

Testing the predictive power of features from linguistic theory: A study on the interpretation of deverbal compounds


Lonneke van der Plas
University of Malta

Joint work with Gianina Iordăchioaia, Glorianna Jagfeld, Patrick Ziering, Prajit Dhar, Stefan Mueller

Plan for this talk

- Introduction to compounds
- Overview of work done on compounds in project D11 and beyond
 - Compound discovery
 - Compound bracketing
 - Compound splitting
 - Predicting novel compounds
- Study on the interpretation of deverbal compounds

Plan for this talk

- Introduction to compounds 
- Overview of work done on compounds in project D11 and beyond
 - Compound discovery
 - Compound bracketing
 - Compound splitting
 - Predicting novel compounds
- Study on the interpretation of deverbal compounds

What are compounds?

The formation of a new lexeme by adjoining two or more lexemes

[Bauer, 2003]

What are compounds?

The formation of a new lexeme by adjoining two or more lexemes

[Bauer, 2003]

Let's look at some examples (gathered during my holidays)

Some examples

trail mix,



Some examples

trail mix, dry bag,



Some examples

trail mix, dry bag, forest firefighter,



Some examples

trail mix, dry bag, forest firefighter, chipmunk,



By Eborutta at German Wikipedia - Own work Transferred from de.wikipedia to Commons by Saperaud commonswiki., CC BY-SA 3.0

<https://commons.wikimedia.org/w/index.php?curid=327802>

Some examples

trail mix, dry bag, forest firefighter, chipmunk, salmon consumption,



Some examples

trail mix, dry bag, forest firefighter, chipmunk, salmon consumption, mosquito spray,



Some examples

trail mix, dry bag, forest firefighter, chipmunk, salmon consumption, mosquito spray, pit toilet,



Some examples

trail mix, dry bag, forest firefighter, chipmunk, salmon consumption, mosquito spray, pit toilet, float plane,



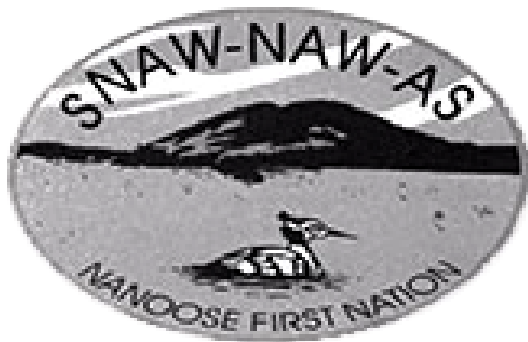
Some examples

trail mix, dry bag, forest firefighter, chipmunk, salmon consumption, mosquito spray, pit toilet, float plane, treeline hike,



Some examples

trail mix, dry bag, forest firefighter, chipmunk, salmon consumption, mosquito spray, pit toilet, float plane, treeline hike, First Nations campground,



Some examples

trail mix, dry bag, forest firefighter, chipmunk, salmon consumption, mosquito spray, pit toilet, float plane, treeline hike, First Nations campground, maple syrup



Some examples

trail mix, dry bag, forest firefighter, chipmunk, salmon consumption, mosquito spray, pit toilet, float plane, treeline hike, First Nations campground, maple syrup

noun-noun vs adjective-noun, or verb-noun

Some examples

trail mix, dry bag, forest firefighter, **chipmunk**, salmon consumption, **mosquito spray**, pit toilet, float plane, treeline hike, First Nations campground, maple syrup

More compositional vs **less compositional**

Some examples

trail mix, dry bag, forest firefighter, chipmunk, salmon consumption, mosquito spray, pit toilet, float plane, **treeline hike**, First Nations campground, maple syrup

Frequent phenomenon, but **low token counts**

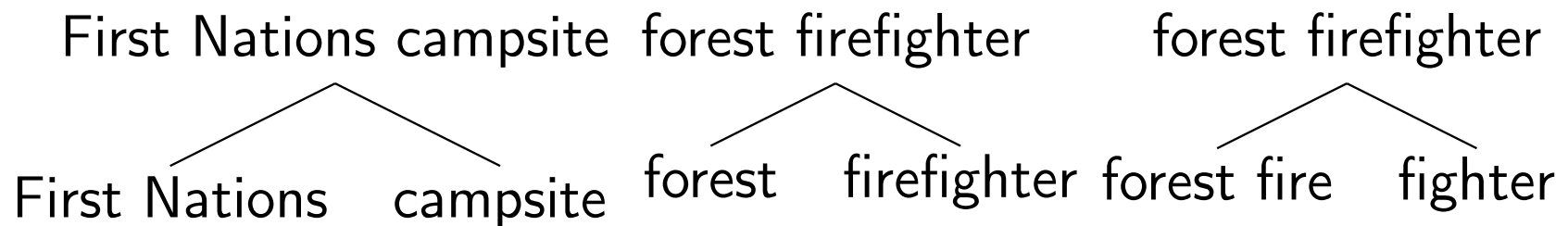
- Compounding is a productive word formation process
- English-speaking children can produce compounds from a very young age
- Hard case for statistical methods in NLP
- Compositional approach, where possible

Some examples

trail mix, dry bag, forest firefighter, chipmunk, salmon consumption, mosquito spray, pit toilet, float plane, treeline hike, First Nations campground, maple syrup

Two constituents vs more constituents

- And how do we parse it?
- Right or left branching?
- Or does it not matter much?



Some examples

trail mix, dry bag, forest firefighter, chipmunk, salmon consumption, mosquito spray, pit toilet, float plane, treeline hike, First Nations campground, maple syrup

Role of constituents,

- Last constituent is the **head**, and other constituents are called **modifiers or non-heads**
- Head → main category, modifier → specialisation
- Nominal compounds are compounds whose head is a noun

Some examples

trail mix, **dry bag**, forest firefighter, chipmunk, salmon consumption, mosquito spray, pit toilet, float plane, treeline hike, First Nations campground, maple syrup

Covert relation between constituents:

A **bag** that keeps things **dry**



Some examples

trail mix, dry bag, forest firefighter, chipmunk, **salmon consumption**, mosquito spray, pit toilet, float plane, treeline hike, First Nations campground, maple syrup

Covert relation between constituents:

If the head of the compound is deverbal (deverbal compound), the argument structure of the verb comes into play

Salmon OBJ **consumption**



Some examples

trail mix, dry bag, forest firefighter, chipmunk, salmon consumption, **mosquito spray**, pit toilet, float plane, treeline hike, First Nations campground, maple syrup

Covert relation between constituents:

A **spray** to be used against **mosquitos**



Some examples

trail mix, dry bag, **forest firefighter**, chipmunk, salmon consumption, mosquito spray, pit toilet, float plane, treeline hike, First Nations campground, maple syrup

Covert relation between constituents:

A person who fights OBJ **fires** in forests



Some examples

trail mix, dry bag, forest firefighter, chipmunk, salmon consumption, mosquito spray, pit toilet, float plane, treeline hike, First Nations campground, maple syrup

Cross-lingual observations:

- Complex nominals (e.g., French: peine de mort)
- Open compounding (e.g., English: trail mix)
- Closed compounding (e.g., German: Kraftfahrzeughaftpflichtversicherung)


Project D11: some facts

- A crosslingual approach to the analysis of compound nouns
- 1 out of 29 projects in the collaborative research centre SFB 732 on Incremental Specification in Context (University of Stuttgart)
- German Research Foundation (2014-2018)
- PhD students:
 - Patrick Ziering
 - Stefan Müller
- Collaborations:
 - Gianina Iordachioaia
- Couple of student assistants:
 - Glorianna Jagfeld,
 - Prajit Dhar
 - ...

Work in project D11

- **Compound discovery** from parallel corpora
[Ziering and Van der Plas, Coling 2014], [Ziering and Van der Plas, *Structure forthcoming*]
- Unsupervised **compound splitting** for Germanic languages using inflectional information and distributional semantics
[Jagfeld et al., ACL 2017], [Ziering and Van der Plas, NAACL 2016], [Ziering, Mueller, and Van der Plas, MWE@ACL 2016]
- Cross-lingual approach to **bracketing compounds**
[Ziering and Van der Plas, Coling 2014], [Ziering and Van der Plas, IWCS 2015], [Ziering and Van der Plas, RANLP 2015]
- Interpretation of **deverbal compounds** based on features of the head
[Iordachioaia et al., GramLex 2016], [Iordachioaia et al., PMWE bookseries *forthcoming*]
- Generating novel compounds
[Dhar and Van der Plas, MWE-WN 2019]

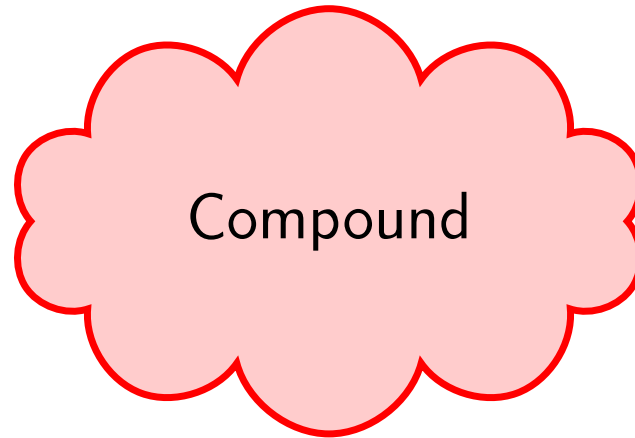
Plan for this talk

- Introduction to compounds
- Overview of work done on compounds in project D11 and beyond
 - Compound discovery 
 - Compound bracketing
 - Compound splitting
 - Predicting novel compounds
- Study on the interpretation of deverbal compounds

