# Automatic prediction of aspectual class of verbs in context (ACL 2014)

**Annemarie Friedrich & Alexis Palmer**

afried,apalmer@coli.uni-saarland.de

Department of Computational Linguistics
Saarland University

University of Edinburgh, May 6th, 2014

# Stative and dynamic meanings of verbs in context

The juice **fills** the glass.
*(stative)*
'something is the case'

# Stative and dynamic meanings of verbs in context

The juice **fills** the glass.
*(stative)*
'something is the case'


She **filled** the glass with juice.
*(dynamic)*
'something happens'

# Ambiguous verb types:
# sometimes both readings are available


The glass was **filled** with juice.
*(both readings)*

# Why predict aspectual class?

- aspectual class of a discourse's finite verbs
  = important factor in conveying / interpreting temporal structure (e.g. Moens & Steedman 1998)
  (other factors: tense, grammatical aspect, mood, completedness)

# Why predict aspectual class?

- aspectual class of a discourse's finite verbs
  = important factor in conveying / interpreting temporal
  structure (e.g. Moens & Steedman 1998)
  (other factors: tense, grammatical aspect, mood, completedness)

$$\Downarrow$$

improve temporal information processing:
improve many NLP tasks / computational semantics approaches (e.g. Lewis & Steedman 2014)

# Why predict aspectual class?

- aspectual class of a discourse's finite verbs
  = important factor in conveying / interpreting temporal structure (e.g. Moens & Steedman 1998)
  (other factors: tense, grammatical aspect, mood, completedness)

$$\Downarrow$$

improve temporal information processing:
  improve many NLP tasks / computational semantics approaches (e.g. Lewis & Steedman 2014)

- one factor in determing situation entity type (Smith 2003)
  – more on this later

# Aspectual class: linguistic literature

- Vendler (1957): time schemata of English verbs (*lexical aspect*)

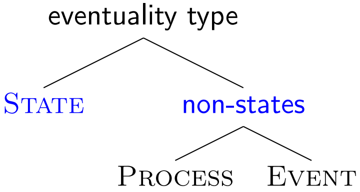
| | | |
|---|---|---|
| **states** | *love, own* | STATIVE |
| **activities** | *run, write letters* | DYNAMIC |
| **accomplishments** | *run a mile, write a letter* | |
| **achievements** | *realize, cross the border* | |

# Aspectual class: linguistic literature

- Vendler (1957): time schemata of English verbs (*lexical aspect*)

<table>
<tr><td>states</td><td>love, own</td><td>STATIVE</td></tr>
<tr><td>activities</td><td>run, write letters</td><td rowspan="3">DYNAMIC</td></tr>
<tr><td>accomplishments</td><td>run a mile, write a letter</td></tr>
<tr><td>achievements</td><td>realize, cross the border</td></tr>
</table>

- Bach (1986): sentences




- ...

# Fundamental aspectual class

## Fundamental aspectual class

= a clause's aspectual class when ignoring any aspectual markers / transformations (following Siegel & McKeown 2000)

= a function of the main verb + a select group of arguments (which may differ per verb)

# Fundamental aspectual class

## Fundamental aspectual class

= a clause's aspectual class when ignoring any aspectual markers / transformations (following Siegel & McKeown 2000)

= a function of the main verb + a select group of arguments (which may differ per verb)

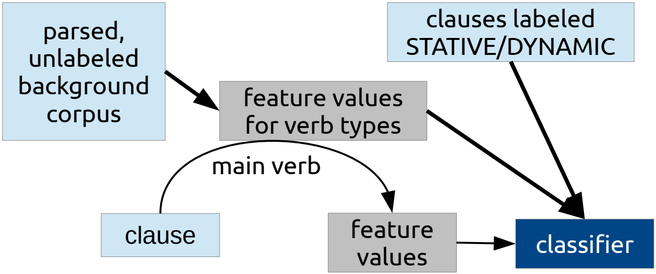
## Example

*John has **kissed** Mary.*

- English perfect → stative (Smith 1991, Katz 2003)
- fundamental aspectual class: *John kiss Mary* → DYNAMIC

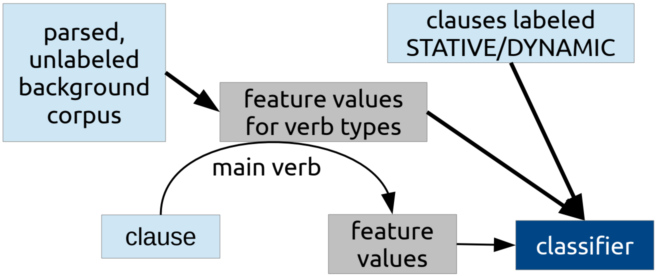
# Aspectual class: computational linguistics

