Automatic Abstract Anaphora Resolution in German

Adam Roussel

September 13, 2017

Ruhr-Universität Bochum

Introduction

Some examples

What do I mean by ‘abstract anaphora’[1]?

(1) Hammond wanted to bring some dinosaurs back to life.
Dr. Grant thought **this** was a bad idea.

[1] Aka. discourse deixis (Webber, 1988), complex anaphors (Consten, Knees, & Schwarz-Friesel, 2007) ...

Some examples

What do I mean by ‘abstract anaphora’[1]?

(1) Hammond wanted to bring some dinosaurs back to life.
Dr. Grant thought **this** was a bad idea.

(2) But Hammond was determined to do **it**.

[1] Aka. discourse deixis (Webber, 1988), complex anaphors (Consten et al., 2007) ...

Some examples

What do I mean by ‘abstract anaphora’[1]?

(1) Hammond wanted to bring some dinosaurs back to life.
Dr. Grant thought **this** was a bad idea.

(2) But Hammond was determined to do **it**.

(3) Dr. Grant was not aware of the **fact**
that he had already done it.

[1] Aka. discourse deixis (Webber, 1988), complex anaphors (Consten et al., 2007) ...

Some examples

What do I mean by ‘abstract anaphora’[1]?

(1) Hammond wanted to bring some dinosaurs back to life.
Dr. Grant thought **this** was a bad idea.

(2) But Hammond was determined to do **it**.

(3) Dr. Grant was not aware of the **fact**
that he had already done it.

(4) Dr. Sattler was not aware of **this** / **this fact** either.

[1] Aka. discourse deixis (Webber, 1988), complex anaphors (Consten et al., 2007) ...

Some examples

What do I mean by ‘abstract anaphora’[1]?

(1) Hammond wanted to bring some dinosaurs back to life.
Dr. Grant thought **this** was a bad idea.

(2) But Hammond was determined to do **it**.

(3) Dr. Grant was not aware of the **fact**
that he had already done it.

(4) Dr. Sattler was not aware of **this** / **this fact** either.

[1] Aka. discourse deixis (Webber, 1988), complex anaphors (Consten et al., 2007) ...

Some examples

What do I mean by ‘abstract anaphora’[1]?

(1) Hammond wanted to bring some dinosaurs back to life.
Dr. Grant thought **this** was a bad idea.

(2) But Hammond was determined to do **it**.

(3) Dr. Grant was not aware of the **fact**
that he had already done it. → cataphoric

(4) Dr. Sattler was not aware of **this** / **this fact** either. → anaphoric

[1] Aka. discourse deixis (Webber, 1988), complex anaphors (Consten et al., 2007) ...

Abstract Anaphora

This system covers two broad classes of German abstract anaphors:

- Pronouns: *dies* ‘this’, *das* ‘that’, and *es* ‘it’
- Shell nouns: *Tatsache* ‘fact’, *Frage* ‘question’, *Problem* ‘problem’, etc.

Both as anaphora and cataphora

Antecedents (or catacedents) may have either verbal or nominal heads

System Design

Classification and Resolution

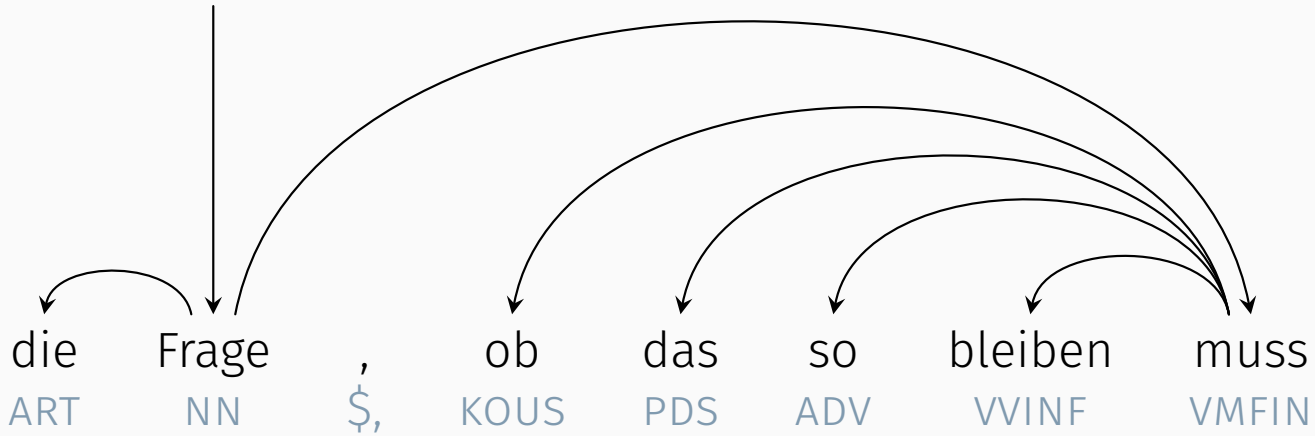
Usual approach involves two steps:

1. Classification: Deciding whether or not an instance requires resolution
2. Resolution: Deciding which potential antecedent belongs with a given anaphor instance

Motivation

- Need to reduce number and variety of candidates
- Effectiveness of sieve-based approaches in coreference resolution (Lee et al., 2013)
- Tendencies of shell nouns to prefer certain patterns (Schmid, 2000)
- Indications that annotator's behavior can be approximated by relatively simple heuristics (Artstein & Poesio, 2006)

Procedure



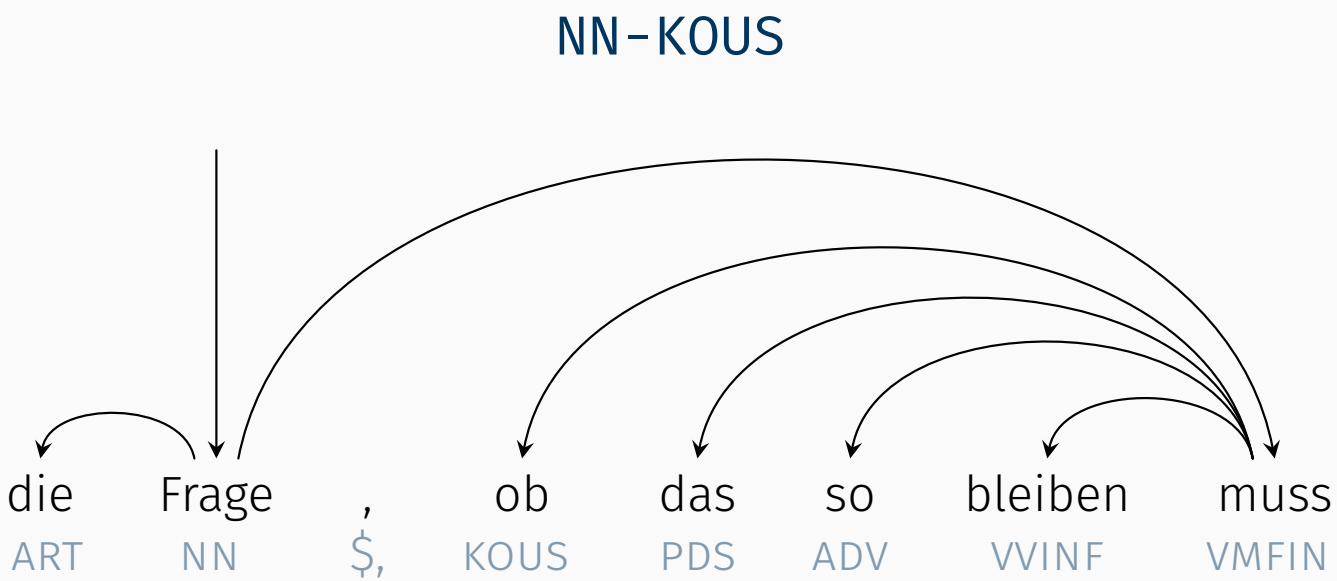
(‘the question whether it has to stay that way’)

Extraction Patterns

Name	Schema	Example
`NN-ist-dass`	sein $\rightarrow$ NN_1 sein $\rightarrow$ x_2 $x_2 \rightarrow$ dass	***Tatsache***$_1$ *ist, dass das nicht* *funktioniert*$_2$
`NN-KOUS`	$NN_1 \rightarrow V_2$ $V_2 \rightarrow$ KOU	*die* ***Frage***$_1$*, ob das so bleiben* *muss*$_2$
`PDS-last-verb`	V_2 ... {das|dies}$_1$	*...nicht* *gibt*$_2$*.* ***Das***$_1$ *wollen sie nun ändern.*
`PDAT-last-sent`	$ROOT_2$ $. NN_1$ $NN_1 \rightarrow$ PDAT	*Er* *hat*$_2$ *es schon getan. Diese* ***Tatsache***$_1$ *war ihr nicht bewusst.*