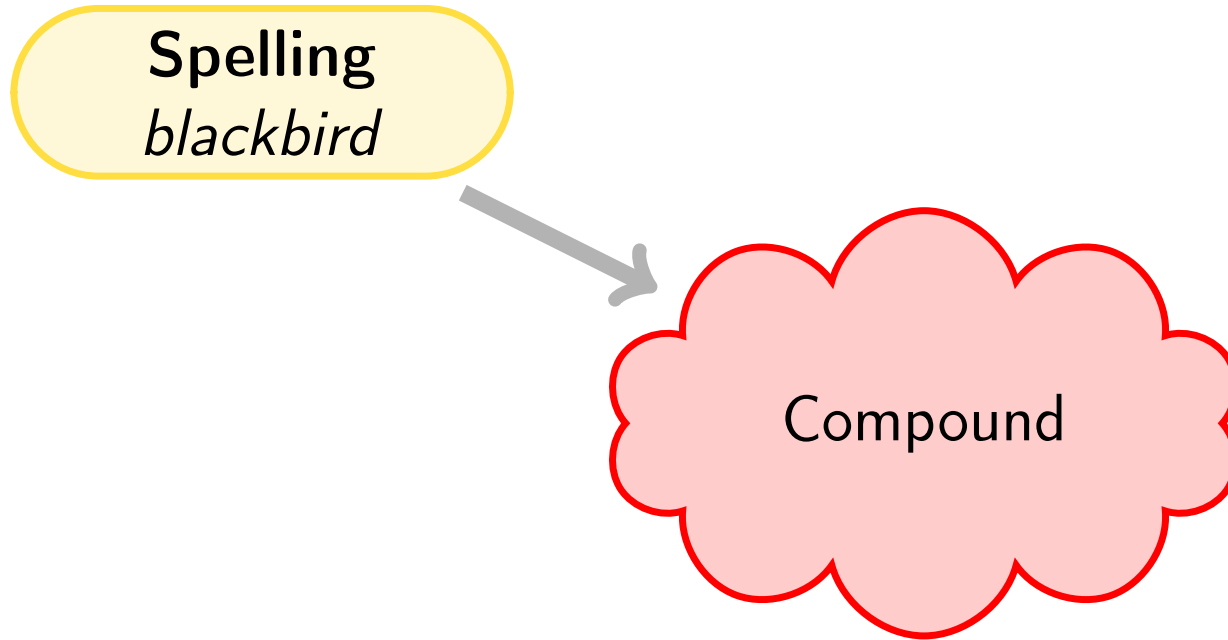
Multilingual compound database

- Aim: Constructing a database of compounds and their translations in several languages
- Need to identify compounds
- Problem: no universally accepted definition of compoundhood
- Hard to distinguish between phrases and compounds
- However, there are some (non-waterproof) linguistic tests

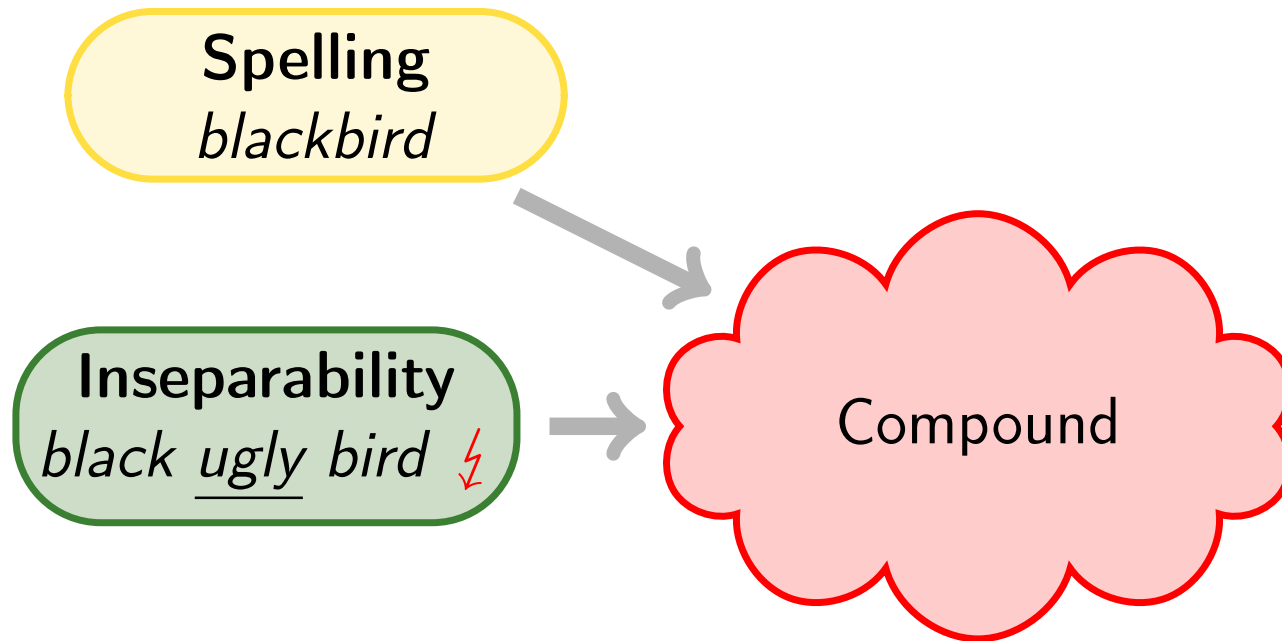
Linguistic Criteria [Lieber and Štekauer, 2009]



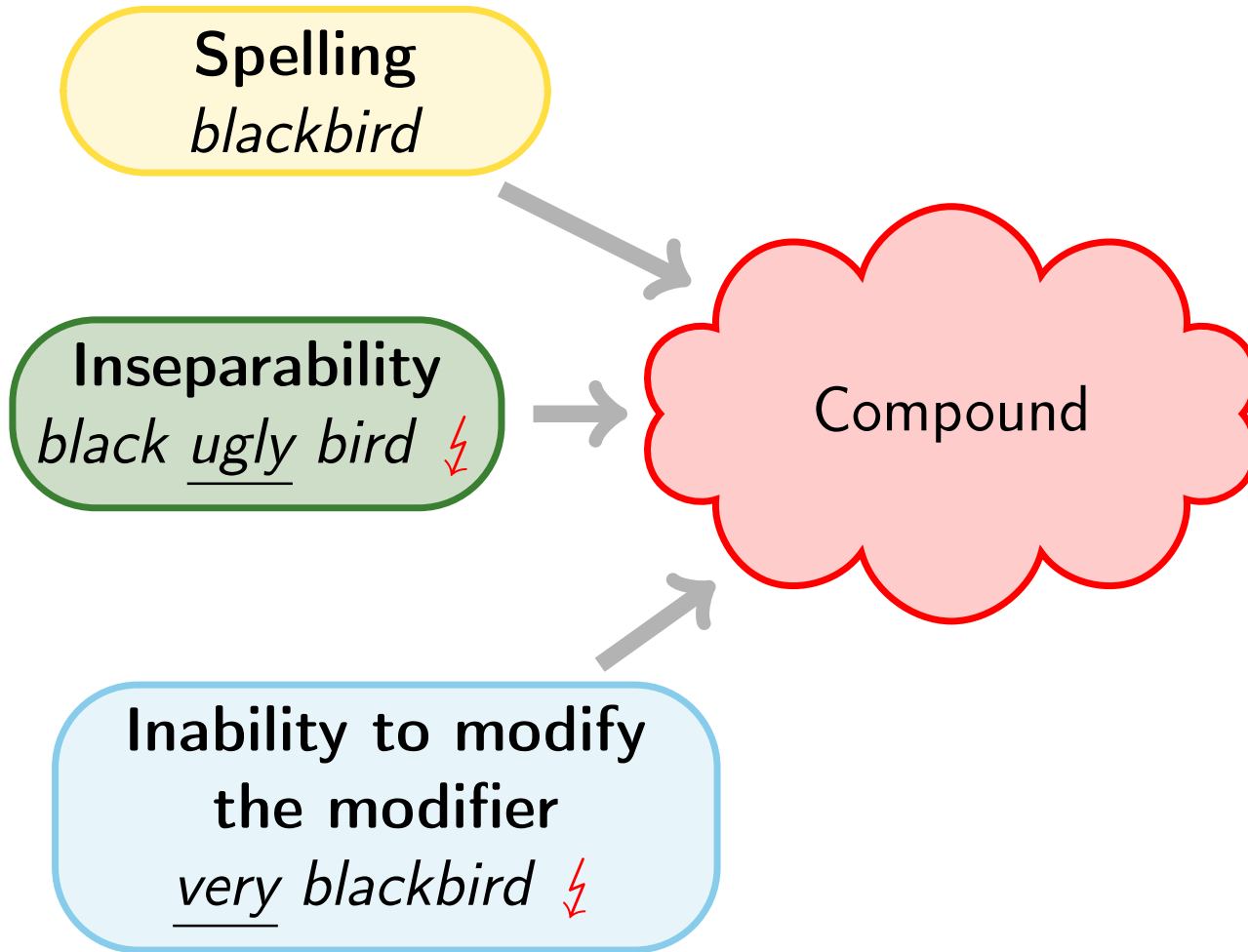
Linguistic Criteria [Lieber and Štekauer, 2009]