## Type-based approaches



# Aspectual class: computational linguistics

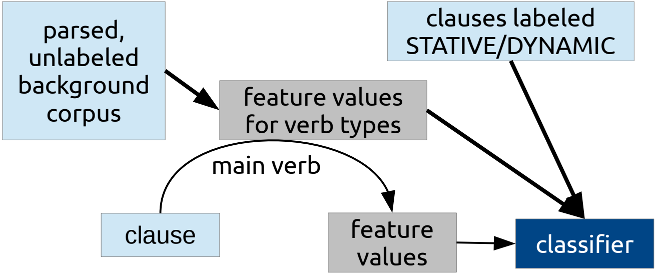
## Type-based approaches




- Klavans & Chodorow (1992): **degrees** of stativity
  $\approx$ tendencies of use are collected for verb types from corpus
  (progressive as indicator)

# Aspectual class: computational linguistics

## Type-based approaches

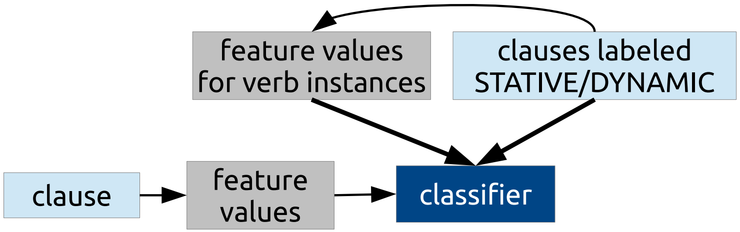



- Klavans & Chodorow (1992): **degrees** of stativity
  $\approx$ tendencies of use are collected for verb types from corpus
  (progressive as indicator)
- Siegel & McKeown (2000): **linguistic indicators**,
  did not beat most-frequent-class baseline $\rightarrow$ main inspiration

# Aspectual class: computational linguistics

**Instance-based approaches**

- Siegel (1998): using direct object of *have*
- TempEval challenges: **event extraction & classification**
  wide range of corpus-based and syntactic-semantic features (e.g. Jung & Stent (2013), Bethard (2013), Chambers (2013))
  Costa & Branco (2013): aspectual features, Portuguese



# Contribution

**This work:**

Automatic prediction of aspectual class of verbs in context

1. creation & analysis of corpora for this task

# Contribution

**This work:**

Automatic prediction of aspectual class of verbs in context

1. creation & analysis of corpora for this task
2. are type-based features sufficient? (no)

# Contribution

**This work:**

Automatic prediction of aspectual class of verbs in context

1. creation & analysis of corpora for this task
2. are type-based features sufficient? (no)
3. do type-based features generalize over verb types? (yes)

# Contribution

**This work:**

Automatic prediction of aspectual class of verbs in context

1. creation & analysis of corpora for this task
2. are type-based features sufficient? (no)
3. do type-based features generalize over verb types? (yes)
4. when are instance-based features helpful?
   (for ambiguous verbs)

# Asp-MASC

- 7875 clauses from the Manually Annotated SubCorpus (MASC) of OANC (Ide et al. 2010): *jokes, letters, news*
  annotated for aspectual class (main verb)

## Example

| | |
|---|---|
| He **assured** her | DYNAMIC |
| that it **was** nonsense, | STATIVE |
| but that Gailhaguet "**knows** my name very well – | STATIVE |

(NYTnewswire2)

# Asp-MASC

- 7875 clauses from the Manually Annotated SubCorpus (MASC) of OANC (Ide et al. 2010): *jokes, letters, news*
  annotated for aspectual class (main verb)

## Example

| | |
|---|---|
| He **assured** her | DYNAMIC |
| that it **was** nonsense, | STATIVE |
| but that Gailhaguet "**knows** my name very well – | STATIVE |

(NYTnewswire2)

- segmentation into clauses: SPADE discourse parser (Soricut & Marcu 2006) + heuristic post-processing
- main verb automatically extracted from dependency parses (Stanford parser), exclude *have/be/none*: 6161 clauses
- 2 annotators: agreed cases → DYNAMIC, STATIVE, BOTH
  disagreed cases: → BOTH

# Asp-MASC: inter-annotator agreement

*Cohen's observed unweighted $\kappa$*

| | complete | | w/o *have/be/none* | |
|---|---|---|---|---|
| **genre** | **clauses** | $\kappa$ | **clauses** | $\kappa$ |
| jokes | 3462 | 0.85 | 2660 | 0.77 |
| news | 2565 | 0.79 | 2075 | 0.69 |
| letters | 1848 | 0.71 | 1444 | 0.62 |
| all | 7875 | 0.80 | 6161 | 0.70 |