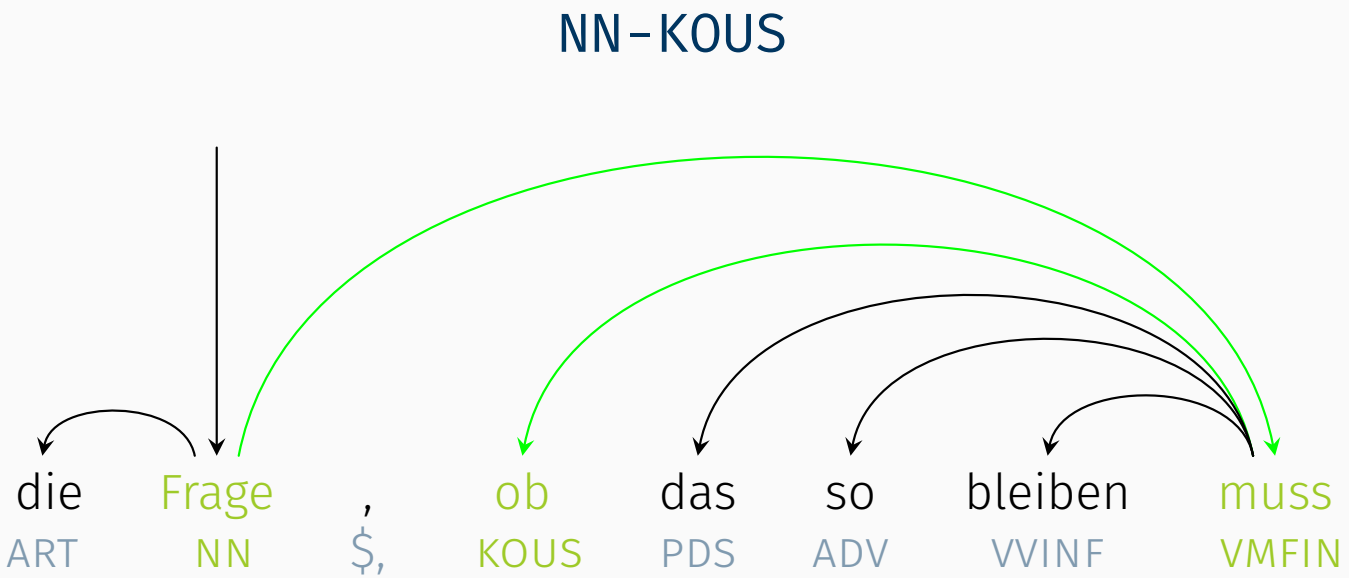
- This implementation includes 14 such extraction patterns
- Patterns ordered according to accuracy/specificity

Procedure

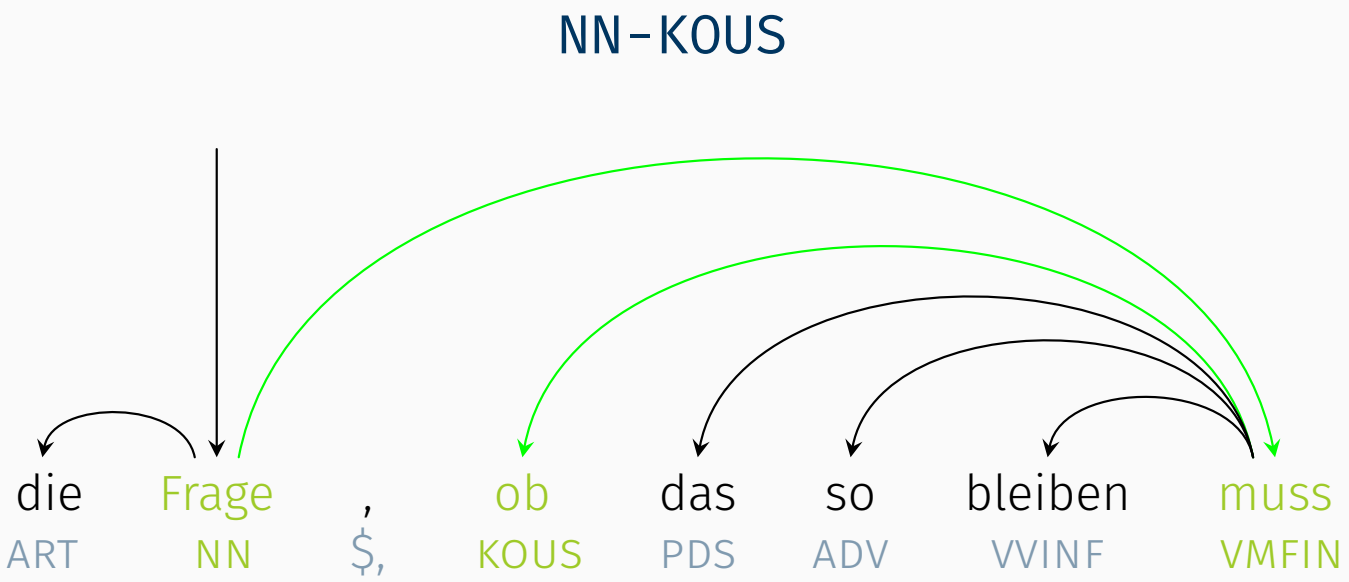


(‘the question whether it has to stay that way’)

Procedure


NN-KOUS
die
ART
Frage
NN
,
$,
ob
KOUS
das
PDS
so
ADV
bleiben
VVINF
muss
VMFIN

Procedure

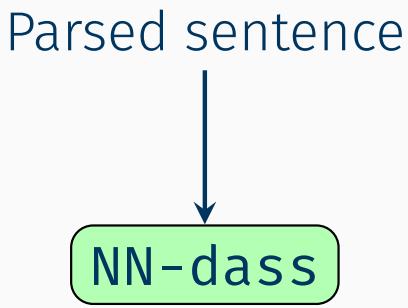


⇒ extracted anaphor–antecedent pair: `(Frage, muss)`

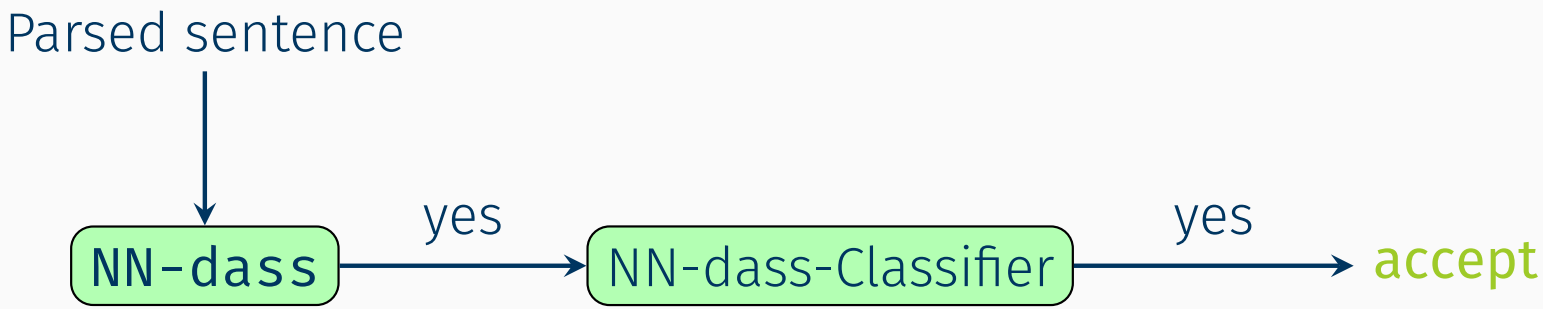
Classification Features

- Lemmas
- Germanet features, e.g.:
  - Semantic field of anaphor and mother of anaphor
  - Whether verbal mother of anaphor could also take a clausal complement
- Syntactic features, e.g:
  - Distance between anaphor & antecedent
  - Grammatical relation of anaphor & antecedent
  - Whether anaphor and antecedent have the same grammatical relation
  - Type of determiner of anaphor, if present
- Surface features, e.g.:
  - Whether head ends in *-ung*, *-keit*, or *-heit* (nominal antecedents)
  - Whether head ends in *-en* (substantivized verbs)