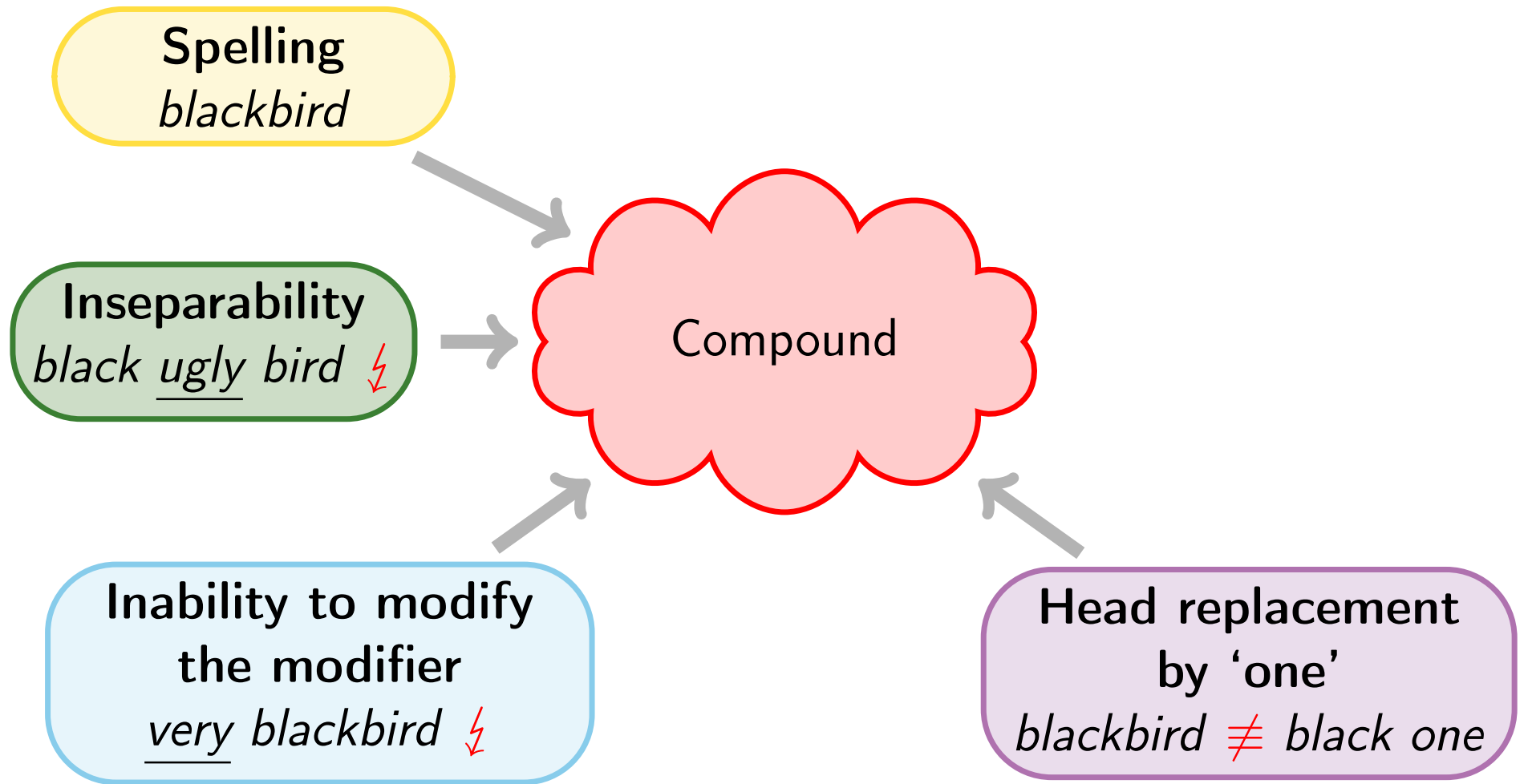
Linguistic Criteria [Lieber and Štekauer, 2009]



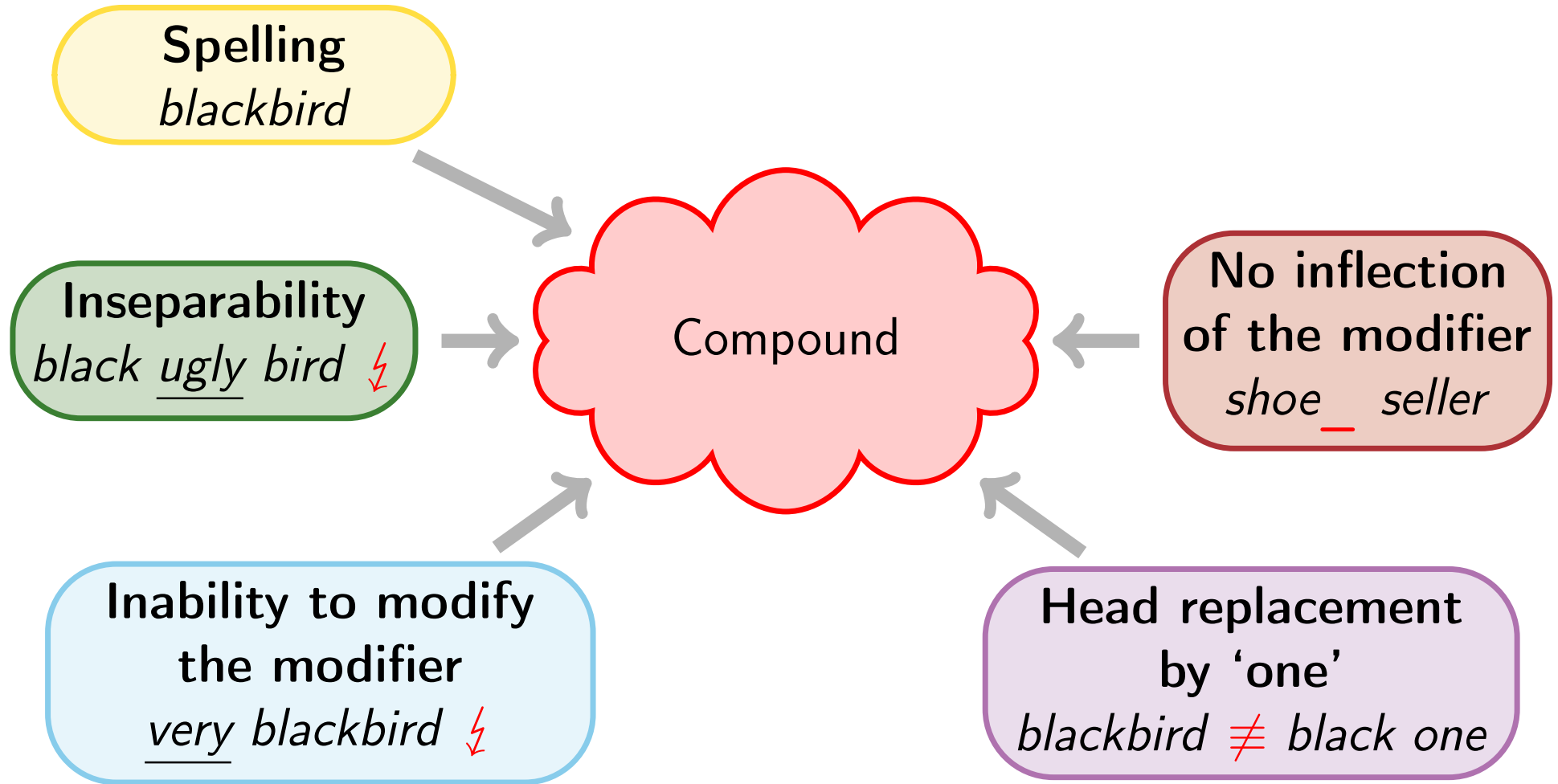
Linguistic Criteria [Lieber and Štekauer, 2009]



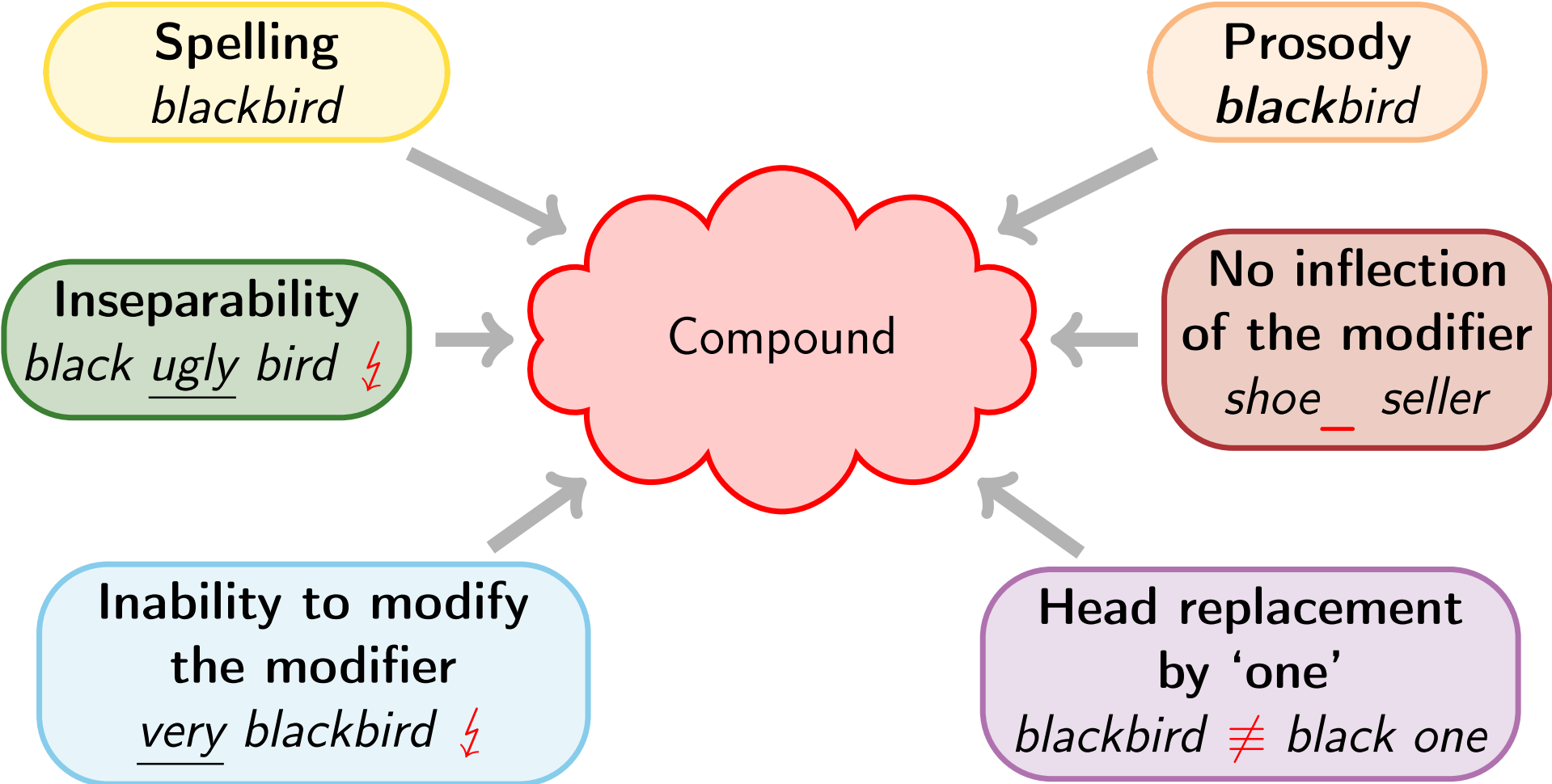
Linguistic Criteria [Lieber and Štekauer, 2009]



Linguistic Criteria [Lieber and Štekauer, 2009]



Linguistic Criteria [Lieber and Štekauer, 2009]



How well defined is compoundhood?