# Asp-MASC: inter-annotator agreement

*Cohen's observed unweighted* $\kappa$

| | complete | | w/o *have/be/none* | |
|---|---|---|---|---|
| **genre** | **clauses** | $\kappa$ | **clauses** | $\kappa$ |
| jokes | 3462 | 0.85 | 2660 | 0.77 |
| news | 2565 | 0.79 | 2075 | 0.69 |
| letters | 1848 | 0.71 | 1444 | 0.62 |
| all | 7875 | 0.80 | 6161 | 0.70 |

## Agreement: findings

- substantial agreement
- 'difficulty': jokes $<$ news $<$ letters
  - jokes $\approx$ narratives, letters $\approx$ persuasive/argumentative
    $\Rightarrow$ different rhetorical style, more statives in letters

# Asp-MASC: annotator preferences

w/o *have/be/none*, confusion matrix:

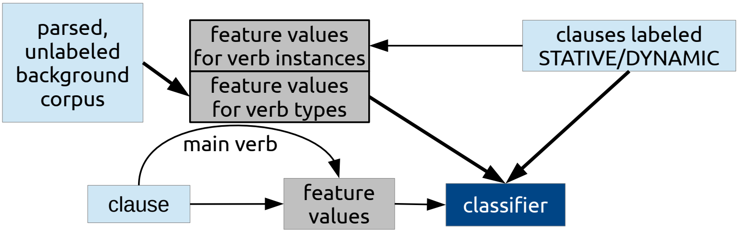
| | | Annotator 2 | | |
|---|---|---|---|---|
| | | DYNAMIC | STATIVE | BOTH |
| Annotator 1 | DYNAMIC | **4464** | 164 | 9 |
| | STATIVE | 434 | **1056** | 29 |
| | BOTH | 5 | 0 | **0** |

## Annotator preferences: findings

- on the disagreed cases, annotator 1 prefers STATIVE, annotator 2 prefers DYNAMIC.
- not uncommon (Beigman Klebanov et al. 2008)

# Method

- three-way classification task: STATIVE/DYNAMIC/BOTH
- (semi-)supervised learning setting:
  Random Forest classifier (Breiman 2001)




- three sets of features:
  1. **linguistic indicator features** (type-based)
  2. **distributional features** (type-based)
  3. **instance-based features**

# Linguistic indicators (LingInd) feature set (type-based)

- idea: **linguistic indicators** (past/present tense, temporal adverb, in-PP, for-PP...) correlate with stative / dynamic readings (Siegel & McKeown 2000)
- count how often instances of a verb type occur with an indicator in a parsed **background corpus**
  $\rightarrow$ we use: Gigaword AFE+XIE, Stanford parser

## Example

verb type: `fill`

feature: `temporal-adverb` feature value: 0.0085

$\Rightarrow$ 0.85% of the occurrences of *fill* in the corpus are modified by one of the temporal adverbs on the list compiled by Siegel (1998).

# Linguistic indicators (LingInd) feature set (type-based)

| Feature | Example |
|---|---|
| frequency | - |
| present | *says* |
| past | *said* |
| future | *will say* |
| perfect | *had won* |
| progressive | *is winning* |
| negated | *not/never* |
| particle | *up/in/...* |
| no subject | - |

Loaiciga et al. (2014)

| Feature | Example (lists) |
|---|---|
| continuous adverb | *continually* *endlessly* |
| evaluation adverb | *better* *horribly* |
| manner adverb | *furiously* *patiently* |
| temporal adverb | *again* *finally* |
| in-PP | *in an hour* |
| for-PP | *for an hour* |

Siegel & McKeown (2000)

# Verb type seed sets

- Lexical Conceptual Structures (LCS) database of English verbs (Dorr 2001)
- Dorr & Olsen (1997): rules to extract 'dynamicity' feature for verb types – we extract three lists of verbs:

## LCS verb type seed sets

1. 188 verbs whose entries in LCS are all stative
   *belong, cost, possess,...*
2. 3760 verbs whose entries in LCS are all dynamic
   *alter, knock, resign,...*
3. 215 verbs that have both stative and dynamic entries in LCS
   *fill, stand, take,...*

# Distributional features (type-based)

- **idea**: use LCS seed lists, but make up for noise / verbs not on list by averaging over distributional similarities

# Distributional features (type-based)

- **idea**: use LCS seed lists, but make up for noise / verbs not on list by averaging over distributional similarities
- **Distributional model** (Thater et al. 2011) trained on part of Gigaword, produces syntactically informed vectors representing contexts in which the verb type occurs.

# Distributional features (type-based)

- **idea**: use LCS seed lists, but make up for noise / verbs not on list by averaging over distributional similarities
- **Distributional model** (Thater et al. 2011) trained on part of Gigaword, produces syntactically informed vectors representing contexts in which the verb type occurs.