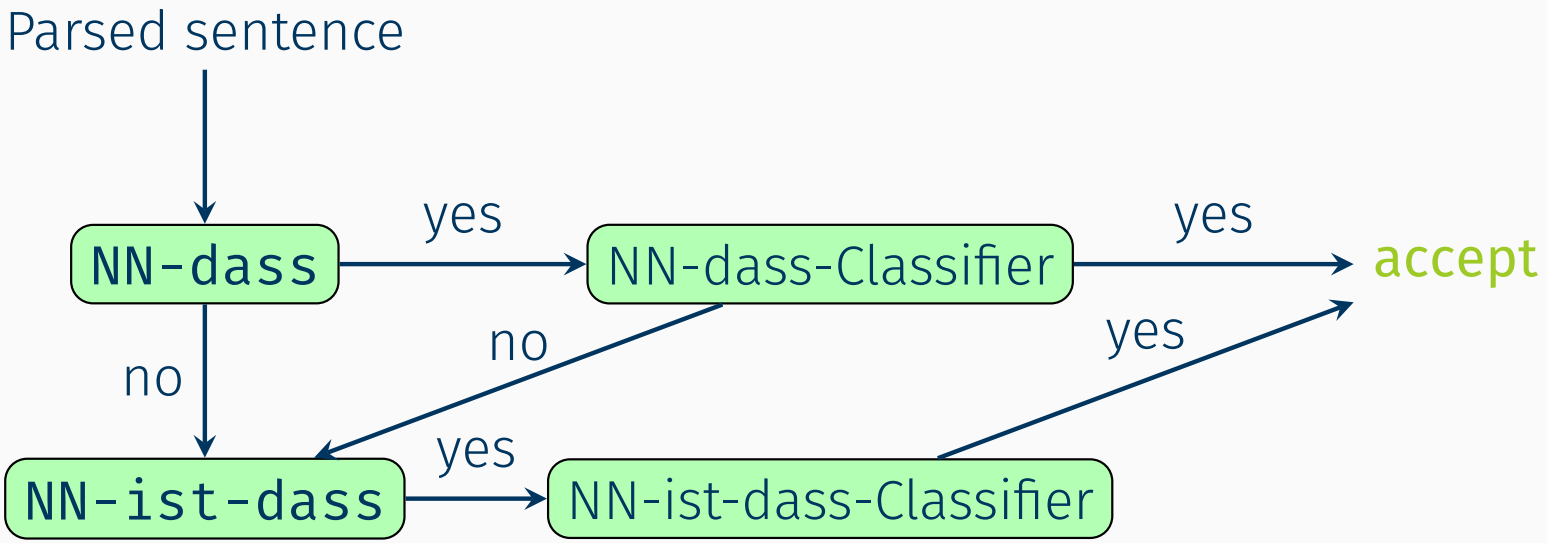
Architecture



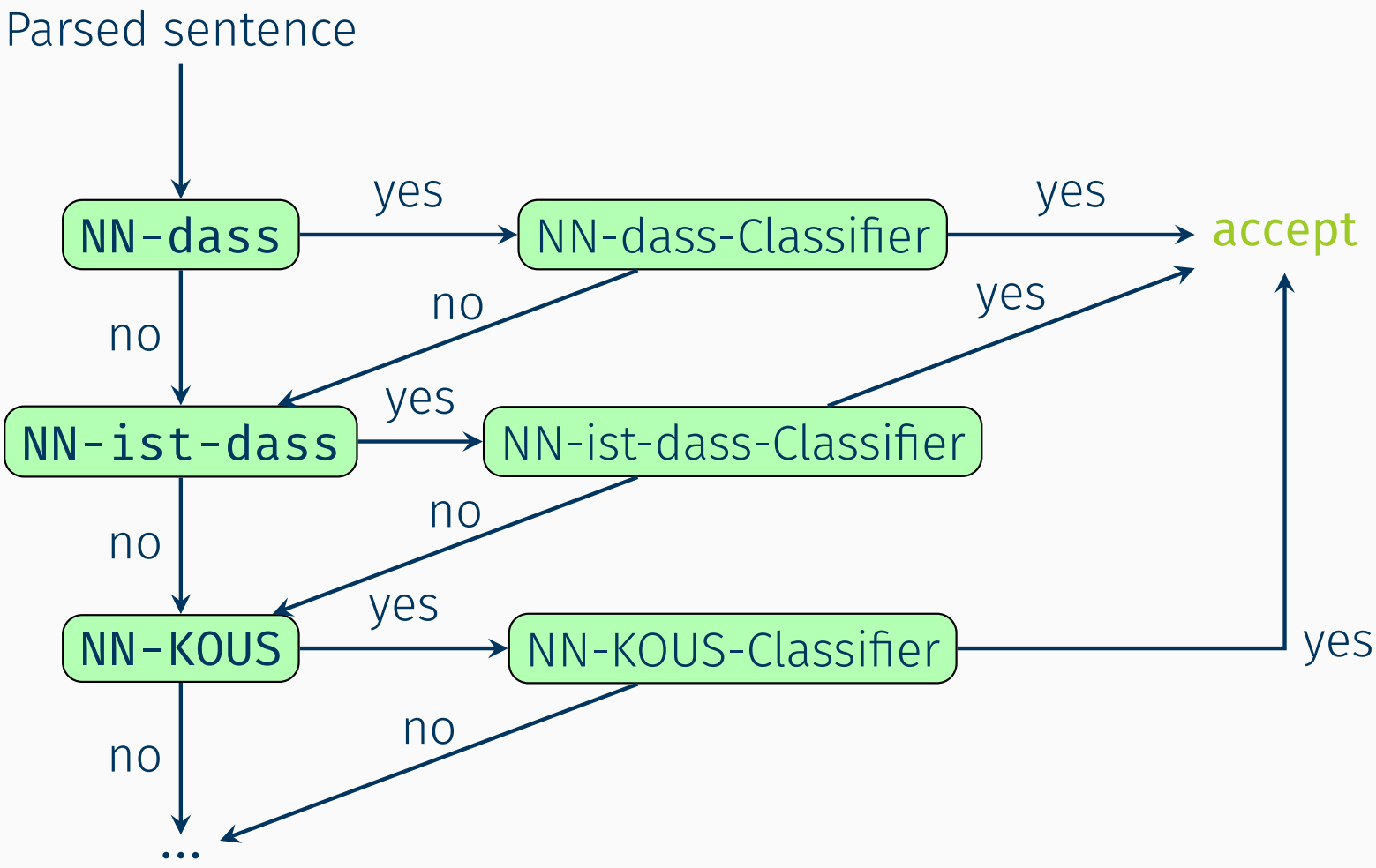
Architecture



Architecture



Architecture



Evaluation

Data

Stems from two annotation projects:

- Dipper and Zinsmeister (2012) annotating the pronouns *dies*, *das*, and *es*
- Simonjetz and Roussel (2016) annotating German (and English) shell nouns

Data

Stems from two annotation projects:

- Dipper and Zinsmeister (2012) annotating the pronouns *dies*, *das*, and *es*
- Simonjetz and Roussel (2016) annotating German (and English) shell nouns

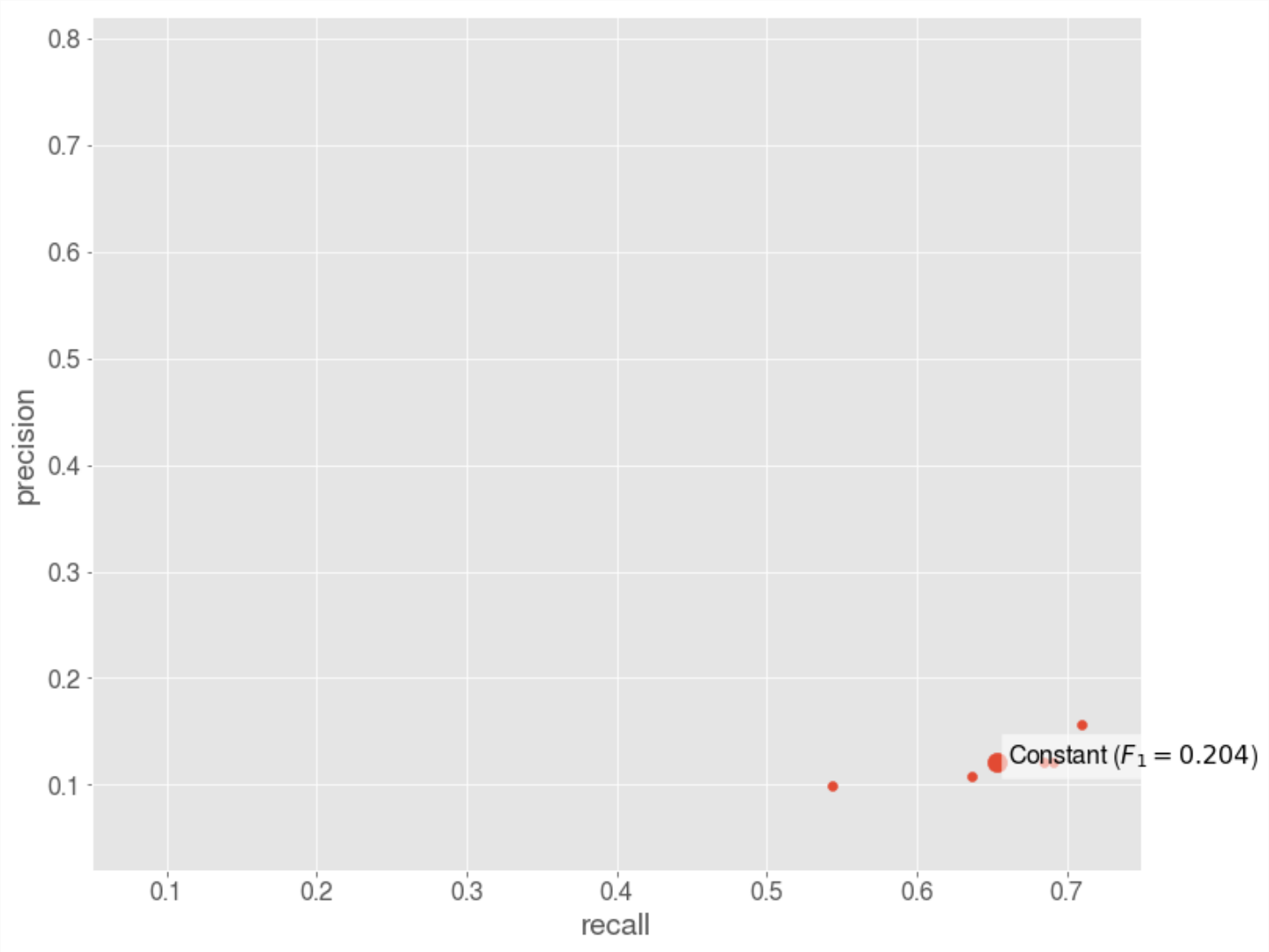
Corpus used here contains 1734 annotated German instances of abstract reference. Of these,

- 1086 are shell nouns
- 249 involve either *dies* or *das*, and
- 375 involve instances of *es*.
- The remaining 24 instances involve pronominal adverbs, such as *deshalb*, and are not covered by this study.

