*

physically urge an animal to go in a direction  
drive cattle, horses

# *Constellations of related meanings*

physically urge an animal to go in a direction  
drive cattle, horses

force a vessel to move in a direction  
storms and tides driving boats ashore

# *Constellations of related meanings*

move or provide power for the motion  
of a mechanism  
steam driving the engine

force a vessel to move in a direction  
storms and tides driving boats ashore

physically urge an animal to go in a direction  
drive cattle, horses

# *Constellations of related meanings*

- operate a vehicle, controlling its motion
- travel in a vehicle a certain distance
- transport something or someone

drive a car

drive twenty miles

drive a friend home

move or provide power for the motion  
of a mechanism

steam driving the engine

force a vessel to move in a direction

storms and tides driving boats ashore

physically urge an animal to go in a direction

drive cattle, horses

# *Constellations of related meanings*

- operate a vehicle, controlling its motion
- travel in a vehicle a certain distance
- transport something or someone

drive a car

drive twenty miles

drive a friend home

move or provide power for the motion  
of a mechanism

steam driving the engine

force a vessel to move in a direction

storms and tides driving boats ashore

physically urge an animal to go in a direction

drive cattle, horses

force an adversary to leave

competitors away, enemy off the battlefield

# *Constellations of related meanings*

- operate a vehicle, controlling its motion
- travel in a vehicle a certain distance
- transport something or someone

drive a car

drive twenty miles

drive a friend home

move or provide power for the motion  
of a mechanism

steam driving the engine

force a vessel to move in a direction

storms and tides driving boats ashore

physically urge an animal to go in a direction

drive cattle, horses

force an adversary to leave

competitors away, enemy off the battlefield

cause to enter a state or force into an activity

drive into poverty, into despair, to commit crimes

# *Constellations of related meanings*

- operate a vehicle, controlling its motion
- travel in a vehicle a certain distance
- transport something or someone

drive a car  
drive twenty miles  
drive a friend home

move or provide power for the motion  
of a mechanism

steam driving the engine

force a vessel to move in a direction

storms and tides driving boats ashore

physically urge an animal to go in a direction

drive cattle, horses

push a sharp object into another object

drive a nail, a stake into the ground

force an adversary to leave

competitors away, enemy off the battlefield

cause to enter a state or force into an activity

drive into poverty, into despair, to commit crimes

# Constellations of related meanings

- operate a vehicle, controlling its motion
- travel in a vehicle a certain distance
- transport something or someone

drive a car  
drive twenty miles  
drive a friend home

move or provide power for the motion  
of a mechanism

steam driving the engine

force a vessel to move in a direction

storms and tides driving boats ashore

physically urge an animal to go in a direction

drive cattle, horses

push a sharp object into another object

drive a nail, a stake into the ground

strike or throw an object of play

drive the ball into the corner

force an adversary to leave

competitors away, enemy off the battlefield

cause to enter a state or force into an activity

drive into poverty, into despair, to commit crimes



# *Motivation*

- Different ambiguities require different kinds of contextual information to be resolved
- Sense-tagged corpora typically do not address the question of what factors allow the speakers to identify a particular sense
  - impossible to evaluate contribution of different factors (different contextual cues) to sense differentiation
- Consequently, it is difficult to perform adequate error analysis for automatic sense detection (WSD/WSI) systems

# *Sources of Sense Differentiation*

Within the scope of a sentence, there are two main sources of sense differentiation for verbs:

- syntactic frame
- semantics of the arguments

# *Sources of Sense Differentiation*

- Syntactic frame

The authorities denied the *visa* to the prime minister (refuse to give)

The authorities denied the *attack* (proclaim false)

- Semantics of the arguments

The general fired four *rounds* (shoot)

The general fired four *lieutenant-colonels* (dismiss)

The customer will absorb this *tax* (pay)

The customer will absorb this *information* (learn)

# *Sources of Sense Differentiation*

- Typically, it is easier for people to distinguish senses that are linked to specific syntactic patterns:

Achilles, denied his *attack*, had to stay in camp, brooding

- ditransitive construction makes sense recognition easy

# *Sources of Sense Differentiation*

- When sense distinctions are linked to semantics of the verb's arguments, things are often less clear
  - e.g. *Security camera footage showed the suspect getting into a car.*  
(pictorially represent)
  - e.g. *The study showed a dependency between X and Y*  
(demonstrate by reasoning)
  - e.g. *The diagram shows a dependency between X and Y*  
-- both?