## 3 numeric features per verb type:

1. average cosine similarity with verbs on ‘stative’ LCS seed list
2. average cosine similarity with verbs on ‘dynamic’ LCS seed list
3. average cosine similarity with verbs on ‘mixed’ LCS seed list

# Instance-based features

- extracted from the **clause** to be classified
  (tense/progressive/perfect/voice: see Loaiciga et al. (2014))

| Feature | Values |
|---|---|
| part-of-speech tag of the verb | VB, VBG, VBN, ... |
| tense | present, past, future |
| progressive | true/false |
| perfect | true/false |
| voice | active/passive |
| grammatical dependents | WordNet lexname/POS |

## Example

*A little girl had just* ***finished*** *her first week of school.*

`tense`:*past* `perfect`:*true* `voice`:*active*

`dobj`:*noun.time* `subj`:*noun.person* `particle`:*none*

# Experiments: overview

| feature sets |
|---|
| type-based:<br>- linguistic indicators<br>- distributional features |
| instance-based |

| data |
|---|
| labeled clauses |

# Experiments: overview

| feature sets |
|---|
| type-based:<br>- linguistic indicators<br>- distributional features |
| instance-based |

| data |
|---|
| labeled clauses |

| classification settings |
|---|
| verb types SEEN in training data |
| verb types UNSEEN in training data |
| ONE-LABEL VS. MULTI-LABEL verbs |
| INSTANCE-BASED classification:<br>ambiguous verb types |

# Experiments: overview

| **feature sets** |
|---|
| type-based:<br>- linguistic indicators<br>- distributional features |
| instance-based |

| **data** |
|---|
| labeled clauses |

| **classification settings** |
|---|
| verb types SEEN in training data |
| verb types UNSEEN in training data |
| ONE-LABEL VS. MULTI-LABEL verbs |
| INSTANCE-BASED classification:<br>ambiguous verb types |

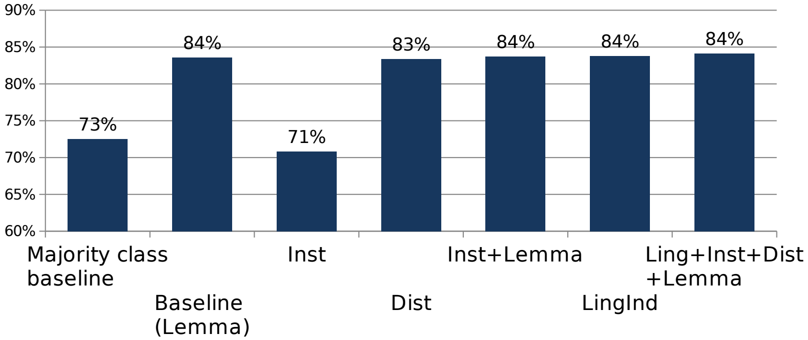
- How effective are the feature sets in each setting?
- Significance testing: McNemar's test with Yates' correction of continuity, $p < 0.01$

# SEEN verbs: labeled training data for verb type available

- Asp-MASC, Random Forest classifier, 10-fold cross validation, distributing instances of verb types over folds
- Lemma = lemma of main verb used as additional feature
- Baseline (Lemma) memorizes most frequent class of verb type in training folds

# SEEN verbs: labeled training data for verb type available

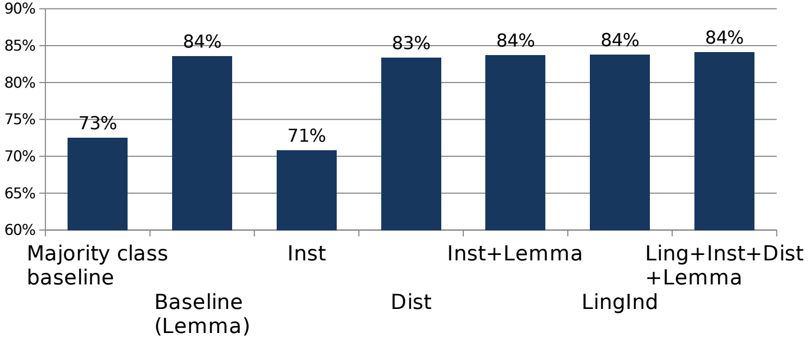
- Asp-MASC, Random Forest classifier, 10-fold cross validation, distributing instances of verb types over folds
- Lemma = lemma of main verb used as additional feature
- Baseline (Lemma) memorizes most frequent class of verb type in training folds



# Seen verbs: labeled training data for verb type available

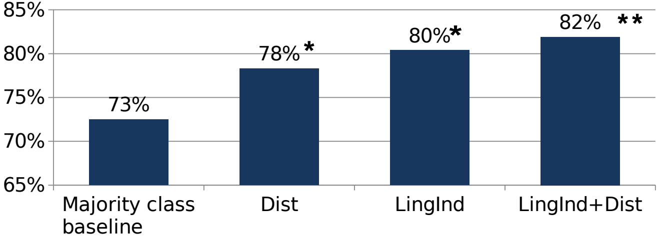
## Seen verb types: findings

All type-based features result in the same performance as using the most frequent class of the type in the training data (no significant improvements).