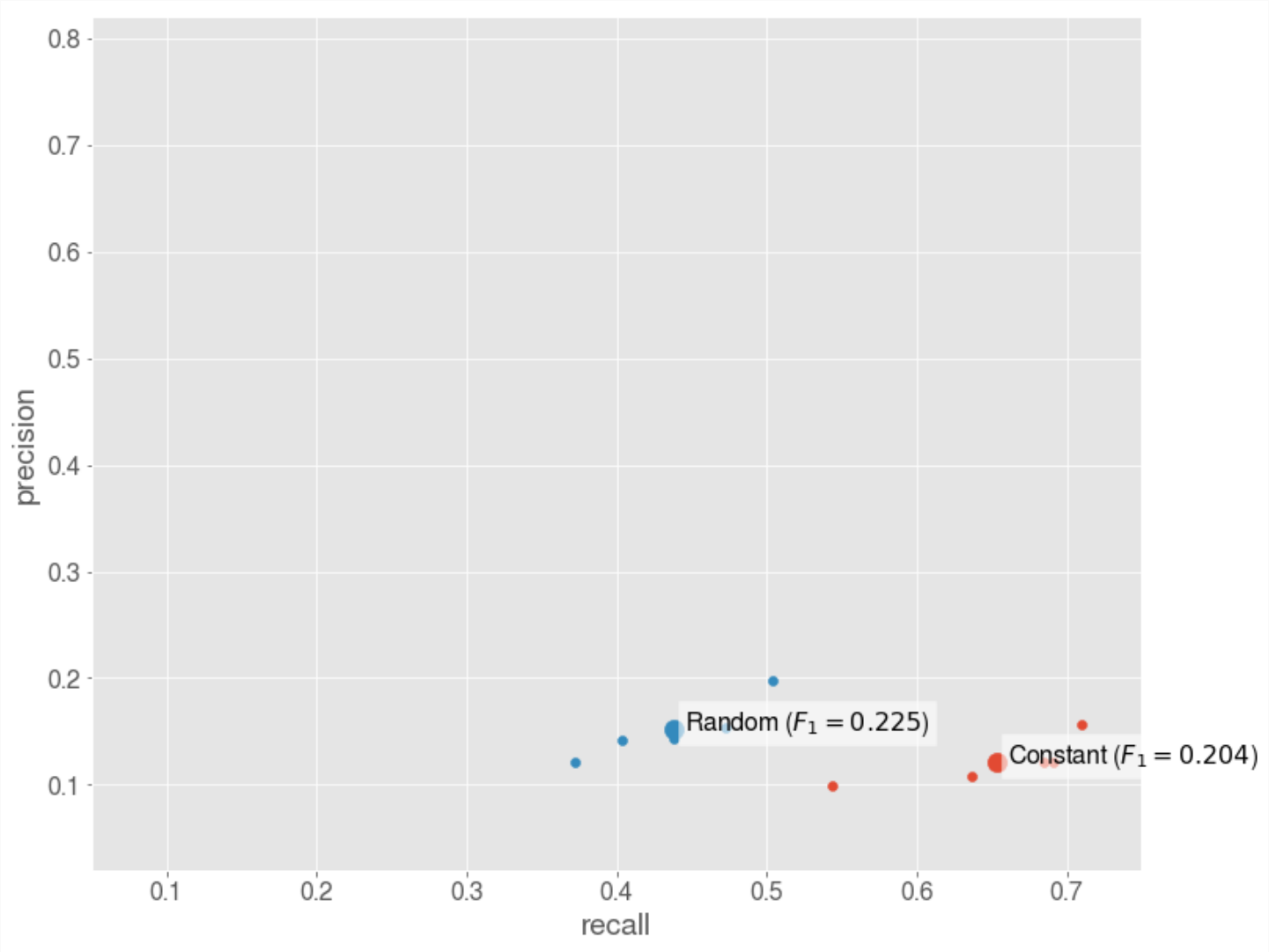
Evaluation

- Comparisons with existing work
  - Classification $\rightarrow$ Was the instance correctly approved by some classifier/pattern?
  - Resolution $\rightarrow$ Was the instance correctly approved by the correct classifier/pattern?
- F_1 scores for classification
- Only accuracy for resolution