- Which linguistic criteria are **most relevant** for **compoundhood**?
- Do people **agree** when identifying compounds?

Approach

- Let two trained annotators identify candidate compounds in extracted sentences and rate these for
 - Degree of compoundhood
 - All six linguistic criteria
- Look at agreements
- Machine learning experiments to determine strongest predictors for compoundhood

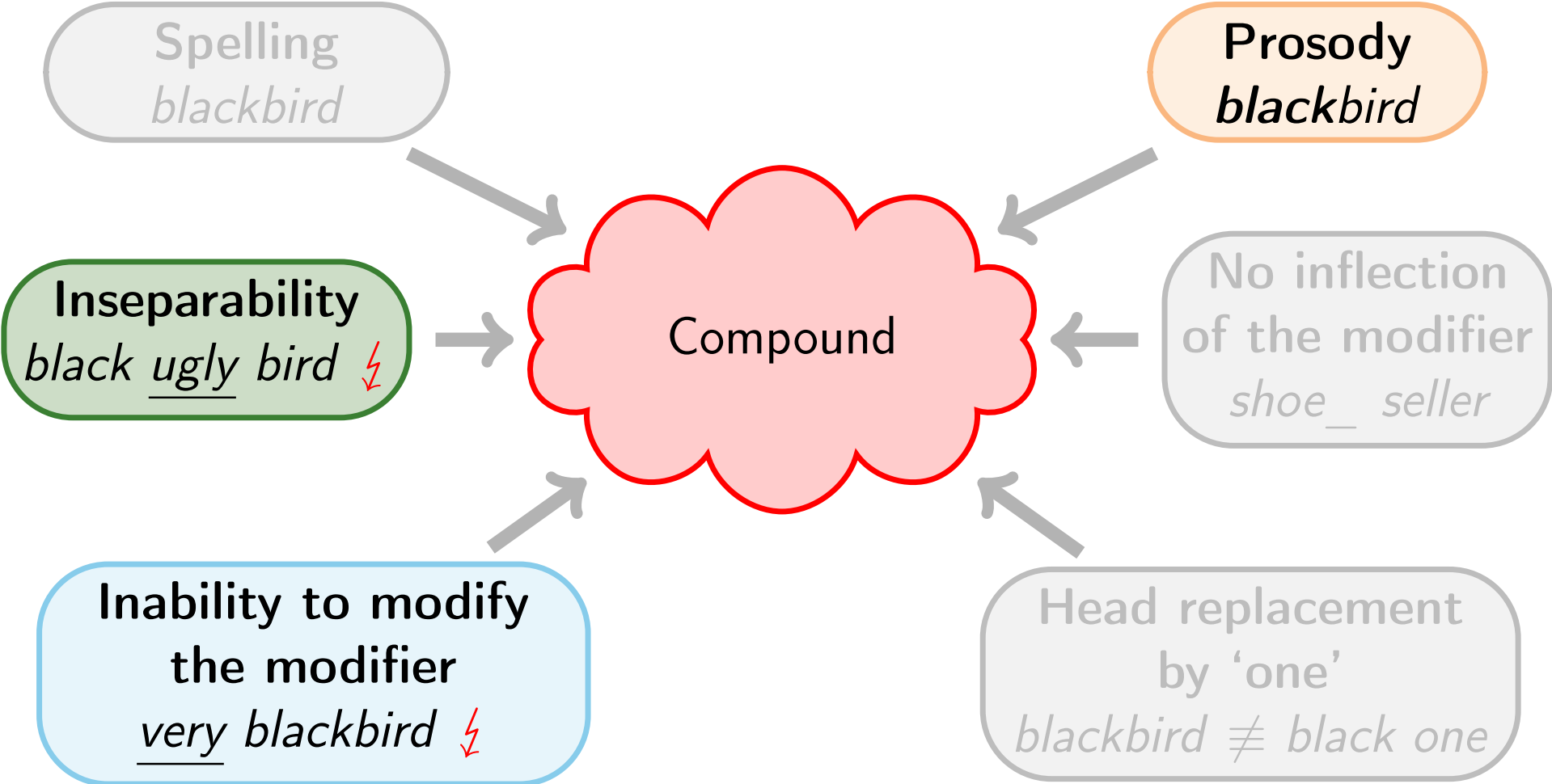
Agreement in identifying compounds

Moderate agreement

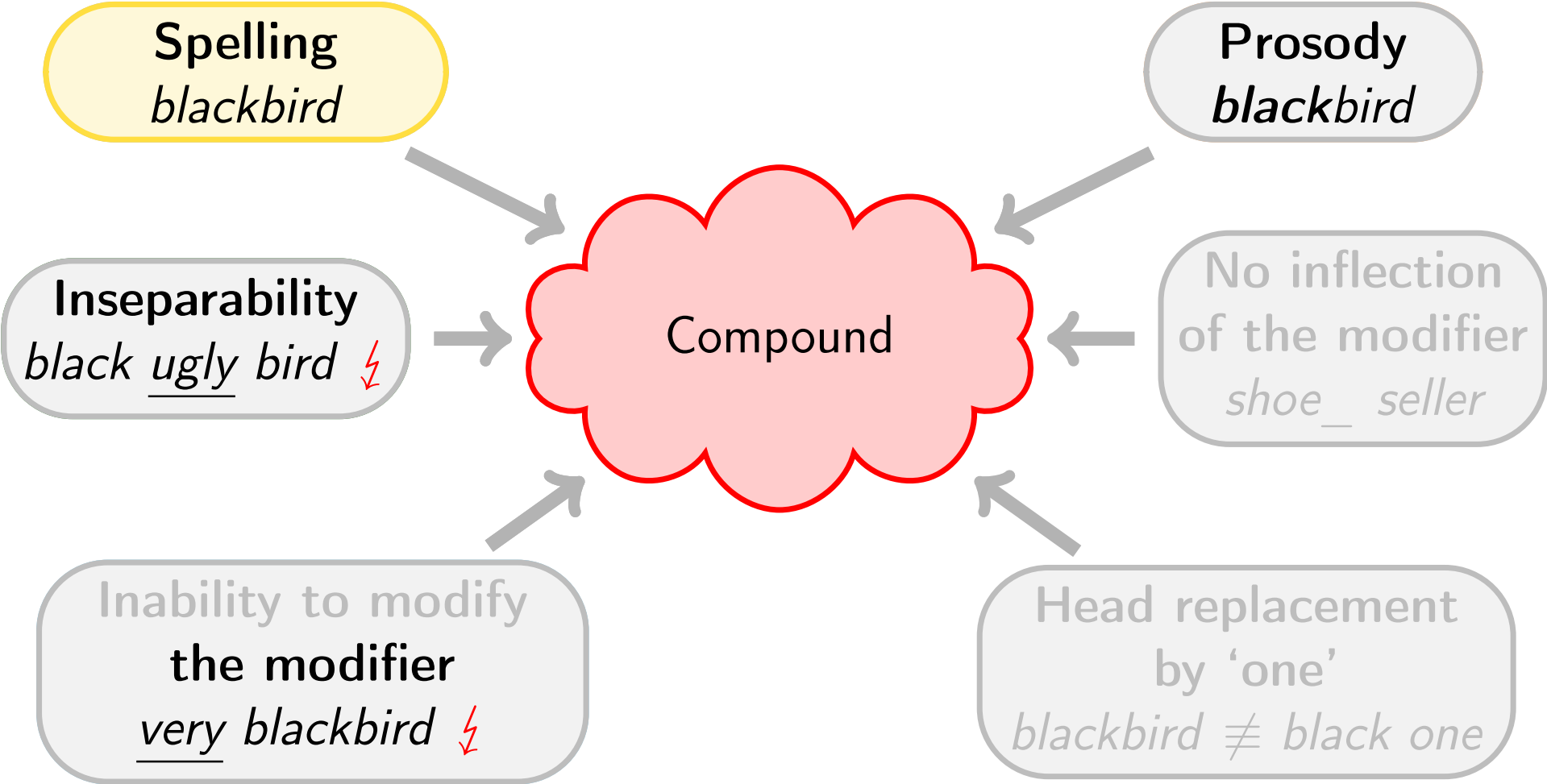
# extractions			Jacc	P	R	F ₁
Anno1	\cap	Anno2				
119	56	88	0.371	0.636	0.471	0.541

- Only moderate agreement on what compounds are
 - First annotator selected many adj-noun combinations as compounds: *economic policy, political forces*
 - Second annotator selected acronyms such as USA

Relevance of Linguistic Criteria



Relevance of Linguistic Criteria



Cross-lingual surface patterns in compounding



Let us look at 'spelling' from a cross-lingual angle

Cross-lingual surface patterns in compounding


- Noun-noun compound gold standard [Ó Séaghdha, 2007]
→ Matches with Europarl: 468 types, 11,793 tokens
- Word alignment to 9 European languages:

- PoS tagging, compound splitting

Results

- ~ 80% of **Romance** equivalents are **complex nominals**
death penalty ⇒  peine de mort
- ~ 70% of **Germanic** equivalents are **closed compounds**
death penalty ⇒  Todesstrafe

Closed compounding as naturally occurring supervision in parallel corpora [Ziering and Van der Plas, Coling 2014]

- So, if you have a compound in English, chances are (70%) you get a closed compound as an equivalent in Germanic languages
- How about the other way around?
- If a multi-word unit in English is translated as one word in Germanic languages, is this a good indicator for compoundhood?
- Difficult case: strong wind ⇒  Starkwind
- Add to that: mistakes in alignment

Closed compounding as naturally occurring supervision in parallel corpora [Ziering and Van der Plas, Coling 2014]

Implemented a system based on PoS patterns filtered by level of closed compounding in several Germanic languages

High level of closed compounding: **automobile sector**

da: "bilindustriens" de: "Automobilsektor" nl: "autosector"

sv: "bilindustrins"

Low level of closed compounding: **small dealer**


da: "små forhandlere" de: "Kleinhändler" nl: "kleine dealer"

sv: "mindre återförsäljarna"