## *Case Study: Sense Distinctions linked to Semantics of a Single Argument*

- Select the verbs that have sense distinctions that can be detected by looking at semantics of a single argument
- Use sense inventories that contain the relevant senses

# *Talk Outline*

- Introduction
- Motivation
- Annotation task
  - Data set
  - Annotation guidelines
- Design of sense inventories
- Relations between senses
- Annotation decisions
- Conclusions

# *Annotation Task (Setup)*

- Standard sense-annotation setup, similar to Senseval Lexical Sample tasks
  - the target word is disambiguated by the annotators, one sense is assigned to each occurrence
  - annotators are given a context for each occurrence (sentence)



# *Data Set Construction*

20 (verb, grammatical relation) pairs

- **dobj**: absorb, acquire, admit, assume, claim, conclude, cut, deny, dictate, drive, edit, enjoy, fire, grasp, know, launch
- **subj**: explain, fall, lead
- **iobj\_with**: meet

# *Verb Selection*

- British National Corpus
- Sketch Engine (Kilgarriff et al., 2004)
  - lexicographic tool that gives a ranked listing of words that co-occur with the target in the specified grammatical relation
- Sense inventory was created for each (verb, relation) pair using a modification of the CPA technique

# *Corpus Pattern Analysis (CPA) (P. Hanks)*

- lexicographic technique that aims to capture norms of usage for individual words using full context specification
- including argument structure, minor categories (locatives, adjuncts, etc.), subphrasal cues (genitives, partitives, bare plural/determiner, infinitivals, negatives, etc.)
- semantics of the arguments specified in terms of basic semantic features (PhysObj, Abstract, Event) or *lexical sets* (collections of lexical items)

# CPA Patterns for “absorb”

The customer will absorb the cost.

Mr. Clinton wanted energy producers to absorb the tax.

PATTERN 1: [[Abstract] | [Person]] absorb [[Asset]]

They quietly absorbed all this new information.

Meanwhile, I absorbed a fair amount of management skills.

PATTERN 2: [[Person]] absorb {(QUANT)} [[Abstract= Concept]}

Water easily absorbs heat.

The SO<sub>2</sub> cloud absorbs solar radiation.

PATTERN 3: [[PhysObj] | [Substance]] absorb [[Energy]]

The villagers were far too absorbed in their own affairs.

He became completely absorbed in struggling for survival.

PATTERN 4: [[Person]] {be | become} absorbed {in [[Activity]][[Abstract]]}

# Sketch Engine

[Home](#) [Concordance](#) [Word Sketch](#) [Thesaurus](#) [Sketch-Diff](#)