# UNSEEN verbs: no labeled training data available

- Logistic regression, Asp-MASC, 10-fold cross validation: all occurrences of a verb type in the same fold
- Baseline: most frequent class (DYNAMIC)

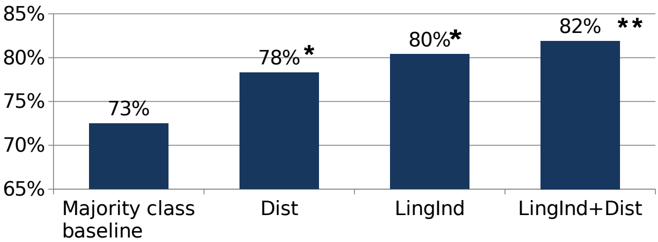



* significantly better than baseline

** significantly better than LingInd

# UNSEEN verbs: no labeled training data available

- Logistic regression, Asp-MASC, 10-fold cross validation: all occurrences of a verb type in the same fold
- Baseline: most frequent class (DYNAMIC)




* significantly better than baseline

** significantly better than LingInd

## UNSEEN verb types: findings

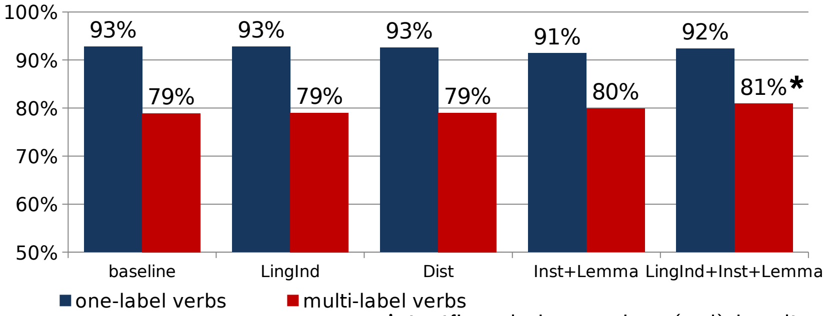
LingInd + Dist features generalize across verb types. Combination works best.

## One-label vs. multi-label verbs

- one-label verbs: all instances in Asp-MASC have the same label (1966 instances, 806 verb types)
- multi-label verbs: instances have differing labels (4195 instances, 264 verb types)
- ‘seen’ setting: Random Forest, 10-fold cross validation

# One-label vs. multi-label verbs

- one-label verbs: all instances in Asp-MASC have the same label (1966 instances, 806 verb types)
- multi-label verbs: instances have differing labels (4195 instances, 264 verb types)
- ‘seen’ setting: Random Forest, 10-fold cross validation

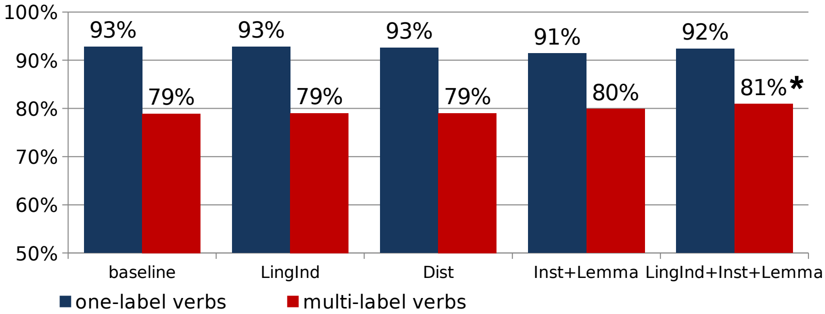



*significantly better than (red) baseline

# One-label vs. multi-label verbs: **Accuracy**

## Multi-label verb types: findings

Type-based features: always select predominant class; but useful bias.
Instance-based features: essential for classifying ambiguous verbs.
⇒ motivation for further investigating ambiguous verbs




*significantly better than (red) baseline

# Instance-based classification: Asp-Ambig data set

- 20 frequent verbs that can occur as either stative or dynamic (selected from LCS list of 'mixed' verb types)
- for each: 138 sentences randomly extracted from Brown corpus
- two annotators mark the aspectual class of the verb in question (highlighted)