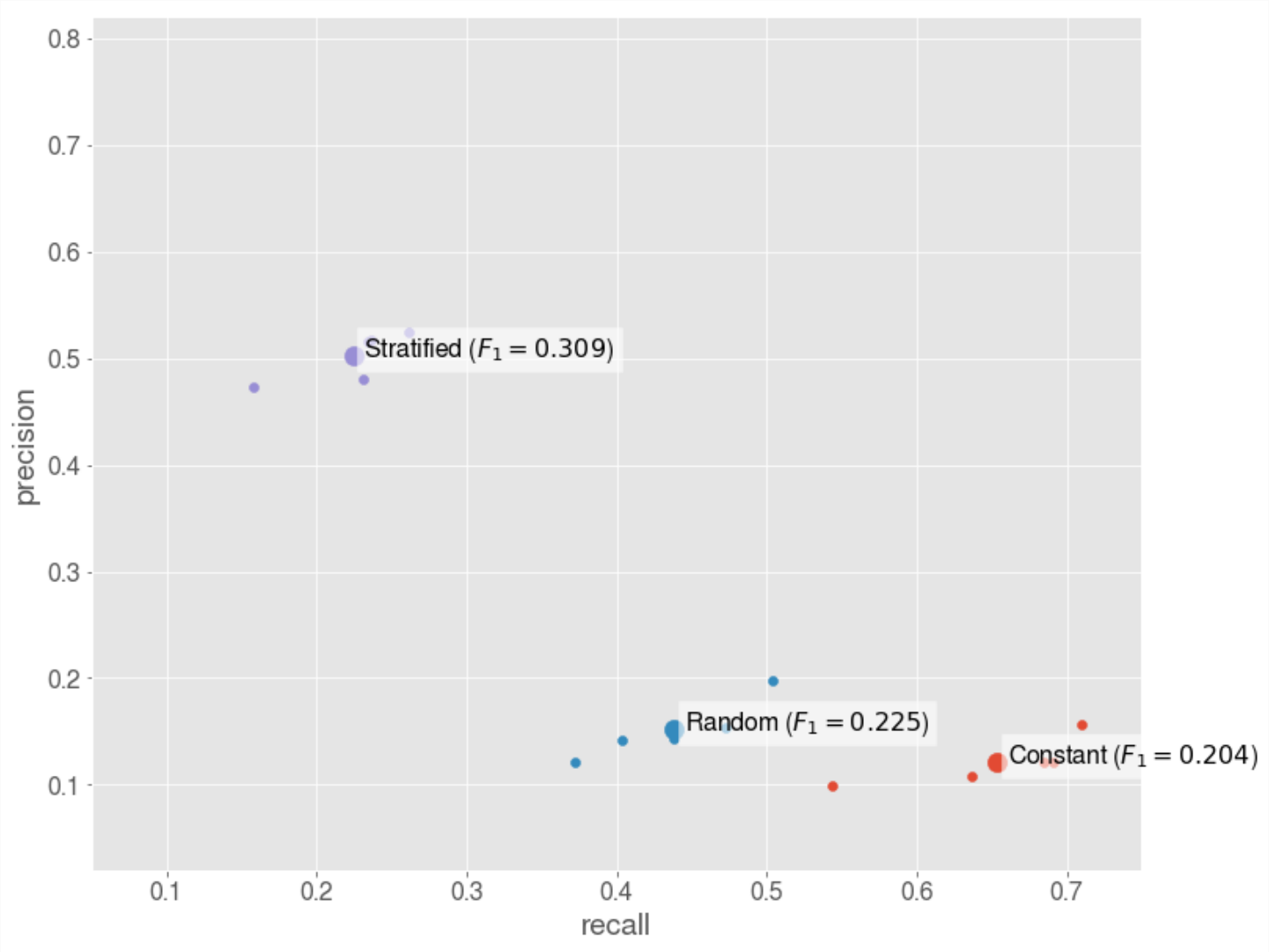
Classifiers and Baselines



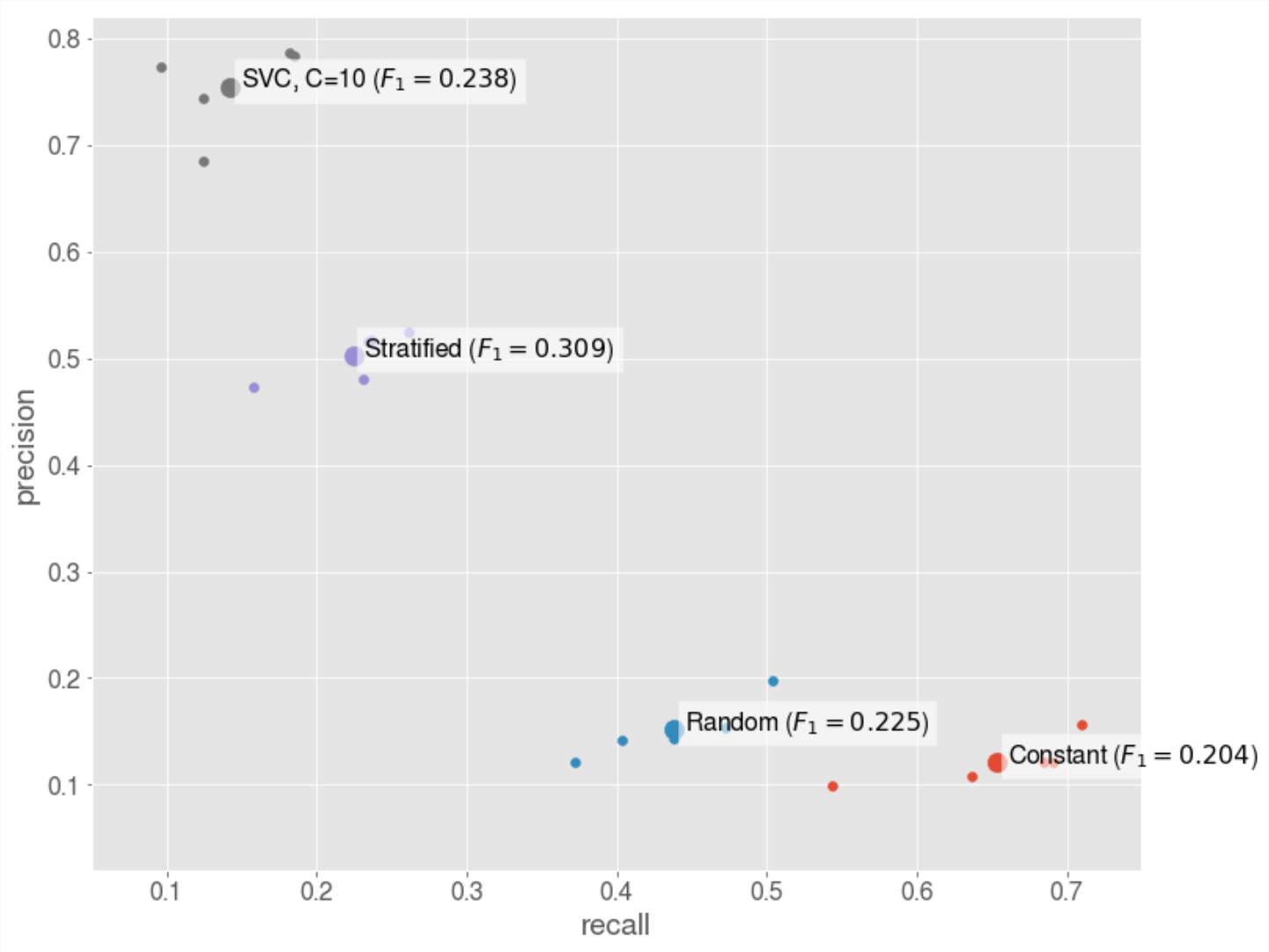
Classifiers and Baselines



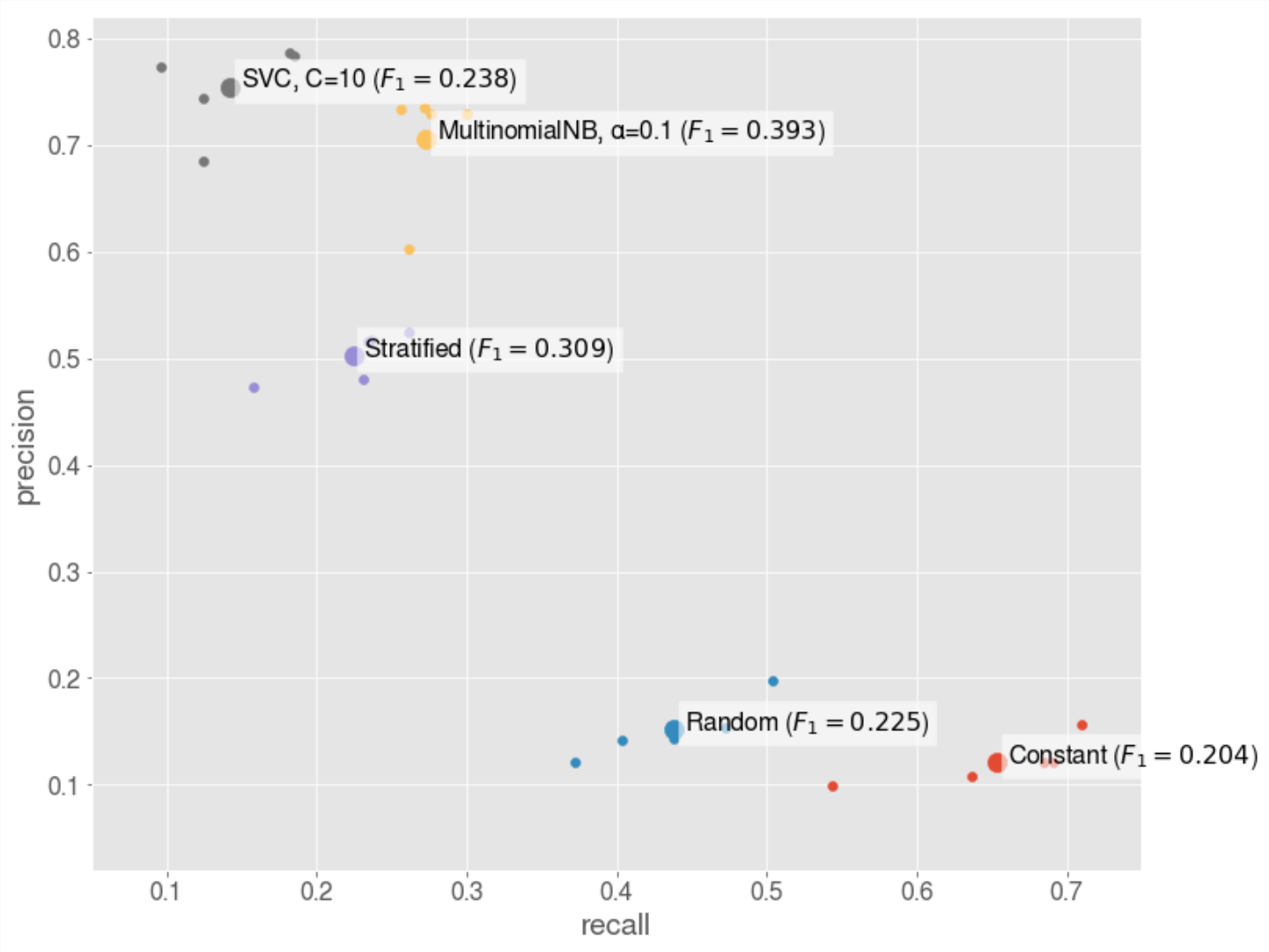
Classifiers and Baselines



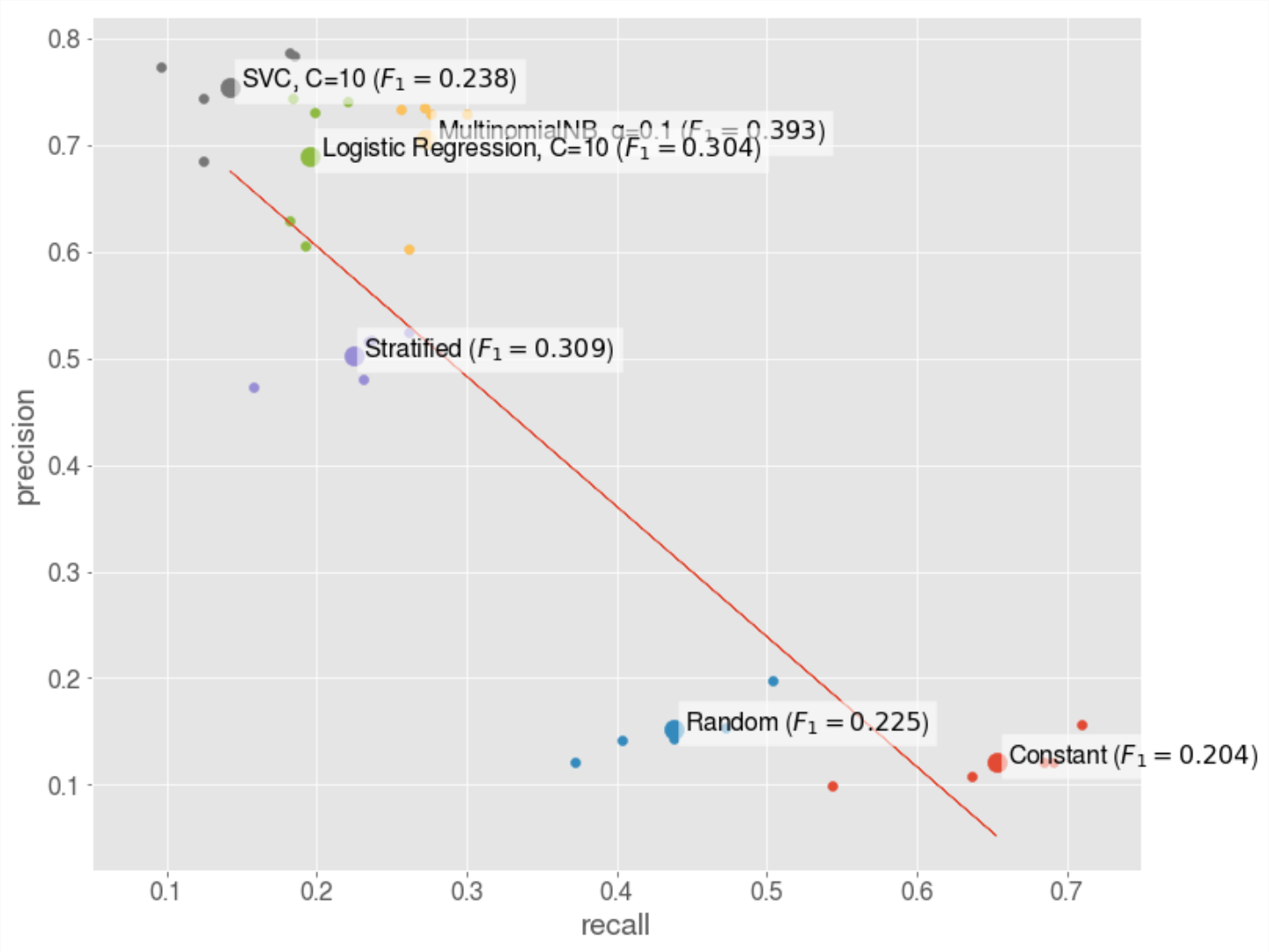
Classifiers and Baselines



Classifiers and Baselines



Classifiers and Baselines



Cataphoric Shell Nouns

- Most similar study, Kolhatkar and Hirst (2014):
  - Baseline (patterns alone), 57%
  - Additional heuristics, 69%
- This system:
  - Baseline (patterns alone): $F_1 = 0.104$, $P = 0.059$, $R = 0.478$
  - Baseline resolution accuracy, 72.2%
  - $\Rightarrow$ 34.5% overall