dictate **bnc freq = 1264**

<b>object</b>	<b>646</b>	<b>4.2</b>	<b>subject</b>	<b>520</b>	<b>6.8</b>	<b>modifier</b>	<b>175</b>	<b>2.2</b>
letter	<a href="#">25</a>	20.51	circumstance	<a href="#">14</a>	18.39	otherwise	<a href="#">11</a>	25.83
pace	<a href="#">12</a>	19.87	consideration	<a href="#">10</a>	15.82	partly	<a href="#">9</a>	24.2
term	<a href="#">26</a>	19.1	custom	<a href="#">7</a>	15.19	largely	<a href="#">7</a>	18.48
choice	<a href="#">14</a>	16.57	prudence	<a href="#">3</a>	14.77	also	<a href="#">14</a>	16.25
policy	<a href="#">21</a>	15.33	tradition	<a href="#">7</a>	12.64	often	<a href="#">8</a>	15.4
shape	<a href="#">9</a>	13.71	sense	<a href="#">10</a>	12.49	externally	<a href="#">2</a>	13.01
caution	<a href="#">4</a>	12.51	conscience	<a href="#">4</a>	12.41	still	<a href="#">7</a>	12.6
intifada	<a href="#">2</a>	10.95	availability	<a href="#">4</a>	11.56	only	<a href="#">7</a>	11.89
content	<a href="#">5</a>	10.51	logic	<a href="#">4</a>	11.47	clearly	<a href="#">4</a>	11.89
form	<a href="#">10</a>	9.15	wisdom	<a href="#">3</a>	9.93	necessarily	<a href="#">3</a>	11.49
tactic	<a href="#">3</a>	9.08	arithmetic	<a href="#">2</a>	9.22	probably	<a href="#">4</a>	11.37
action	<a href="#">8</a>	9.03	Quinn	<a href="#">2</a>	8.85	even	<a href="#">5</a>	11.04
passage	<a href="#">4</a>	8.86	paradigm	<a href="#">2</a>	8.44	merely	<a href="#">3</a>	10.96
memoirs	<a href="#">2</a>	8.86	function	<a href="#">5</a>	8.36	partially	<a href="#">2</a>	10.15
geometry	<a href="#">2</a>	8.67	convenience	<a href="#">2</a>	8.28	already	<a href="#">4</a>	9.83
pattern	<a href="#">6</a>	8.4	fashion	<a href="#">3</a>	7.92	ultimately	<a href="#">2</a>	9.81
format	<a href="#">3</a>	8.25	Brussels	<a href="#">2</a>	7.87	always	<a href="#">4</a>	9.26
need	<a href="#">7</a>	8.14	need	<a href="#">6</a>	7.8	virtually	<a href="#">2</a>	8.89
treatment	<a href="#">5</a>	7.84	interest	<a href="#">7</a>	7.78	effectively	<a href="#">2</a>	8.45
answer	<a href="#">4</a>	7.64	policy	<a href="#">7</a>	7.67	both	<a href="#">3</a>	8.27

# Sketch Engine

[Home](#) [Concordance](#) [Word Sketch](#) [Thesaurus](#) [Sketch-Diff](#)

dictate **bnc freq = 1264**

<b>object</b>	<b>646</b>	<b>4.2</b>	<b>subject</b>	<b>520</b>	<b>6.8</b>	<b>modifier</b>	<b>175</b>	<b>2.2</b>
letter	<u>25</u>	20.51	circumstance	<u>14</u>	18.39	otherwise	<u>11</u>	25.83
pace	<u>12</u>	19.87	consideration	<u>10</u>	15.82	partly	<u>9</u>	24.2
term	<u>26</u>	19.1	custom	<u>7</u>	15.19	largely	<u>7</u>	18.48
choice	<u>14</u>	16.57	prudence	<u>3</u>	14.77	also	<u>14</u>	16.25
policy	<u>21</u>	15.33	tradition	<u>7</u>	12.64	often	<u>8</u>	15.4
shape	<u>9</u>	13.71	sense	<u>10</u>	12.49	externally	<u>2</u>	13.01
caution	<u>4</u>	12.51	conscience	<u>4</u>	12.41	still	<u>7</u>	12.6
intifada	<u>2</u>	10.95	availability	<u>4</u>	11.56	only	<u>7</u>	11.89
content	<u>5</u>	10.51	logic	<u>4</u>	11.47	clearly	<u>4</u>	11.89
form	<u>10</u>	9.15	wisdom	<u>3</u>	9.93	necessarily	<u>3</u>	11.49
tactic	<u>3</u>	9.08	arithmetic	<u>2</u>	9.22	probably	<u>4</u>	11.37
action	<u>8</u>	9.03	Quinn	<u>2</u>	8.85	even	<u>5</u>	11.04
passage	<u>4</u>	8.86	paradigm	<u>2</u>	8.44	merely	<u>3</u>	10.96
memoirs	<u>2</u>	8.86	function	<u>5</u>	8.36	partially	<u>2</u>	10.15
geometry	<u>2</u>	8.67	convenience	<u>2</u>	8.28	already	<u>4</u>	9.83
pattern	<u>6</u>	8.4	fashion	<u>3</u>	7.92	ultimately	<u>2</u>	9.81
format	<u>3</u>	8.25	Brussels	<u>2</u>	7.87	always	<u>4</u>	9.26
need	<u>7</u>	8.14	need	<u>6</u>	7.8	virtually	<u>2</u>	8.89
treatment	<u>5</u>	7.84	interest	<u>7</u>	7.78	effectively	<u>2</u>	8.45
answer	<u>4</u>	7.64	policy	<u>7</u>	7.67	both	<u>3</u>	8.27