# Instance-based classification: Asp-Ambig data set

- 20 frequent verbs that can occur as either stative or dynamic (selected from LCS list of 'mixed' verb types)
- for each: 138 sentences randomly extracted from Brown corpus
- two annotators mark the aspectual class of the verb in question (highlighted)
- 2667 instances, $\kappa = 0.6$ (Asp-MASC $\kappa = 0.7$)
  1444 DYNAMIC, 697 STATIVE, 526 BOTH

# Instance-based classification

- Asp-Ambig (20 verbs, $\approx$ 138 instances each)
- Random Forest, Leave-One-Out cross validation

| Features | Micro-avg. accuracy |
|---|---|
| majority class baseline | 66.3% |
| type-based features | |
| Inst | 58.1% |
| Inst+Lemma | 71.0% |
| Inst+Lemma+LingInd+Dist | 72.0% |

# Instance-based classification

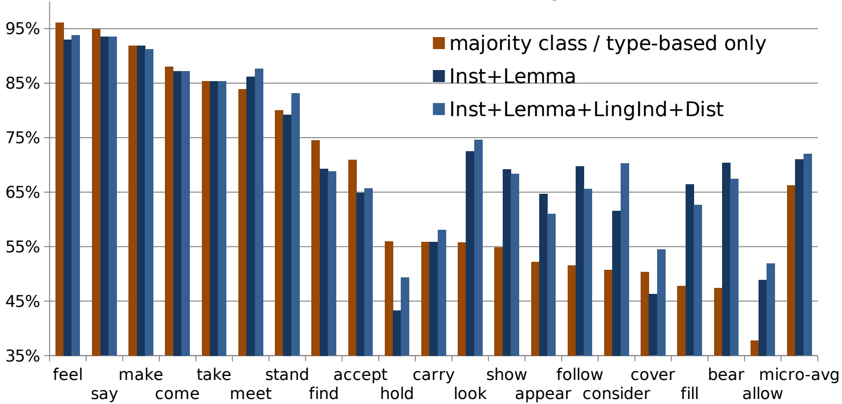
- Asp-Ambig (20 verbs, $\approx$ 138 instances each)
- Random Forest, Leave-One-Out cross validation

| Features | Micro-avg. accuracy |
|---|---|
| majority class baseline | 66.3% |
| type-based features | |
| Inst | 58.1% |
| Inst+Lemma | 71.0% |
| Inst+Lemma+LingInd+Dist | 72.0% |

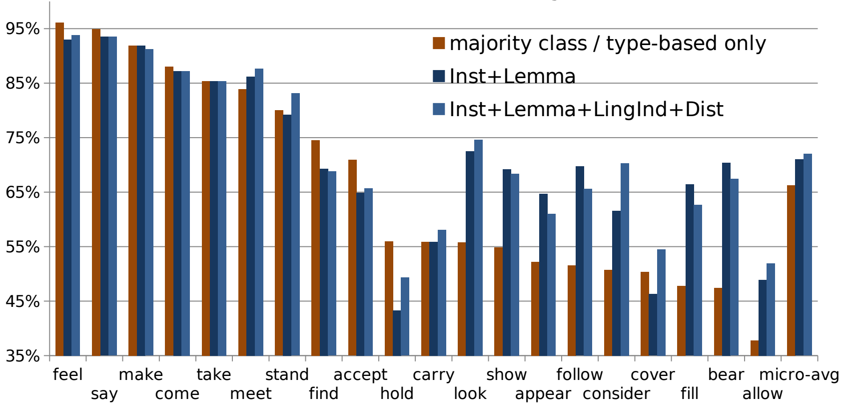
## INSTANCE-BASED classification: findings

Inst features do not generalize across verb types.
Only useful as a feature in combination with the verb type.

# Instance-based classification: **Accuracy**



# Instance-based classification: **Accuracy**




## INSTANCE-BASED classification: findings

The more 'ambiguous' the verb type, the more essential are instance-based features.

Type-based features (bias) helpful? $\rightarrow$ depends on the verb type

# Summary

- **context-aware** approach to automatically predicting aspectual class, new set of distributional features
- two new **corpora**: Asp-MASC & Asp-Ambig

# Summary

- **context-aware** approach to automatically predicting aspectual class, new set of distributional features
- two new **corpora**: Asp-MASC & Asp-Ambig
- labeled training data available $\rightarrow$ improvement over most-frequent-class baseline can only be reached by integrating **instance-based** features

# Summary

- **context-aware** approach to automatically predicting aspectual class, new set of distributional features
- two new **corpora**: Asp-MASC & Asp-Ambig
- labeled training data available $\rightarrow$ improvement over most-frequent-class baseline can only be reached by integrating **instance-based** features
- **type-based features** can provide useful prior & are useful to predict predominant aspectual class for 'unseen' verb types