  - Best classifier: $F_1 = 0.413$, $P = 0.899$, $R = 0.272$
  - With resolution accuracy of 87%
  - $\Rightarrow$ 23.7% overall

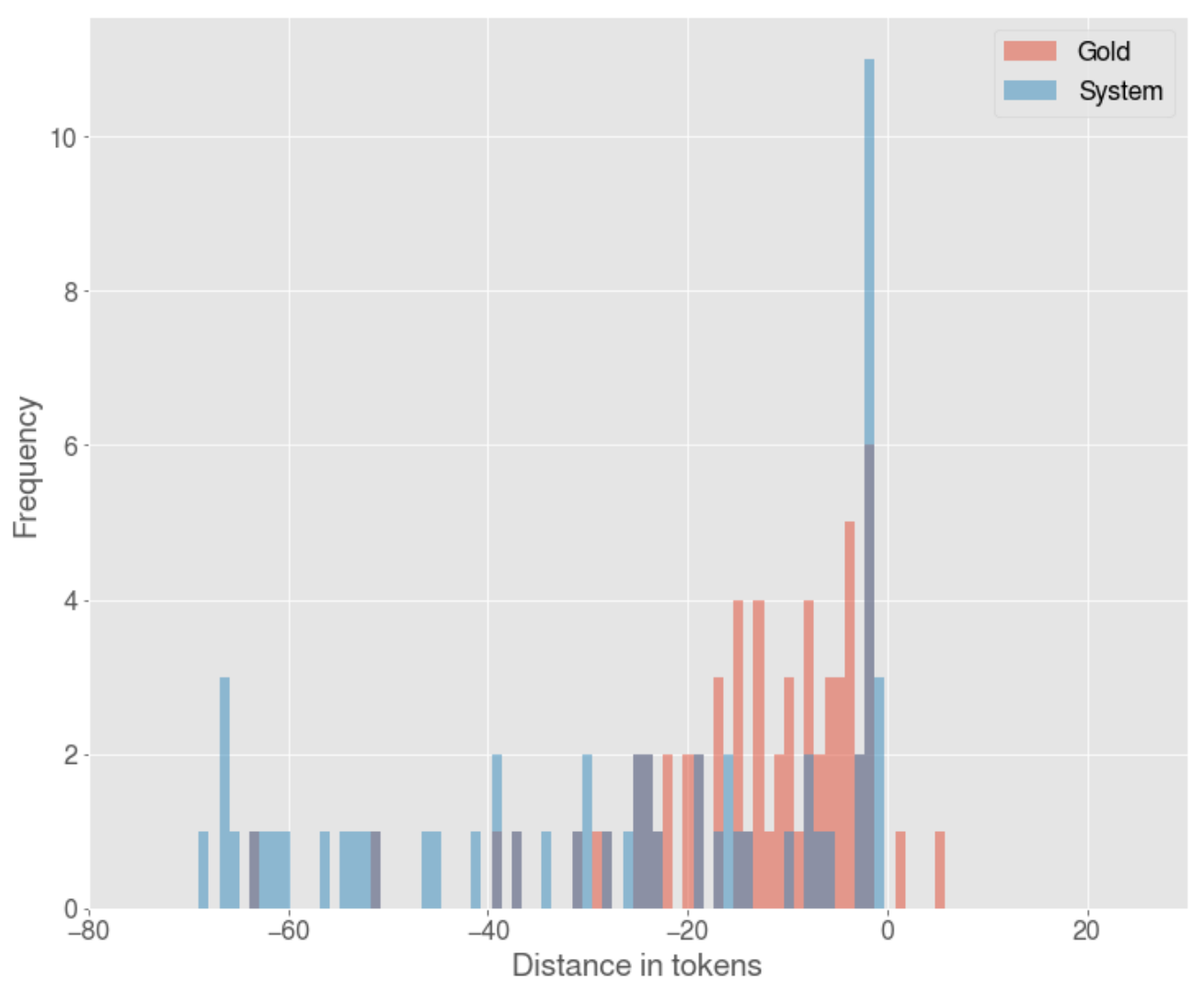
Anaphoric Shell Nouns

- Most similar study, Kolhatkar and Hirst (2012) examined instances of *this issue*
  - Baseline (adjacent sentence): 22.93%
  - With classifier, 59.91%
- This system:
  - Baseline classifier: $F_1 = 0.041$, $P = 0.021$, $R = 0.640$
  - Baseline resolution accuracy, 14.00%
  - $\Rightarrow$ 8.96% overall
  - Best classifier, $F_1 = 0.263$, $P = 0.354$, $R = 0.214$
  - Resolution accuracy of 30.7%
  - $\Rightarrow$ 6.57% overall