# Sketch Engine

[Home](#) [Concordance](#) [Word Sketch](#) [Thesaurus](#) [Sketch-Diff](#)

dictate **bnc freq = 1264**

<b>object</b>	<b>646</b>	<b>4.2</b>	<b>subject</b>	<b>520</b>	<b>6.8</b>	<b>modifier</b>	<b>175</b>	<b>2.2</b>
letter	<u>25</u>	20.51	circumstance	<u>14</u>	18.39	otherwise	<u>11</u>	25.83
pace	<u>12</u>	19.87	consideration	<u>10</u>	15.82	partly	<u>9</u>	24.2
term	<u>26</u>	19.1	custom	<u>7</u>	15.19	largely	<u>7</u>	18.48
choice	<u>14</u>	16.57	prudence	<u>3</u>	14.77	also	<u>14</u>	16.25
policy	<u>21</u>	15.33	tradition	<u>7</u>	12.64	often	<u>8</u>	15.4
shape	<u>9</u>	13.71	sense	<u>10</u>	12.49	externally	<u>2</u>	13.01
caution	<u>4</u>	12.51	conscience	<u>4</u>	12.41	still	<u>7</u>	12.6
intifada	<u>2</u>	10.95	availability	<u>4</u>	11.56	only	<u>7</u>	11.89
content	<u>5</u>	10.51	logic	<u>4</u>	11.47	clearly	<u>4</u>	11.89
form	<u>10</u>	9.15	wisdom	<u>3</u>	9.93	necessarily	<u>3</u>	11.49
tactic	<u>3</u>	9.08	arithmetic	<u>2</u>	9.22	probably	<u>4</u>	11.37
action	<u>8</u>	9.03	Quinn	<u>2</u>	8.85	even	<u>5</u>	11.04
passage	<u>4</u>	8.86	paradigm	<u>2</u>	8.44	merely	<u>3</u>	10.96
memoirs	<u>2</u>	8.86	function	<u>5</u>	8.36	partially	<u>2</u>	10.15
geometry	<u>2</u>	8.67	convenience	<u>2</u>	8.28	already	<u>4</u>	9.83
pattern	<u>6</u>	8.4	fashion	<u>3</u>	7.92	ultimately	<u>2</u>	9.81
format	<u>3</u>	8.25	Brussels	<u>2</u>	7.87	always	<u>4</u>	9.26
need	<u>7</u>	8.14	need	<u>6</u>	7.8	virtually	<u>2</u>	8.89
treatment	<u>5</u>	7.84	interest	<u>7</u>	7.78	effectively	<u>2</u>	8.45
answer	<u>4</u>	7.64	policy	<u>7</u>	7.67	both	<u>3</u>	8.27