# Summary

- **context-aware** approach to automatically predicting aspectual class, new set of distributional features
- two new **corpora**: Asp-MASC & Asp-Ambig
- labeled training data available → improvement over most-frequent-class baseline can only be reached by integrating **instance-based** features
- **type-based features** can provide useful prior & are useful to predict predominant aspectual class for ‘unseen’ verb types

**Future work**

- a globally well-performing system: multi-stage approach, treating verbs differently according to whether the verb’s aspectual class distribution is highly skewed
- gather more data & apply more features

# Thanks to:

Ambika Kirkland & Ruth Kühn

Omri Abend, Mike Lewis, Annie Louis, Manfred Pinkal,
Mark Steedman, Stefan Thater, Bonnie Webber

# References I

Emmon Bach. 1986. The algebra of events. Linguistics and philosophy, 9(1):5–16.

Beata Beigman Klebanov, Eyal Beigman, and Daniel Diermeier. 2008. Analyzing disagreements. In Procceedings of the Workshop on Human Judgements in Computational Linguistics, pages 2–7. Association for Computational Linguistics.

Steven Bethard. 2013. ClearTK-TimeML: A minimalist approach to TempEval 2013. In Second Joint Conference on Lexical and Computational Semantics (* SEM), volume 2, pages 10–14.

Leo Breiman. 2001. Random forests. Machine Learning, 45(1):5–32.

Nathanael Chambers. 2013. Navytime: Event and time ordering from raw text. In Second Joint Con- ference on Lexical and Computational Semantics (*SEM), volume 2, pages 73–77.

Francisco Costa and Antonio Branco. 2012. Aspectual type and temporal

# References II

relation classification. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pages 266–275. Association for Computational Linguistics.

Bonnie J. Dorr and Mari Broman Olsen. 1997. Deriving verbal and compositional lexical aspect for NLP applications. In Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, pages 151–158. Association for Computational Linguistics.

Bonnie J. Dorr. 2001. LCS verb database. Online software database of Lexical Conceptual Structures, University of Maryland, College Park, MD.

Nancy Ide, Christiane Fellbaum, Collin Baker, and Rebecca Passonneau. 2010. The manually annotated sub-corpus: a community resource for and

# References III

by the people. In Proceedings of the ACL 2010 Conference Short Papers, pages 68–73. Association for Computational Linguistics.

Graham Katz. 2003. On the stativity of the English perfect. Perfect explorations, pages 205–234.

Judith L. Klavans and Martin Chodorow. 1992. Degrees of stativity: the lexical representation of verb aspect. In Proceedings of the 14th conference on Computational Linguistics, pages 1126–1131. Association for Computational Linguistics.

Sharid Loaiciga, Thomas Meyer, and Andrei Popescu-Belis. 2014. English-French Verb Phrase Alignment in Europarl for Tense Translation

# References IV

Modeling. In Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland.

Marc Moens and Mark J. Steedman. 1988. Temporal ontology and temporal reference. Computational Linguistics, 14(2):15–28.

Eric V. Siegel. 1998a. Disambiguating verbs with the WordNet category of the direct object. In Proceedings of Workshop on Usage of WordNet in Natural Language Processing Systems, Universite de Montreal.

Eric V. Siegel and Kathleen R. McKeown. 2000. Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights. Computational Linguistics, 26(4):595–628.

Radu Soricut and Daniel Marcu. 2003. Sentence level discourse parsing using syntactic and lexical information. In Proceedings of the 2003 Conference of the North American Chapter of the Association for

# References V

Computational Linguistics on Human Language Technology-Volume 1, pages 149–156. Association for Computational Linguistics.

Stefan Thater, Hagen Fürstenau, and Manfred Pinkal. 2011. Word meaning in context: A simple and effective vector model. In IJCNLP, pages 1134–1143.

Zeno Vendler, 1957. Linguistics in Philosophy, chapter Verbs and Times, pages 97–121. Cornell University Press, Ithaca, New York.