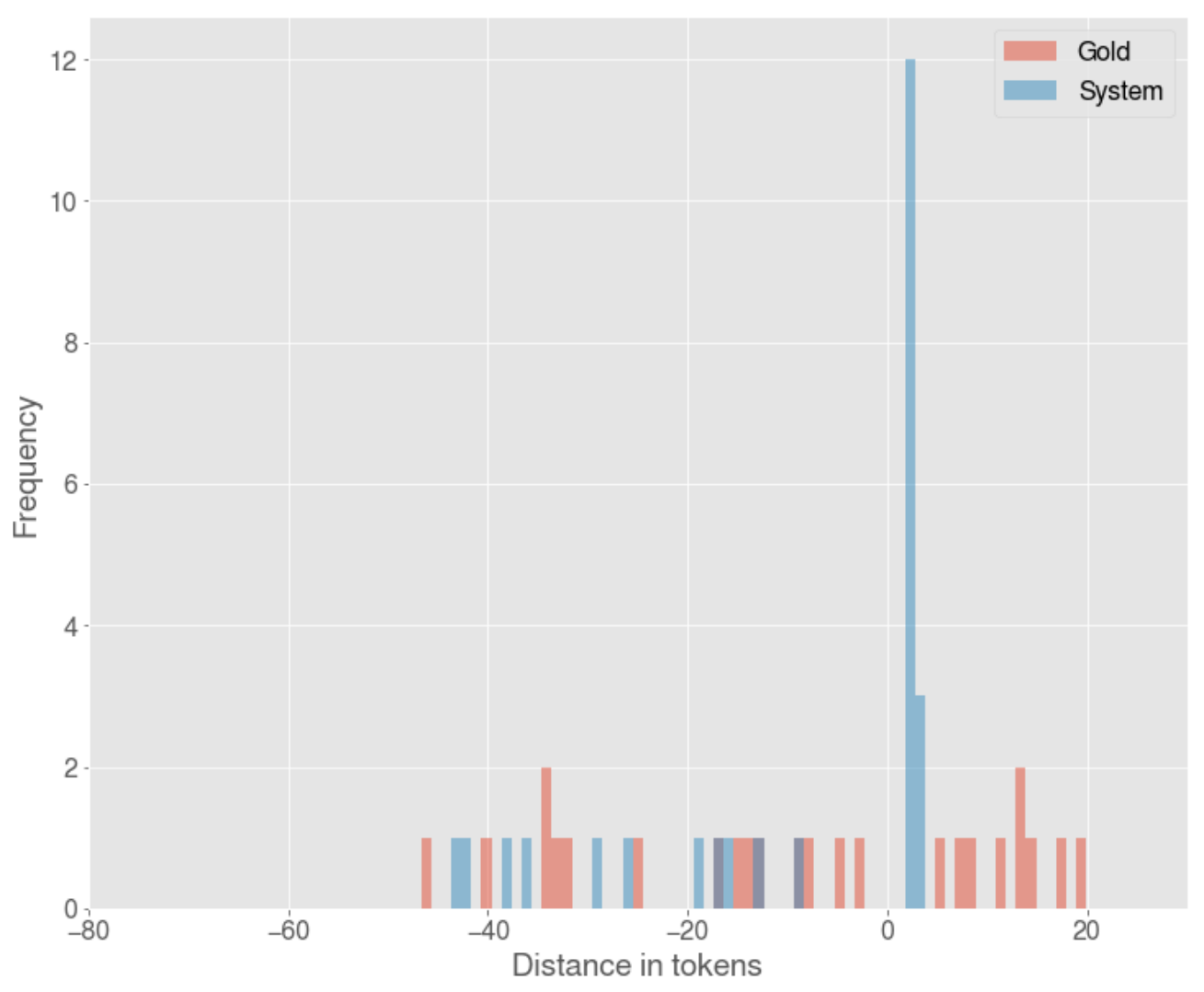
Pronouns

- Most similar study, Jauhar et al. (2015):
  - Classification Baseline: $F_1 = 0.217$, $P = 0.121$, $R = 1.000$
  - Class. + Resolution Baseline: $F_1 = 0.165$, $P = 0.153$, $R = 0.179$
  - Classification: $F_1 = 0.386$, $P = 0.352$, $R = 0.429$
  - Class. + Resolution: $F_1 = 0.222$, $P = 0.226$, $R = 0.218$
- This system:
  - Baseline: $F_1 = 0.590$, $P = 0.430$, $R = 0.946$
  - Baseline resolution accuracy: 12.6%
  - Best classifier, $F_1 = 0.762$, $P = 0.691$, $R = 0.853$
  - Resolution accuracy of 15.4%
  - $\Rightarrow$ 13.14% overall

Extraction pattern errors (Pronouns)



Extraction pattern errors (Shell nouns)



Summary

- Overall, 20–25% of cases correctly assigned some antecedent; of these cases, 50–55% contain the correct antecedent

Summary

- Overall, 20–25% of cases correctly assigned some antecedent; of these cases, 50–55% contain the correct antecedent
- System shows least improvement over baseline for pronouns
  $\rightarrow$ Lack of relevant features
- These features work much better for shell nouns

Summary

- Overall, 20–25% of cases correctly assigned some antecedent; of these cases, 50–55% contain the correct antecedent
- System shows least improvement over baseline for pronouns
  $\rightarrow$ Lack of relevant features
- These features work much better for shell nouns
- Ideas for future work:
  - Better features targeting pronouns
  - Mixing data between similar patterns (i.e., *this* $\approx$ *this NN*)
  - Integrating NP coreference information

Thanks!

https://github.com/ajroussel/aaarg

References i

Artstein, R., & Poesio, M. (2006). Identifying reference to abstract objects in dialogue. In *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue* (pp. 56–63). Potsdam, Germany.

Consten, M., Knees, M., & Schwarz-Friesel, M. (2007). The function of complex anaphors in texts: Evidence from corpus studies and ontological considerations. In M. Schwarz-Friesel, M. Consten, & M. Knees (Eds.), *Anaphors in text: Cognitive, formal and applied approaches to anaphoric reference* (Vol. 86, pp. 81–102). Amsterdam, Netherlands: John Benjamins.

Dipper, S., & Zinsmeister, H. (2012). Annotating abstract anaphora. *Language Resources and Evaluation*, *46*(1), 37–52.

Jauhar, S. K., Guerra, R. D., Gonzàlez, E., & Recasens, M. (2015). Resolving discourse-deictic pronouns: A two-stage approach to do *It*. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics* (pp. 299–308). Denver, CO, USA.

Kolhatkar, V., & Hirst, G. (2012). Resolving "this-issue" anaphora. In *Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing* (pp. 1255–1265). Jeju Island, Korea.

Kolhatkar, V., & Hirst, G. (2014). Resolving shell nouns. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp 2014)* (pp. 499–510). Doha, Qatar.

Lee, H., Chang, A., Peirsman, Y., Chambers, N., Surdeanu, M., & Jurafsky, D. (2013). Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Computational Linguistics*, *39*(4), 885–916.

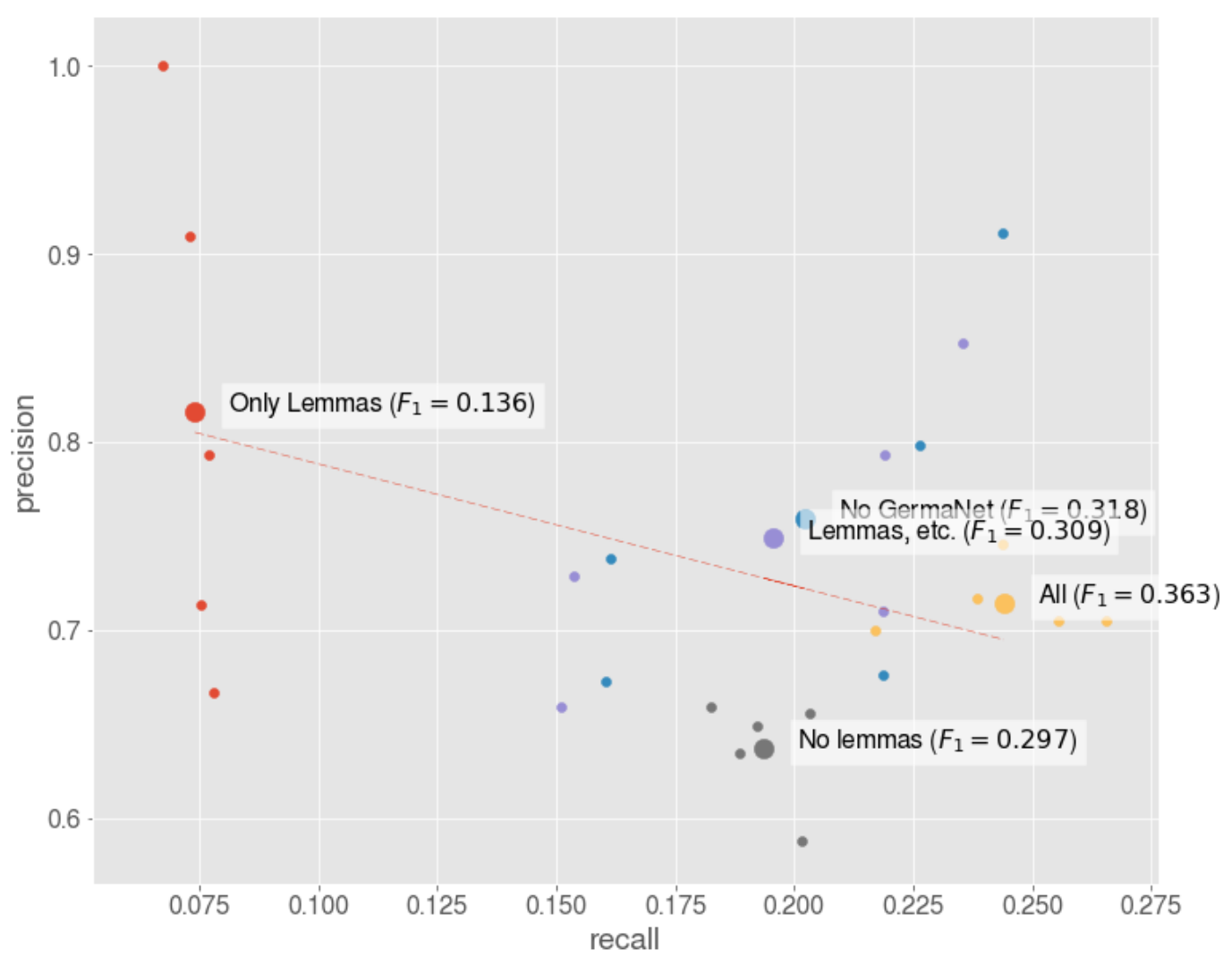
References ii

Schmid, H.-J. (2000). *English abstract nouns as conceptual shells: From corpus to cognition*. Berlin, Germany: de Gruyter.

Simonjetz, F., & Roussel, A. (2016). Crosslinguistic annotation of German and English shell noun complexes. In *Proceedings of the 13th Conference on Natural Language Processing (KONVENS)* (pp. 265–278). Bochum, Germany.

Webber, B. L. (1988). *Discourse deixis and discourse processing* (Tech. Rep. No. 610). University of Pennsylvania.