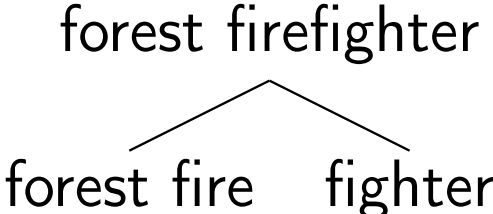
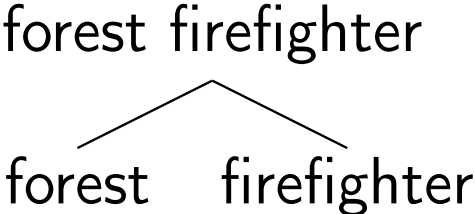
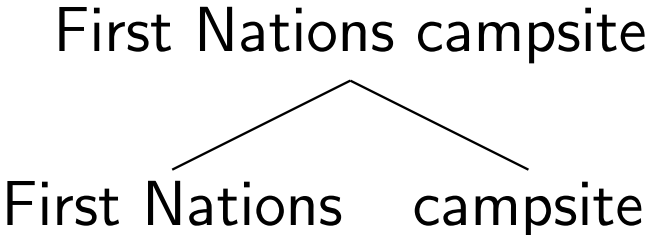
Closed compounding as naturally occurring supervision in parallel corpora [Ziering and Van der Plas, Coling 2014]

- Performance of compound extraction (EN) is 18 percentage points below the human upper bound (F-measure).
- Resource: A multilingual database of compound nouns and their translations
- Translations in up to ten languages : Danish, Dutch, French, German, Greek, Italian, Portuguese, Romanian, Spanish Swedish
- Available from:
<http://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/EuroparlNominalCompoundDatabase/index.html>
- We are currently building a larger data base (based on the DGT parallel corpus) with these languages: bg, cs, da, de, el, en, es, et, fi, fr, ga, hu, it, lt, lv, mt, nl, pl, pt, ro, sh, sk, sl, sv

Plan for this talk

- Introduction to compounds
- Overview of work done on compounds in project D11 and beyond
 - Compound discovery
 - Compound bracketing 
 - Compound splitting
 - Predicting novel compounds
- Study on the interpretation of deverbal compounds

Bracketing compounds



Bracketing compounds

Previous work:

- Unsupervised:
 - N-gram statistics or lexical patterns [Lauer, 1995; Nakov and Hearst, 2005; Barrière and Ménard, 2014]
 - web counts [Lapata and Keller, 2004]
 - semantic relations [Kim and Baldwin, 2013]
- Supervised [Vadas and Curran, 2007a,b ; Pitler et al., 2010]

Bracketing compounds

Previous work:

- Unsupervised:
 - N-gram statistics or lexical patterns [Lauer, 1995; Nakov and Hearst, 2005; Barrière and Ménard, 2014]
 - web counts [Lapata and Keller, 2004]
 - semantic relations [Kim and Baldwin, 2013]
- Supervised [Vadas and Curran, 2007a,b ; Pitler et al., 2010]

Supervised beats unsupervised.


Can we get supervision for free?


Open compounding as naturally occurring supervision for bracketing

As said, 80% of Romance equivalents of English compounds are complex nominals, and 30% of Germanic equivalents as well


What do these equivalents look like?

Open compounding as naturally occurring supervision for bracketing

health policy measures ⇒  *sundhedspolitiske
foranstaltninger*

world heritage status ⇒  *statut de patrimoine mondial*

Open compounding as naturally occurring supervision for bracketing


health policy measures ⇒  *sundhedspolitiske
foranstaltninger*

world heritage status ⇒  *statut de patrimoine mondial*

→ Behaghel's (1909) **First Law:**

Words that belong close together intellectually are also positioned close together


Open compounding as naturally occurring supervision for bracketing

health policy measures ⇒  *sundhedspolitiske
foranstaltninger*

world heritage status ⇒  *statut de patrimoine mondial*

→ Applied cross-lingually: If translations of the constituents of a compound are close together, the constituents in the source language belong close together.

Open compounding as naturally occurring supervision for bracketing

health policy measures ⇒  *sundhedspolitiske
foranstaltninger*

world heritage status ⇒  *statut de patrimoine mondial*

→ We used the distance between aligned words for the automatic bracketing of compounds [Ziering and Van der Plas, Coling 2014], [Ziering and Van der Plas, IWCS 2015],[Ziering and Van der Plas, RANLP 2015]

Aligned Word Distance Bracketing in Action

[Ziering and Van der Plas, IWCS 2015]

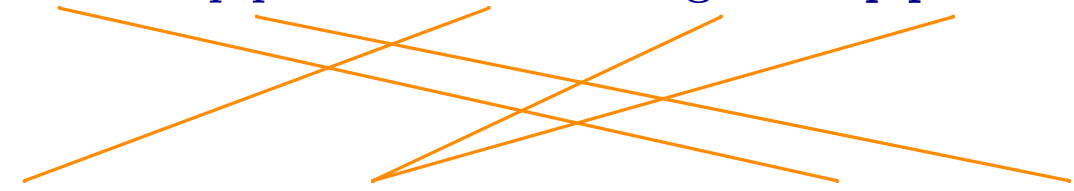
Bottom-up Parser

twinn pipe undersea gas pipeline

Word Alignment

onderzeese gaspijpleiding met dubbele pijp

Aligned Word Distance



Aligned Word Distance Bracketing in Action

[Ziering and Van der Plas, IWCS 2015]

Bottom-up Parser

twinn pipe undersea gas pipeline

Word Alignment

onderzeese gaspijpleiding met dubbele pijp

Aligned Word Distance

0

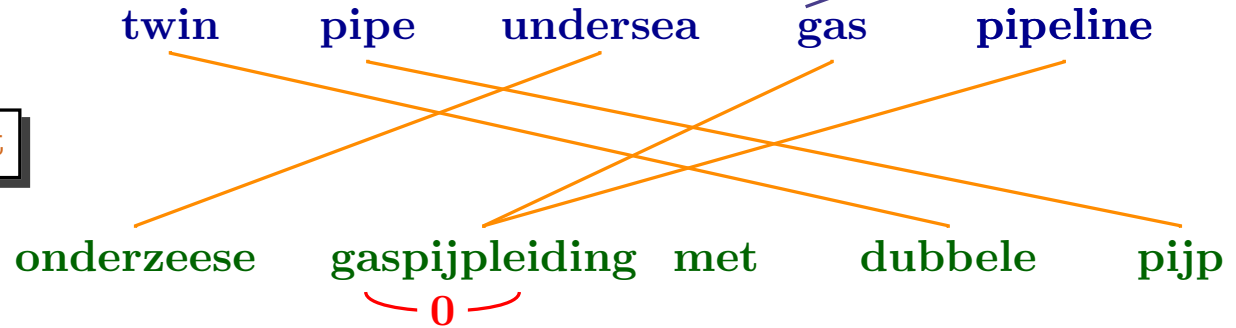
Aligned Word Distance Bracketing in Action

[Ziering and Van der Plas, IWCS 2015]

Bottom-up Parser



Word Alignment



Aligned Word Distance

0

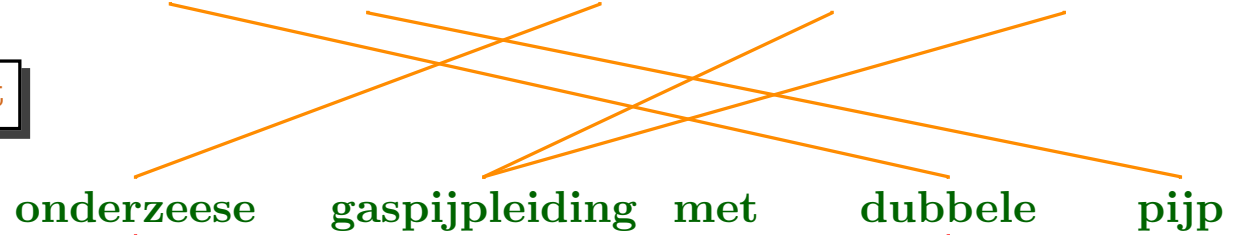
Aligned Word Distance Bracketing in Action

[Ziering and Van der Plas, IWCS 2015]

Bottom-up Parser



Word Alignment



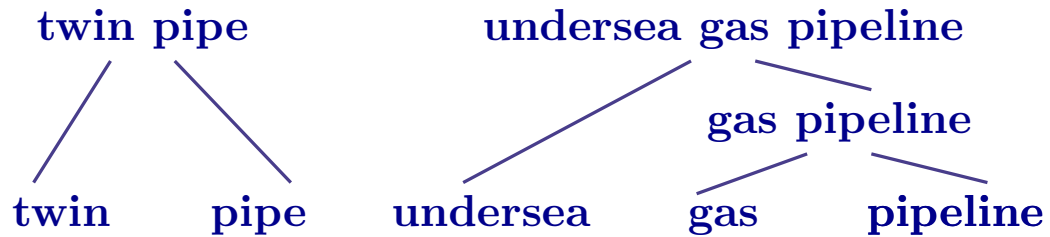
Aligned Word Distance



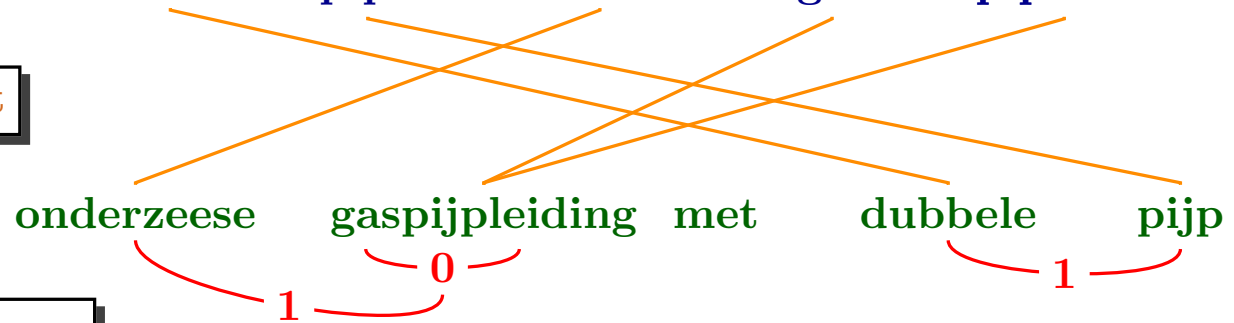
Aligned Word Distance Bracketing in Action