Backup Slides

# SEEN verbs: labeled training data for verb type available

- Asp-MASC, Random Forest classifier, 10-fold cross validation, distributing instances of verb types over folds
- Lemma = lemma of main verb used as additional feature

| Features | Accuracy (%) |
|---|---|
| Majority class baseline (DYNAMIC) | 72.5 |
| Baseline (Lemma) * | 83.6 |
| LingInd | 83.8 |
| Inst | 70.8 |
| Inst+Lemma | 83.7 |
| Dist | 83.4 |
| LingInd+Inst+Dist+Lemma | 84.1 |

*memorizes most frequent class of verb type in training folds

## SEEN verb types: findings

All type-based features result in the same performance as using the most frequent class of the type in the training data (no significant

# Instance-based classification

- Asp-Ambig, Random Forest, Leave-One-Out cross validation
- using Inst features alone: acc. 58.1%

| Verb | # of inst. | Majority Class | | Inst +Lemma | Inst +Lemma +LingInd +Dist |
|---|---|---|---|---|---|
| *feel* | 128 | **96.1** | stat | 93.0 | 93.8 |
| *say* | 138 | **94.9** | dyn | 93.5 | 93.5 |
| *make* | 136 | **91.9** | dyn | 91.9 | 91.2 |
| *come* | 133 | **88.0** | dyn | 87.2 | 87.2 |
| *take* | 137 | **85.4** | dyn | 85.4 | 85.4 |
| *meet* | 130 | 83.9 | dyn | 86.2 | **87.7** |
| *stand* | 130 | 80.0 | stat | 79.2 | **83.1** |
| *find* | 137 | **74.5** | dyn | 69.3 | 68.8 |
| *accept* | 134 | **70.9** | dyn | 64.9 | 65.7 |
| *hold* | 134 | **56.0** | both | 43.3 | 49.3 |
| *carry* | 136 | 55.9 | dyn | 55.9 | **58.1** |
| *look* | 138 | 55.8 | dyn | 72.5 | **74.6** |
| *show* | 133 | 54.9 | dyn | **69.2** | 68.4 |
| *appear* | 136 | 52.2 | stat | **64.7** | 61.0 |
| *follow* | 122 | 51.6 | both | **69.7** | 65.6 |
| *consider* | 138 | 50.7 | dyn | 61.6 | **70.3** |
| *cover* | 123 | 50.4 | stat | 46.3 | **54.5** |
| *fill* | 134 | 47.8 | dyn | **66.4** | 62.7 |
| *bear* | 135 | 47.4 | dyn | **70.4** | 67.4 |
| *allow* | 135 | 37.8 | dyn | 48.9 | **51.9** |
| micro-avg. | 2667 | 66.3 | | **71.0*** | **72.0*** |

# Asp-Ambig: confusion matrix

| | | Annotator 2 | | |
|---|---|---|---|---|
| | | DYNAMIC | STATIVE | BOTH |
| Annotator 1 | DYNAMIC | **1444** | 201 | 54 |
| | STATIVE | 168 | **697** | 20 |
| | BOTH | 44 | 31 | **8** |

# UNSEEN verbs: no labeled training data available

- Logistic regression, Asp-MASC, 10-fold cross validation: all occurrences of a verb type in the same fold
- Baseline: most frequent class (DYNAMIC)

| | Features | Accuracy (%) |
|---|---|---|
| 1 | Baseline | 72.5 |
| 2 | Dist | 78.3* |
| 3 | LingInd | 80.4* |
| 4 | LingInd+Dist | **81.9*†** |

## UNSEEN verb types: findings

LingInd + Dist features generalize across verb types.
Combination works best.

# ONE-LABEL vs. MULTI-LABEL verbs

| Data | Features | Acc. (%) |
|---|---|---|
| one-label verbs (1966 inst.) | Baseline | 92.8 |
| | LingInd | 92.8 |
| | Dist | 92.6 |
| | Inst+Lemma | 91.4* |
| | LingInd+Inst+Lemma | 92.4 |
| multi-label verbs (4195 inst.) | Baseline | 78.9 |
| | LingInd | 79.0 |
| | Dist | 79.0 |
| | Inst | 67.4* |
| | Inst+Lemma | 79.9 |
| | LingInd+Inst+Lemma | **80.9*** |
| | LingInd+Inst+Lemma+Dist | 80.2* |

## MULTI-LABEL verb types: findings

Type-based features always select predominant class.

# ONE-LABEL vs. MULTI-LABEL verbs

| Class | Acc.(%) | P | R | F |
|---|---|---|---|---|
| Baseline (Lemma) | | | | |
| micro-avg. | 78.9 | 0.75 | 0.79 | 0.76 |
| LingInd+Inst+Lemma | | | | |
| DYNAMIC | | 0.84 | 0.95 | 0.89 |
| STATIVE | | 0.76 | 0.69 | 0.72 |
| BOTH | | 0.51 | 0.24 | 0.33 |
| micro-avg. | 80.9* | 0.78 | 0.81 | 0.79 |

Table: **Experiment 3**: 'multi-label', precision, recall and F-measure, detailed class statistics for the best-performing system from Table

## MULTI-LABEL verb types: findings

Significant gains of 2% in accuracy and 3% in F-measure (absolute). 'Difficulty': DYNAMIC< STATIVE< BOTH