Feature Set Comparison



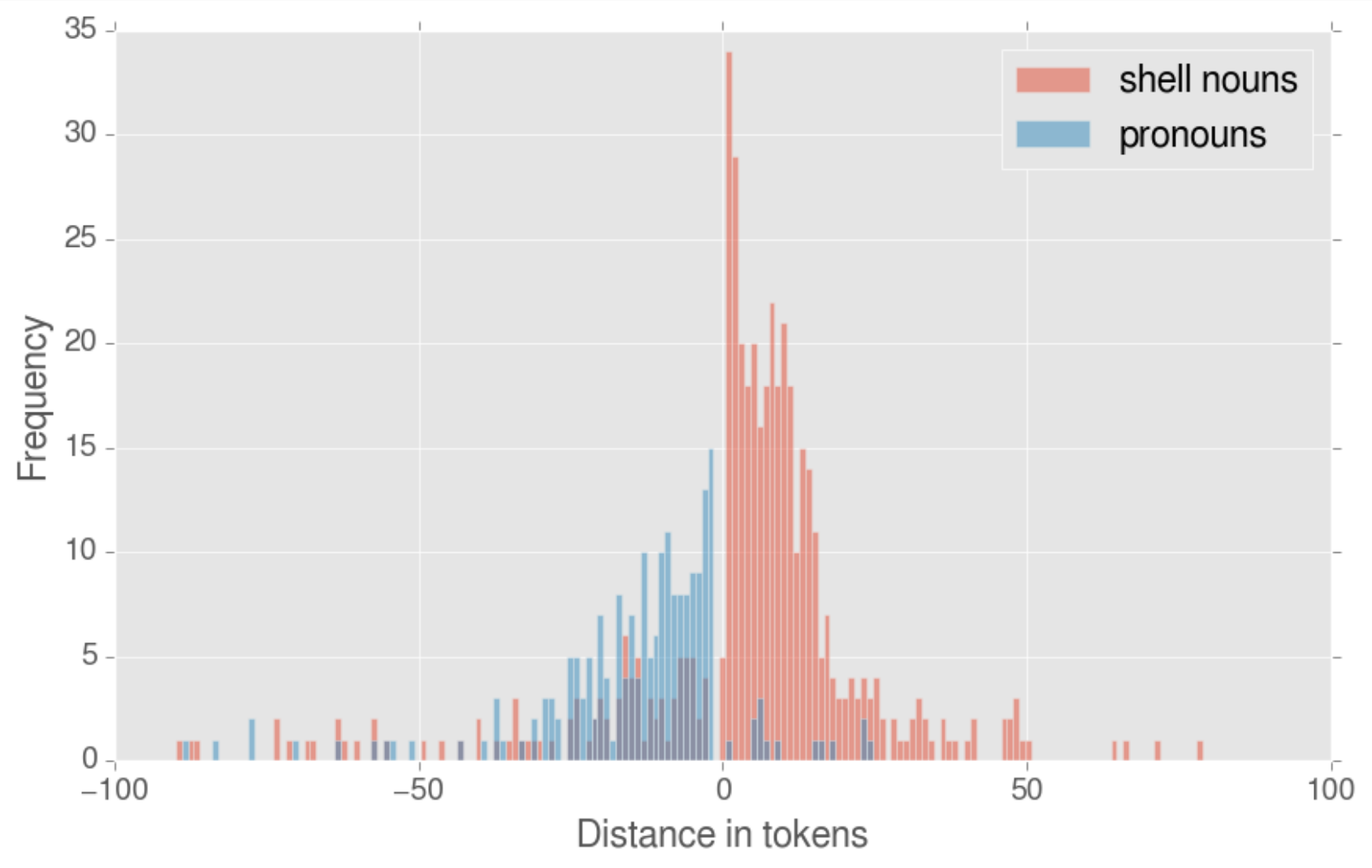
Complete Feature Set

Feature	Examples
Anaphor	
Lemma	*das*, *es*, *Umstand*
Number	Sing./Pl.
Grammatical function	*subj*, *obja*
Whether parent precedes anaphor	Yes/No
Whether parent is subjunctive	Yes/No
Whether parent is clausal verb	Yes/No
Semantic field	*Attribut*, *Kommunikation*
Parent semantic field	*Gefühl*, *Perzeption*
Semantic fields of dep. adjectives	*Bewegung*, *Menge*
Whether dep. article is definite or indefinite	Yes/No
Dep. determiners	*dieser*, *kein*, *beiden*

Complete Feature Set

Antecedent	
Dependent preposition lemmas	*zu*, *für*, *nach*
Dependent complementizers	*dass*, *ob*, *weil*
Grammatical function	*root*, *objc*
Length	No. of tokens
Gender	*Masc*, *Fem*, *Neut*
Semantic field	*Attribut*, *Kommunikation*
Embedding depth	No. of deps. to sentence root
If nominal, ending	*-ung*, *-heit*, *-en*
Whether antecedent contains question mark	Yes/No
Relation	
Distance between anaphor/antecedent	No. of tokens
Whether anaphor precedes antecedent	Yes/No
Whether anaphor/antecedent funcs. match	Yes/No
Whether colon between anaphor/antecedent	Yes/No

Abstract Anaphora Distribution



Classifiers and Baselines

	Classification			Resolution
Name	Precision	Recall	F_1-score	Accuracy
Constant	0.121	**0.654**	0.204	0.353
Random	0.155	0.446	0.230	0.359
Stratified	0.518	0.234	0.322	0.259
MultinomialNB, $\alpha = 0.1$	0.674	0.244	**0.356**	0.525
SVC, $C = 10$	0.742	0.132	0.224	0.340
Logistic Regression, $C = 10$	0.722	0.185	0.292	0.559
Voting	**0.774**	0.179	0.288	**0.563**

Table 1: Classification performance and resolution accuracy

Cataphoric Shell Nouns

	Classification			Resolution
Name	Precision	Recall	F_1-score	Accuracy
Constant	0.059	**0.478**	0.104	0.722
Random	0.062	0.255	0.100	0.658
Stratified	0.233	0.102	0.140	0.910
MultinomialNB, $\alpha = 0.1$	0.729	0.269	0.391	0.836
SVC, $C = 10$	0.736	0.128	0.216	**0.928**
Logistic Regression, $C = 10$	**0.899**	0.272	**0.413**	0.870
Voting	0.893	0.254	0.390	0.885

Table 2: System performance for cataphoric shell noun instances

Anaphoric Shell Nouns

	Classification			Resolution
Name	Precision	Recall	F_1-score	Accuracy
Constant	0.021	**0.640**	0.041	0.140
Random	0.020	0.272	0.046	0.113
Stratified	0.056	0.083	0.084	**0.375**
MultinomialNB, $\alpha = 0.1$	0.354	0.214	0.263	0.307
SVC, $C = 10$	0.000	0.000	NaN	NaN
Logistic Regression, $C = 10$	**0.442**	0.119	**0.293**	0.083
Voting	0.428	0.109	0.274	0.083

Table 3: System performance for anaphoric shell noun instances

Pronouns

	Classification			Resolution
Name	Precision	Recall	F_1-score	Accuracy
Constant	0.430	**0.946**	0.590	0.126
Random	0.499	0.767	0.603	**0.164**
Stratified	0.658	0.835	0.733	0.135
MultinomialNB, $\alpha = 0.1$	0.691	0.819	0.749	0.150
SVC, $C = 10$	0.691	0.853	**0.762**	0.154
Logistic Regression, $C = 10$	0.693	0.808	0.745	0.160
Voting	**0.697**	0.820	0.752	0.158

Table 4: System performance for pronominal abstract anaphora

Per-anaphor classification performance

Anaphor	*N*	Recall	Precision	F_1-score
Zusicherung 'pledge'	1	1.000	1.000	1.000
Notwendigkeit 'need'	7	0.714	1.000	0.833
Tatsache 'fact'	21	0.684	1.000	0.813
Überzeugung 'conviction'	8	0.667	1.000	0.800
Versuch 'attempt'	7	0.571	1.000	0.727
Ansicht 'view'	42	0.471	0.889	0.615
Forderung 'demand'	26	0.450	0.900	0.600
dies 'this'	59	0.420	0.913	0.575
das 'that'	191	0.430	0.829	0.566
Meinung 'opinion'	30	0.364	1.000	0.533
Argument 'argument'	5	0.333	1.000	0.500

Table 5: Classification performance of Naive Bayes classifier for particular anaphors

Difficult cases

(5) **Eines** ist auch klar, und **dazu** stehen wir auch: *Kontrolle ist gut, Vertrauen in Kontrolle ist besser*. Aber **das** wird der Vorschlag dieser vier Staaten bei den Bürgerinnen und Bürgern in dieser Form nicht erreichen.
‘One thing is clear and we stand behind this too: Control is good, trust in control is better. But the proposal of these four states in this form won’t achieve this.’

Difficult cases

(6) *Wir*, die Vertreter der Bürger, der Menschen Europas, erreichen innerhalb der Institutionen etwas, und wir debattieren darüber kurz vor Mitternacht und werden ermahnt, wir sollen uns kurz fassen. **Das** ist ein Wiederholungsfall!

'Within the institutions, we – the representatives of the citizens, of the people of Europe – are actually achieving something, and, as we debate it shortly before midnight, we are enjoined to keep things brief. This is not the first time this has happened.'

Beamer Theme Credits

Get the source of this theme and the demo presentation from

`github.com/matze/mtheme`