# Sketch Engine

[Home](#) [Concordance](#) [Word Sketch](#) [Thesaurus](#) [Sketch-Diff](#)

dictate **bnc freq = 1264**

<b>object</b>	<b>646</b>	<b>4.2</b>	<b>subject</b>	<b>520</b>	<b>6.8</b>	<b>modifier</b>	<b>175</b>	<b>2.2</b>
letter	<u>25</u>	20.51	circumstance	<u>14</u>	18.39	otherwise	<u>11</u>	25.83
pace	<u>12</u>	19.87	consideration	<u>10</u>	15.82	partly	<u>9</u>	24.2
term	<u>26</u>	19.1	custom	<u>7</u>	15.19	largely	<u>7</u>	18.48
choice	<u>14</u>	16.57	prudence	<u>3</u>	14.77	also	<u>14</u>	16.25
policy	<u>21</u>	15.33	tradition	<u>7</u>	12.64	often	<u>8</u>	15.4
shape	<u>9</u>	13.71	sense	<u>10</u>	12.49	externally	<u>2</u>	13.01
caution	<u>4</u>	12.51	conscience	<u>4</u>	12.41	still	<u>7</u>	12.6
intifada	<u>2</u>	10.95	availability	<u>4</u>	11.56	only	<u>7</u>	11.89
content	<u>5</u>	10.51	logic	<u>4</u>	11.47	clearly	<u>4</u>	11.89
form	<u>10</u>	9.15	wisdom	<u>3</u>	9.93	necessarily	<u>3</u>	11.49
tactic	<u>3</u>	9.08	arithmetic	<u>2</u>	9.22	probably	<u>4</u>	11.37
action	<u>8</u>	9.03	Quinn	<u>2</u>	8.85	even	<u>5</u>	11.04
passage	<u>4</u>	8.86	paradigm	<u>2</u>	8.44	merely	<u>3</u>	10.96
memoirs	<u>2</u>	8.86	function	<u>5</u>	8.36	partially	<u>2</u>	10.15
geometry	<u>2</u>	8.67	convenience	<u>2</u>	8.28	already	<u>4</u>	9.83
pattern	<u>6</u>	8.4	fashion	<u>3</u>	7.92	ultimately	<u>2</u>	9.81
format	<u>3</u>	8.25	Brussels	<u>2</u>	7.87	always	<u>4</u>	9.26
need	<u>7</u>	8.14	need	<u>6</u>	7.8	virtually	<u>2</u>	8.89
treatment	<u>5</u>	7.84	interest	<u>7</u>	7.78	effectively	<u>2</u>	8.45
answer	<u>4</u>	7.64	policy	<u>7</u>	7.67	both	<u>3</u>	8.27



# *Senses for dictate, dobj*

- (1) *verbalize to be recorded* (letter, passage, memoir)
- (2) *determine the character of or serve as a motivation for* (terms, policy, etc.)

# *Sense Inventory Construction*

- Sense inventory for each verb cross-checked against
  - WordNet, PropBank, Merriam-Webster, Oxford English dictionary, and existing correspondences in FrameNet, OntoNotes, and CPA Patterns.
- We performed test annotations of 100 instances, with sense inventories additionally modified upon examining the results of annotation.

# *Sense inventory for acquire, dobj*

1. *take on certain characteristics*

e.g. importance, meaning; also: reputation

2. *learn*

e.g. language, manners, knowledge, skill

3. *purchase or become the owner of property*

e.g. land, stocks, business

4. *become associated with something, often newly brought into being*

e.g. cities acquiring new jobs

# *Sense inventory for fall, subj*

1. physically drop; move or extend downward

e.g. physical objects falling; also: extending downward: rainbow, light, hair

2. decrease (e.g. price, inflation, profits, attendance)

3. lose power or suffer a defeat (e.g. Roman Empire, Napoleon, France)

4. for a state to come or commence (e.g. darkness, silence, night)

5. be categorized or fall into a range

e.g. cases falling into categories, into types, into a range

6. be associated with or get assigned to a person, location, or time

e.g. Birthdays, lunches, celebrations falling on a date

e.g. Stress or emphasis falling on a syllable or a topic

e.g. Responsibility, luck, suspicion falling on or to a person

# *Talk Outline*

- Introduction
- Motivation
- Annotation task
  - Data set
  - Annotation guidelines
- Design of sense inventories
- Relations between senses
- Annotation decisions
- Conclusions