[Ziering and Van der Plas, IWCS 2015]

Bottom-up Parser



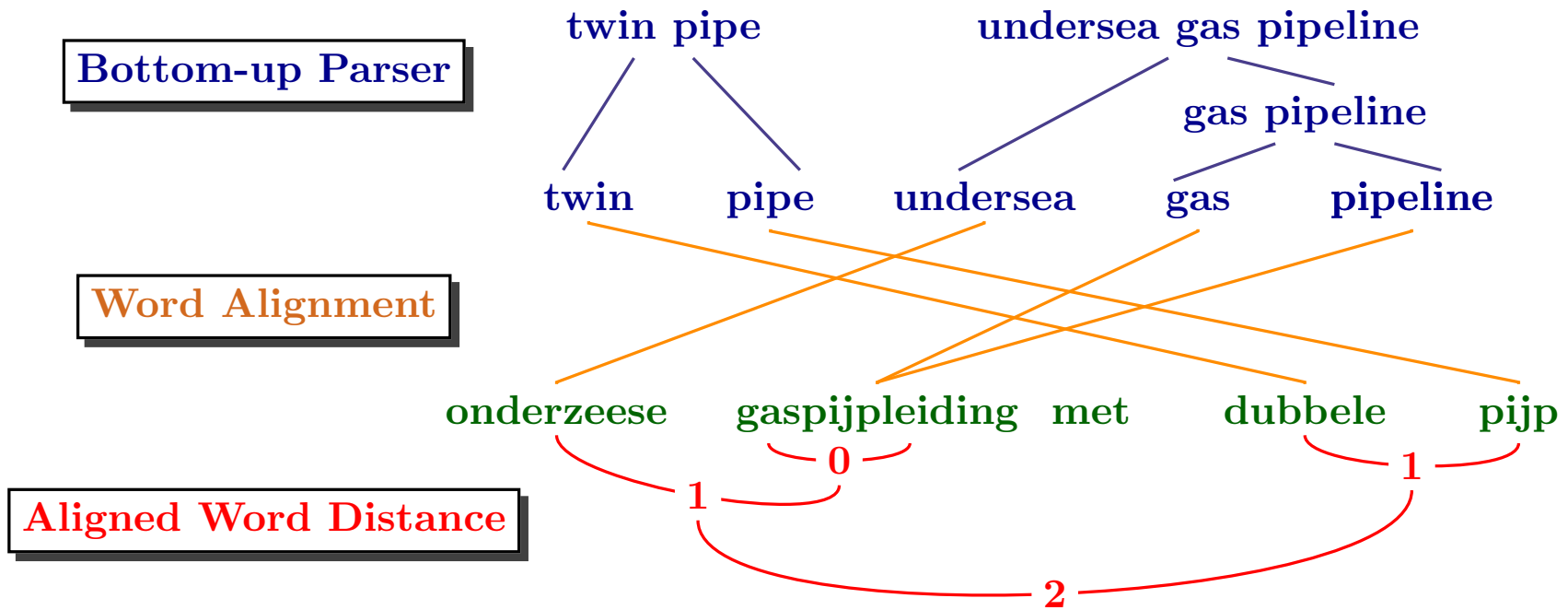
Word Alignment



Aligned Word Distance

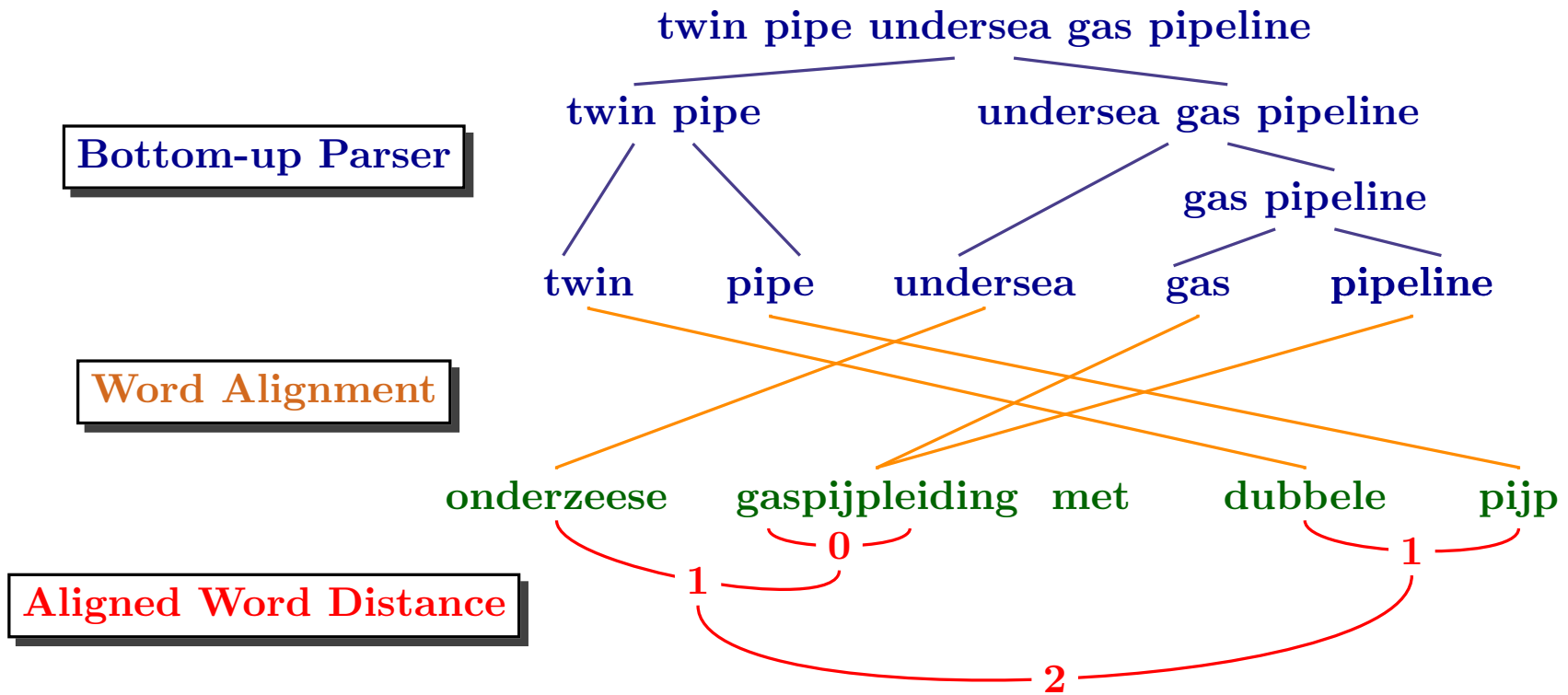
Aligned Word Distance Bracketing in Action

[Ziering and Van der Plas, IWCS 2015]



Aligned Word Distance Bracketing in Action

[Ziering and Van der Plas, IWCS 2015]




Results

System	MRP
AWD	93.7%
freq	84.6%
chance	62.5%

- AWD outperforms the baselines
- Performance comparable to human performance
- As a result, we have over 26K tokens in context annotated with high-quality bracketing structures .
- The manual annotation by Vadas and Curran (2007a) resulted in 5582 three-word NPs, that were successfully used to train supervised learners

Plan for this talk

- Introduction to compounds, and project D11
- Overview of work done on compounds in project D11 and beyond
 - Compound discovery
 - Compound bracketing
 - Compound splitting 
 - Predicting novel compounds
- Study on the interpretation of deverbal compounds

Kraftfahrzeug | *haftpflichtversicherung*
'motor car liability insurance'

Kraft | *fahrzeug*
'motor vehicle'

Haftpflicht | *versicherung*
'liability insurance'

Kraft
'force'

Fahr | *zeug*
'vehicle'

Haft | *pflicht*
'liability'

Versicherung
'insurance'

fahren
'to drive'

Zeug
'stuff'

Haft
'responsibility'

Pflicht
'obligation'

BEACH INSTRUCTIONS

WARNING No life guard on duty

- * Children under 12 Years old must be accompanied by adult
- * Diving is not allowed
- * Do not swim away from the shore
- * Swimming on your back recommended
- * Do not let sea water enter your eyes or mouth
- * Swimming is at your own risk

Strandin formationen

WARNUNG

Baden auf eigene gefahr keine

- * Keine aufsicht durch einen ademeister
- * kinder unter 12 jahren muessen in Begleitung eines erwachsenen sein
- * Tauchen ist nicht erlaubt
- * schwimmen sie nich zu weit nach draussen
- * lassen sie moeglichst kein wasser in ihre augen und mund kommen

ATTENTION

*Plage non surveillée

- * Enfants de moins de 12 ans doivent être accompagnés
- * Plonger est interdit
- * Ne vous éloignez pas du rivage
- * Il est plus prudent de nager sur le dos
- * N'avez pas d'eau de mer dans vos yeux

تعليمات السباحة في البحر الميت

- * لا يوجد منقذ
- * الأطفال دون سن 12 سنة يجب اصطحابهم من قبل شخص بالغ
- * الفطس ممنوع
- * عدم السباحة بعيد عن الشاطئ
- * السباحة على الظهر أكثر امانا
- * لا تسمح بدخول مياه البحر الى العين والفع
- * السباحة على مسؤوليتك الخاصة

BEACH INSTRUCTIONS

WARNING No life guard on duty

- * Children under 12 Years old must be accompanied by adult
- * Diving is not allowed
- * Do not swim away from the shore
- * Swimming on your back recommended
- * Do not let sea water enter your eyes or mouth
- * Swimming is at your own risk

Strandin formationen

WARNUNG

Baden auf eigene gefahr keine

- * Keine aufsicht durch einen ademeister
- * kinder unter 12 jahren muessen in Begleitung eines erwachsenen sein
- * Tauchen ist nicht erlaubt
- * schwimmen sie nich zu weit nach draussen
- * lassen sie moeglichst kein wasser in ihre augen und mund kommen

ATTENTION

*Plage non surveillée

- * Enfants de moins de 12 ans doivent être accompagnés
- * Plonger est interdit
- * Ne vous éloignez pas du rivage
- * Il est plus prudent de nager sur le dos
- * N'avez pas d'eau de mer dans vos yeux

تعليمات السباحة في البحر الميت

- * لا يوجد منقذ
- * الأطفال دون سن ١٢ سنة يجب اصطحابهم من قبل شخص بالغ
- * الفطس ممنوع
- * عدم السباحة بعيد عن الشاطئ
- * السباحة على الظهر أكثر امانا
- * لا تسمح بدخول مياه البحر الى العين والوجه
- * السباحة على مسؤوليتك الخاصة

Compound splitting

[Jagfeld et al., ACL 2017], [Ziering and Van der Plas, NAACL 2016]

Procedure

 *Hühnersuppe* 'chicken soup' → *Hühner* | *suppe*
→ *Huhn* + *Suppe*
⇒ *u* → *ü* & ⊕ *er*

Previous work

Compound splitters relying on morphological analyzers
[Fritzing and Fraser, 2010]

→ language-dependent

Statistical/corpus-based compound splitters [Koehn and Knight, 2003]

→ still contain a hand-crafted set of morphological rules
(e.g., linking morphemes)

Learning morphological transformations from parallel data
[Macherey et al., 2011]

→ relies on parallel data

Multilingual compound splitting using word inflection

- Linking operations strongly conform with word inflection (though no one-to-one correspondence)
verbal conjugation, pluralization, case marking, ...
- Assumption : we can learn linking operations from word inflection (lemmatised corpus)
⇒ Representation: Morphological Operation Pattern (MOP)

Morphological Operation Patterns (MOPs)

- We represent morphological operations as a sequence of character replacements: **Morphological Operation Patterns**

Language	MOP	Examples
German	u/ü:\$/er\$	<Huhn, Hühner> 'hen', <Buch, Bücher> 'book'
	um\$/en\$	<Studium, Studien> 'study', <Medium, Medien> 'medium'
Dutch	\$/en\$	<arts, artsen> 'doctor', <band, banden> 'tyre')
Afrikaans	\$/se\$	<proses, prosesse> 'process'

Learning MOPs from word inflection




- ① **Input:** Lemmatized tokens (i.e., <lemma, word-form> pairs)
- ② **Process:** Backtrace of Levenshtein edit distance (ED)
- ③ **Output:** Grouped ED replacement operations
→ $\text{freq}(\text{MOP}) \equiv \text{type-based freq}(\text{lemmatization})$

Experimental setup

Monolingual corpora

 / 	Wikipedia	665M / 114M	tokens
	Taalkommissie corpus	57M	tokens

Compound gold standard

	German	[Henrich and Hinrichs, 2011]	51K
	Dutch	[Verhoeven et al., 2014]	22K
	Afrikaans	[Verhoeven et al., 2014]	17K

Evaluation measures

- Split point accuracy (**SPA**cc): split points
- Normalization accuracy (**NormA**cc): split points and lemmas

Results compound splitting

German:

Our resource-lean approach (needs lemmatised corpus only) outperforms previous work in split point accuracy

It is competitive with previous work (lexicon and rule-based) w.r.t. normalisation accuracy

Outperforms previous work for Dutch

Does not outperform previous work on Afrikaans due to small corpus

Split re-ranking using semantic support

[Ziering, Mueller, and Van der Plas, MWE@ACL 2016]


- Most splitters follow generate-and-rank procedure (ranking based on constituent frequency)
 - neglect semantic plausibility
 - Ei* | *dotter* ‘egg yolk’ vs. *Eid* | *otter* ‘oath otter’
- Distributional similarity between constituents and compound
 - Previously, used to determine compound compositionality [Reddy et al., (2011); Schulte im Walde et al., (2013)]
 - Can be used directly to measure plausibility of candidate splits
 - Eidotter* ‘egg yolk’ is more semantically similar to *Dotter* ‘yolk’ than to *otter* ‘otter’

Evaluation of re-ranking

Accuracy	SPA cc	Norm Acc
[ZvdP_2016]	97.5%	87.4%
Re-ranking	98.0% †	89.0% †

- ⇒ Re-ranking outperforms the initial ranking [Ziering and Van der Plas, NAACL 2016] significantly
- ⇒ Has been applied successfully to re-rank output of several splitters

Plan for his talk

- Introduction to compounds, and project D11
- Overview of work done on compounds in project D11 and beyond
 - Compound discovery
 - Compound bracketing
 - Compound splitting
 - Predicting novel compounds 
- Study on the interpretation of deverbal compounds

Creativity a much needed skill

- “We need new ideas to solve our country’s pressing problems”
“We need workers who can ‘think outside the box’ - especially in science and technology - to be competitive in today’s global economy” [Moran, 2010]
- LinkedIn Learning determined the skills companies need most based on their internal data → Creativity on top of list
[<https://learning.linkedin.com>]
- Prime skill of human beings in future post-information society
[Corazza, 2016]

Creativity in NLP

- Creativity has not been on top of the agenda for NLP
- Mostly concerned with analysis
- Interfaces with large amounts of information, still focused on 'finding what you are looking for' → filter bubbles [Pariser, 2011]
- **General idea:** Twist the patterns we find slightly to create novelty

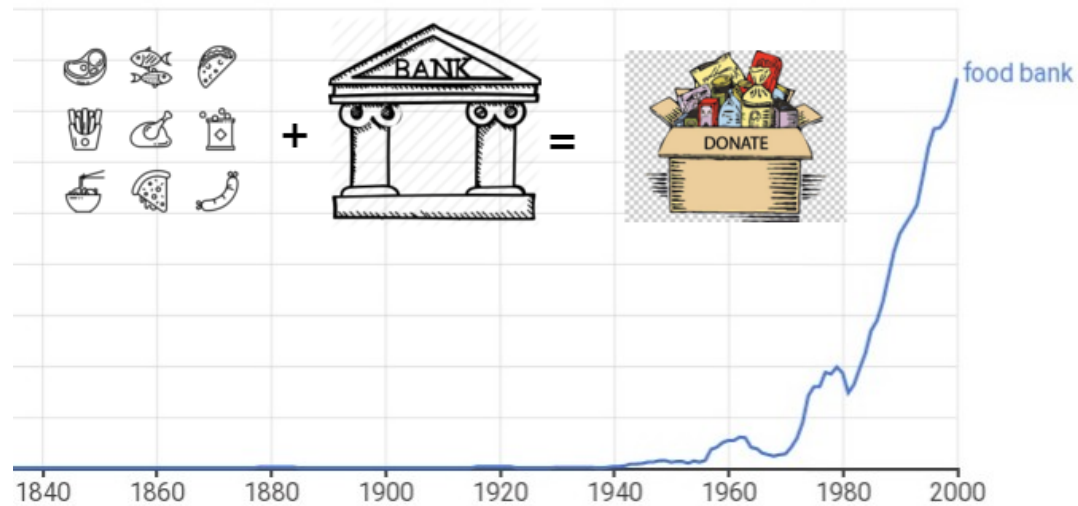
Compounds as vehicles for creative thought

- High level of underspecification
- Flexibility
- Combinatorial power

Novel compound generation

Ngram Viewer

food bank



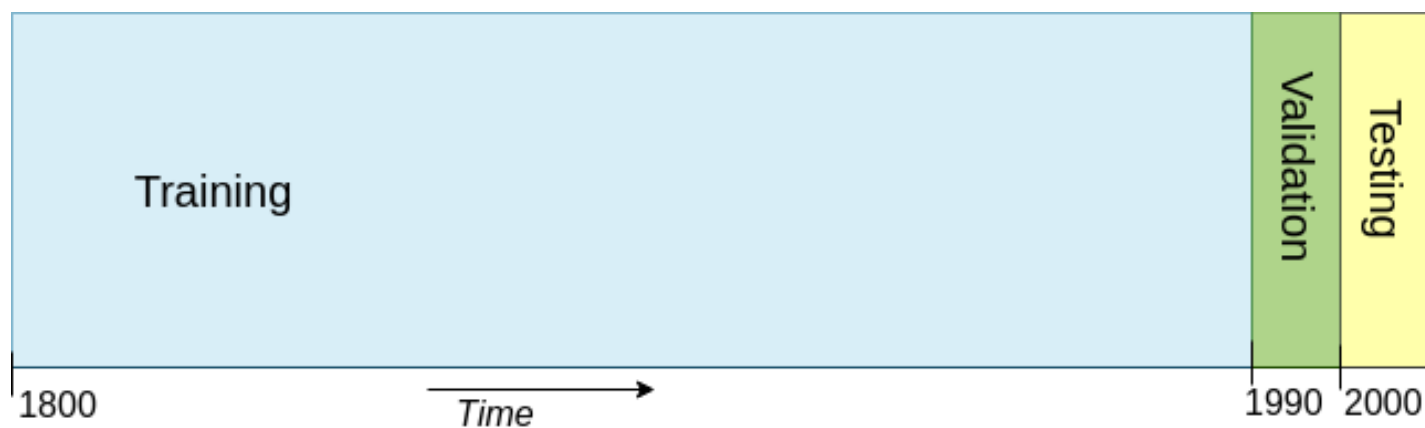
Novel compound generation: How?