# *Annotation Task*

Annotators are given

- a sense inventory for each verb
- a set of sentences to tag

# *Annotation Guidelines*

Annotators were given the following instructions:

- mark each sentence with the most fitting sense
- allowed to mark sentence as “N/A” if
  - sense inventory was missing the relevant sense
  - more than one sense seemed to fit
  - the sense was impossible to determine from context
- with respect to metaphoric senses, instructions were to throw out cases where interpretation was difficult or not immediately clear
- idiomatic expressions and phrasal verbs thrown out

# *Inter-Annotator Agreement*

- Annotators were two linguistics majors
- Inter-annotator agreement was 95%
- Disagreements were resolved in adjudication by the co-authors
- ITA was computed as macro-average of percentage of instances annotated with the same sense by both annotators
- The instances marked as “N/A” by one or both annotators or in adjudication were not included in the computation



# *Talk Outline*

- Introduction
- Motivation
- Annotation task
  - Data set
  - Annotation guidelines
- Design of sense inventories
- Relations between senses
- Annotation decisions
- Conclusions

# *Sense Inventory Problems*

- Parallel sense distinctions
  - Words used in the sense inventory may have sense distinctions parallel to the sense distinctions of the target word being described
    - OntoNotes sense inventory for **fire** has a gloss **ignite or become ignited** under which very divergent examples are grouped:
      - **oil fired the furnace** (literal, primary sense)
      - **curiosity fired my imagination** (metaphoric extension).
- Generic vs. specific senses
  - **acquire a land, a business** (purchase or become the owner of property) vs. **acquire an infection, a boyfriend, a following** (become associated with something, often newly brought into being)
  - **fall** (be associated or get assigned to a person, location or time)

# *Prototypicality in Argument Sets*

- The same sense is often activated by a number of semantically diverse arguments.
- The requisite semantic component may be central to some of them, and accidental to others.
  - absorb oil, oxygen, water vs. dirt, flavour, moisture
    - actual SUBSTANCES vs. other words that activate the same sense
  - take on: “tackle an adversary” vs. “acquire a quality”
    - competition, enemy, opponent, government, world
    - shape, meaning, color, reality
      - competition vs. government
        - [+adversary] component

# *“Senses in Construction”*

- Each decision to split a sense and make another category is to a certain extent an arbitrary decision
  - *drive a nail into the ground* vs. *drive the ball into the corner*
- Distinguishing an alternation as a separate “sense in construction” may be useful for inference
  - knowing which semantic role relative to the described event is expressed by a particular argument
  - *drive a car* vs. *drive twenty miles*

# *Regular Semantic Processes*

- Postulating a separate sense may or may not be justified when there are regular semantic processes that allow complements to satisfy selectional requirements of the verb
  - **conclude visit, tour** vs. **conclude letter, chapter, novel**
    - the latter are coerced into events corresponding to activity that brings them about, i.e. reinterpreted as events of writing
  - **deny allegations, reports** vs. **deny attack, involvement**
    - event nouns coerced into a propositional reading

# *Generative Lexicon (J. Pustejovsky)*

- complex types
  - dealing with regular polysemy of complex nouns such as **book** ( **INFO • PHYSOBJ** ), **building** ( **PROCESS • RESULT** )
- qualia structure, esp. for nouns
  - agentive (how did it come about?)
  - telic (what is it used for?)
  - formal (what is it?)
  - constitutive (what is it made of, what are its parts?)

# *Boundary Cases*

- When verb senses are linked to distinctions in argument semantics, there are almost always boundary cases
  - The *diagram* showed the dependency between X and Y
- Other examples
  - conclude a meeting, investigation, visit (“finish an event”) vs. conclude a treaty, contract, cease-fire (“reach an agreement”)
    - conclude negotiations
  - launch an expedition (“begin an event”) vs. launch a missile (“propel a physical object”)
    - launch a ship

# *Talk Outline*

- Introduction
- Motivation
- Annotation task
  - Data set
  - Annotation guidelines
- Design of sense inventories
- Relations between senses
- Annotation decisions
- Conclusions



## *Relations between Senses: Argument Structure Alternations*

1. Different case roles (frame elements) may be expressed in the same argument position (e.g. dobj), corresponding to different perspectives on the same event.

Example 1: Direct object of *drive* may be **VEHICLE**, **DISTANCE**, or **PHYSOBJ** giving rise to 3 senses:

- a. operate a vehicle controlling its motion
- b. travel in a vehicle a certain distance
- c. transport something or someone

Example 2: for *fire*, **PROJECTILE** or **WEAPON** as dobj give rise to two related senses:

- a. shoot, discharge a weapon
- b. shoot, propel a projectile

## *Relations between Senses: Argument Structure Alternations 2*

- The distinction between propositional and non-propositional complements:
  - a. *admit defeat, inconsistency, offence* (acknowledge the truth or reality of)
  - b. *admit patients, students* (grant entry or allow into a community)
- Mutual dependency between subcategorization features of the complements in different argument positions.

Example: the [+animate] subject of *acquire* may combine with specific complements not available for [-animate]:

  - a. learn: NPsubj [+animate] *acquire* NPdobj (*language, manners, knowledge, skill*)
  - b. take on certain characteristics: NPsubj [-animate] *acquire* NPdobj (*importance, significance*).

## *Relations between Senses: Semantic Underspecification*

- The meaning component ‘**manner of motion**’ gets transformed in different senses of *drive*.
  - Physical senses of *drive* (“**operate a vehicle**”, “**transport something or somebody**”): PRESENT
  - Non-physical use (‘**motivate the progress**’: *drive the economy, drive the market forward*): LOST
  - The value of the agentive role of *drive* becomes semantically weak and the overall meaning of *drive* is transformed to ‘**cause something to move**’.

# *Relations between Senses: Lexical Semantic Features*

- Information about semantic type contained in QS allows apparently diverse elements to activate the same sense of the V.
  - Example: *absorb* (sense ‘learn or incorporate skill or information’) gets as direct objects *values, atmosphere, information, idea, words, lesson, attitudes, culture*.
- Different semantic realizations of the requisite semantic component:
  - complex types with INFORMATION as one of the constituent types: *words* (ACOUSTIC/VISUAL ENTITY•INFO), *lesson* (EVENT•INFO).
  - polysemous direct objects with one of the senses being INFORMATION: *idea*
  - more difficult cases: *culture* and *values* refer to knowledge, and the INFORMATION component is clearly present.
- Consequently, the annotators are able to identify the corresponding sense of *absorb* with a high degree of agreement.

# *Relations between Senses: Metaphor*

Some of the conventionalized extensions with metaphorical flavour:

- a. *grasp object* vis. *grasp meaning*
- b. *launch object* vs. *launch an event* or *launch a product (newspaper, collection)*
- c. *meet with a person* vs. *meet with success, resistance*
- d. *lead somebody to a location* vs. *lead to a consequence*

The distinction between generic and specific senses is one of the effects of the metaphorization

Example: *acquire land, business* (specific sense) vs. *acquire an infection, a boyfriend, a following* (light generic association).

Specificity involving specialization within a certain domain:

- a. *conclude* as 'finish' vs. *conclude* as 'reach an agreement'  
(Law, Politics)
- b. *fire* as 'shoot a weapon or a projectile' vs. *fire* as 'kick or pass an object of play in sports' (Sport)

# *Talk Outline*

- Introduction
- Motivation
- Annotation task
  - Data set
  - Annotation guidelines
- Design of sense inventories
- Relations between senses
- **Annotation decisions**
- **Conclusions**

# *Analysis of Annotation Decisions 1*

Situation 1. A specific meaning is not included into the sense inventory

Annotation decisions:

- use a more general meaning (annoB)
- pick the closest meaning possible (annoA)

Engineers successfully fired thrusters to boost the research satellite to an altitude of 507 km.

annoA: shoot, propel a projectile

annoB: apply fire to

# *Analysis of Annotation Decisions 2*

Situation 2. The appropriate specific sense is available

Annotation decision: annotators choose the more generic sense

Several referrals fell into this category.

annoA: be associated with or get assigned to a person or location or  
for event to fall onto a time

annoB: be categorized as or fall into a range

correct: be categorized as or fall into a range

He acquired a taste for performing in public.

annoA: become associated with something, often newly brought into  
being

annoB: become associated with something...

correct: learn

Not a desirable outcome, since generic senses are introduced in the inventory to account only for semantically underspecified cases.



# *Analysis of Annotation Decisions 3*

Situation 3. Ambiguity of literal and non-literal uses

Annotation decisions:

- metaphoric sense chosen (annoA)
- literal sense chosen (annoB)

She was delighted when the story of Hank fell into her lap

annoA: be associated with or get assigned to a person or location...

annoB: physically drop; move or extend downward

# *Analysis of Annotation Decisions 4*

Situation 4. Impact of subcategorization features on disambiguation: the animacy of the subject activates 2 different subcategorization frames and 2 different senses

The reggae tourist can easily absorb the current reggae vibe.

annoA: absorb energy or impact

annoB: learn or incorporate skill or information

# *Analysis of Annotation Decisions 5*

Situation 5. Semantic type of the relevant argument is not clear

Annotation decisions depend on the interpretation of semantically complex types assumed by each annotator:

**program** [Event•Product], **vehicle** [PhysObj•Product]

The AAA launched education programs.

annoA: *begin or initiate an endeavour* (EVENT)

annoB: *begin to produce or distribute; start a company* (PRODUCT)

France plans to launch a remote-sensing vehicle called Spot.

annoA: *physically propel into the air, water or space* (PhysObj)

annoB: *begin to produce or distribute; start a company* (PRODUCT)

# *Analysis of Annotation Decisions 6*

Situation 6. Influence of wider context: resort to domain-specific clues is necessary to identify the sense

The choice of annoB could be motivated by specific clues referring to military conflict: **rebel control**

The road fell into *rebel control*.

annoA: be associated with or get assigned to a person or location or for event to fall onto a time

annoB: lose power or suffer a defeat

# *Analysis of Annotation Decisions 7*

Situation 7. The senses compatible with a given sentence can be interpreted as having positive or negative connotation.

..help absorb the latest wave of immigrants.

annoA: bear the cost of; take on an expense (NEGATIVE)

annoB: take in or assimilate, making part of a whole or a group (POSITIVE)

For senior management an important lesson was the trade union's capacity to absorb change and to become its agents.

annoA: learn or incorporate skill or information (POSITIVE)

annoB: bear the cost of; take on an expense (NEGATIVE)

# *Analysis of Annotation Decisions 8*

Situation 8. The senses have different presuppositions with respect to pre-existence of the relevant argument

The annotation decision depends on the temporal reference interpretations: for annoB *success* is something that has already happened.

One area where the government can claim some success involves debt repayment.

annoA: come in possession of or claim property you are entitled to

annoB: claim the truth of

# *Discussion*

- Our analysis suggests that theoretical tools must be refined and further developed to give an adequate account to the sense modifications found in real corpus data.
- Appropriate semantic annotation that would allow one to determine which sense distinctions can be detected better by automatic systems does not need to be highly specific and unnecessarily complex, but requires development of robust generalizations about sense relations.
- Data sets need to be explicitly restricted to the instances where humans have no trouble disambiguating between different senses. Prototypical cases can be accounted for reliably, ensuring the clarity of annotated sense distinctions. This decision impacts most strongly those boundary cases which are not reliably disambiguated by human annotators, and which introduce noise into the data set.

*Thank you*



Anna Rumshisky, Olga Batiukova  
Brandeis University