- Using distributed representations for the constituents and modelling their combination
- For example, *glass-bottom boat* is found in early corpora, but not *glass canoe*
- Task: Infer that a *glass canoe* is a plausible concept, given the evidence of seeing *glass-bottom boat*, and the similarity between the constituents



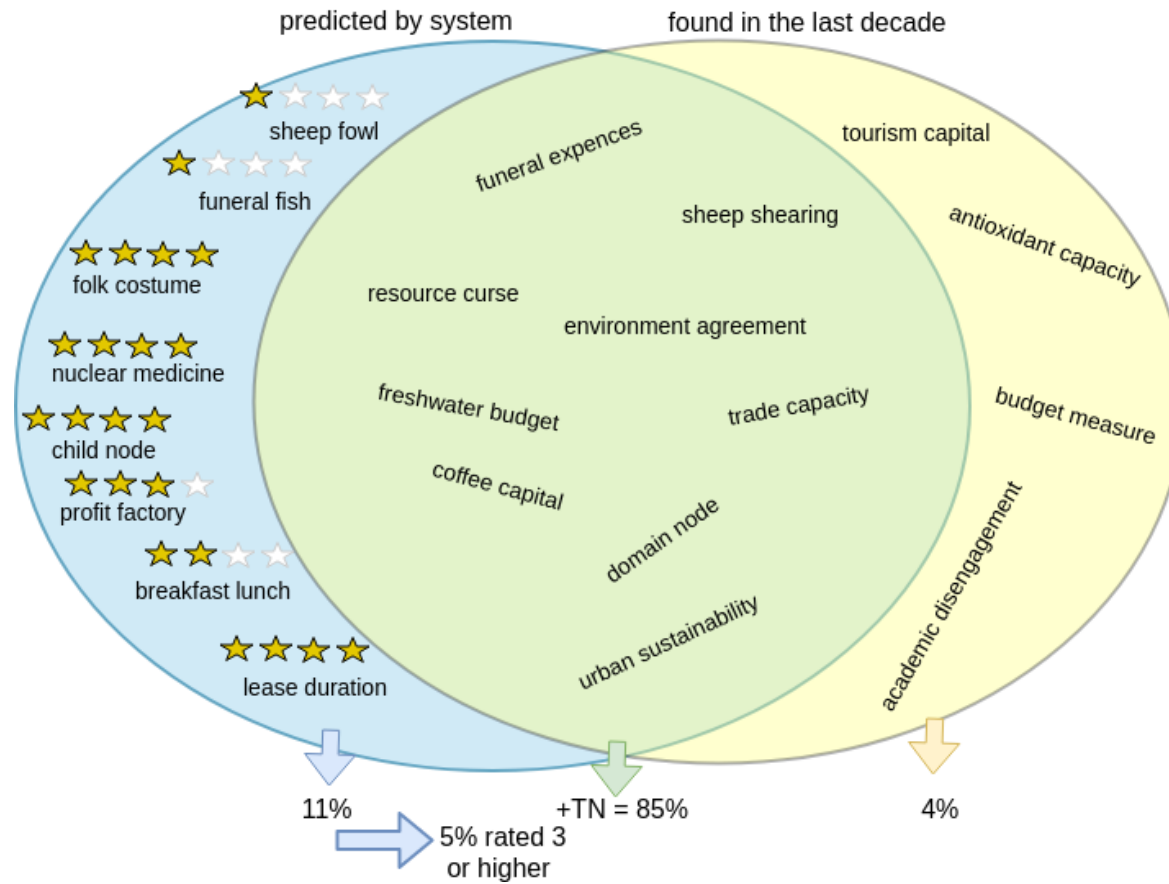
Evaluation in time-stamped corpus

- Train on early parts of time-stamped corpus
- Predict novel compounds in unseen part of corpus.



- Novel compound generation → novel compound prediction

Results



Plan for his talk

- Introduction to compounds, and project D11
- Overview of work done on compounds in project D11 and beyond
 - Compound discovery
 - Compound bracketing
 - Compound splitting
 - Predicting novel compounds
- Study on the interpretation of deverbal compounds



Deverbal nouns, between Argument Structure Nominals (ASNs) and Result Nominals (RNs)

- Grimshaw (1990): Deverbal Ns are ambiguous between compositional verb-like ASN-readings and more lexicalized RN-readings:
 1. a. The **examination**/exam was on the table. (RN)
b. The **examination of the patients** took a long time/*was on the table. (ASN)
- ASNs vs. RNs (presence vs. absence of event/argument structure):

(adapted from Alexiadou & Grimshaw 2008: 3, citing Grimshaw 1990; see Appendix-1 for details)

The Linguistic Debate on Deverbal Compounds

- **Grimshaw (1990): DCs = ASNs**: only the lowest argument (Theme/OBJ) is possible (Agent<Goal<**Theme**):
gift-giving to children - ***child**-giving of gifts (*to give gifts to children*)
book-reading by students - ***student**-reading of books (*Students read books*)
- Root compounds (RCs) may realize agents/subjects, but DCs cannot:
bee sting; **dog** bite (vs. ***bee**-stinging, ***dog**-biting)
- **Borer (2013): DCs ≠ ASNs**; DCs have no arguments or event structure:
 - a. the house demolition (***by the army**) (*in two hours) (DC)
 - b. the demolition of the house **by the army** in two hours (ASN)aspectual modifiers not available with DCs
- As in root compounds, non-heads are context-dependent: agent is OK:
teacher recommendation; **court** investigation; **government** decision

Hypothesis and Predictions

- **Hypothesis:** If a deverbal noun is used more like a compositional ASN or a lexicalized RN, this will be preserved in DCs
- **Grimshaw's prediction:** we should find a correlation between the ASN-properties of the head noun and the OBJ-reading of a DC:
 snow_{OBJ}/waste_{OBJ} removal vs. health_{OBJ}/flood_{OTHER} insurance
 drug_{OBJ}/child_{OBJ} trafficking border_{OBJ}/fire_{OTHER}/police_{SUBJ} protection
- **Borer's prediction:** no such correlation; apparent ASN-properties of a head noun should be compatible with both OBJ and non-OBJ DCs

Our Study

Combine **corpus data automatically extracted** from the Gigaword Corpus (Napoles et al. 2012) and **annotations by three native speakers**

- Use **machine learning techniques** to tell:
 - If ASN-properties of DC-heads are **predictive** in disambiguating OBJ/non-OBJ readings in DCs (cf. Grimshaw vs. Borer)
 - which properties of DC-heads are most relevant
- Our results provide support for Grimshaw's analysis and our hypothesis that **DCs headed by ASN-like nouns receive OBJ readings**
- **Human annotation feature** is the best => we need even larger corpora to capture human intuitions

Outline

- 1) **Our Methodology: Data Extraction and Annotation**
- 2) Case study 1: Corpus-based Morphosyntactic Features
- 3) Case study 2: Human Annotation of Result-vs-Process in DC
- 4) Interpretation of Results
- 5) Conclusion and Future Plans

Methodology

- 1) We created a frequency-balanced list of 25 deverbal nouns for each of the suffixes **-ing**, **-ion**, **-al**, **-ance**, **-ment**
- 2) We then extracted the 25 most frequent compounds that they appeared as heads of => a total of 3111 compounds
- 3) Annotate interpretation of DCs: OBJ, SUBJ, OTHER, ERROR

3) Annotation of interpretation

- Three trained native speaker annotators (US & Canadian English)
- **Annotate** the relation between head and non-head:
 - SUBJ: ext. Arg. (**police** questioning, **designer** creation)
 - OBJ: int. Arg. (**book** writing, **crop** destruction, **hair** removal)
 - OTHER (**contract** killing, **safety** instruction)
 - ERROR (PoS tag errors or uninterpretable compounds: e.g. *face_V abandonment, fond_A remembrance*)
- Allow for ambiguity & preference order: SUBJ – OBJ, SUBJ > OBJ
- Post-processing => binary classification OBJ-NOBJ
- Inter-annotator agreement: **81.5%** (2 annotators), **72.8%** (3 annotators)
- **Result:** **2067 DCs: 1377 OBJ – 690 NOBJ**

Methodology

- 1) We created a frequency-balanced list of 25 deverbal nouns for each of the suffixes **-ing, -ion, -al, -ance, -ment**
- 2) We then extracted the 25 most frequent compounds that they appeared as heads of => a total of 3111 compounds
- 3) Annotate interpretation of DCs: OBJ, SUBJ, OTHER, ERROR
- 4) **Determine ASN-properties of heads by automatically extracting contexts from Gigaword**

4) Morphosyntactic Features

- Features 1-4 ~ Grimshaw's ASN-properties of the head noun when it appears on its own, i.e., *outside* a DC:

Feature label	Description and illustration
1. <i>of_outside_DC</i>	Percentage of the head's occurrences as singular outside DCs, which realize a syntactic relation with an <i>of</i> -phrase: e.g., <i>assignment of problems</i>
2. <i>by_outside_DC</i>	Percentage of the head's occurrences in the singular outside DCs, which realize a syntactic relation with a <i>by</i> -phrase: e.g., <i>assignment (of problems) by teachers</i>
3. <i>sum_adjectives</i>	Percentage of the head's occurrences in a modifier relation with one of the adjectives <i>frequent</i> , <i>constant</i> , <i>intentional</i> , <i>deliberate</i> , or <i>careful</i> .
4. <i>sg_outside_DC</i>	Percentage of the head's occurrences as singular outside DCs.

4) Further Morphosyntactic Features

- Features 5-6: by-phrases and singular form when used **inside** DCs
- Features 7-8: our additions

Feature label	Description and illustration
5. <i>by_inside_DC</i>	Percentage of the head's occurrences in the singular inside DCs, which realize a syntactic relation with a <i>by</i> -phrase: e.g., <i>task assignment by teachers</i>
6. <i>sg_inside_DC</i>	Percentage of the head's occurrences as singular inside DCs.
7. <i>suffix</i>	Suffix of the head noun: AL (rental), ANCE (insurance), ING (killing), ION (destruction), MENT (treatment)
8. <i>head_in_DC</i>	Percentage of the head's occurrences within a compound out of its total occurrences in the corpus.

Automatic analysis of Gigaword

- More than 4 billion words
- Automatic tokenization, lemmatization, PoS tagging, and syntactic parsing
- Performance of parser: F1 score of 91.4% on standard test set (WSJ)

Reliability of extracted features:

2000 sentences manually annotated

Precision 91.0% and recall 90.1% for of-phrases

Precision 85.0% and recall 73.8% for by-phrases

Outline

- 1) Our Methodology: Data Extraction and Annotation
- 2) **Case study 1: Corpus-based Morphosyntactic Features**
- 3) Case study 2: Human Annotation of Result-vs-Process in DC
- 4) Interpretation of Results
- 5) Conclusion and Future Plans

Logistic Regression for Data Analysis

- **Questions** for our experiments:
 - 1) Can the head's properties predict the OBJ/NOBJ reading of DCs?
 - 2) Which properties are the strongest predictors?
- **Predictions:**
 - Grimshaw: **ASN-properties** should have good predictive power identifying **OBJ-DCs**
 - Borer: DCs have no event structure and cannot involve ASN
- **Procedure:**
 - 8 (9) independent variables (the morphosyntactic properties)
 - 1 categorical dependent variable (OBJ-NOBJ)
 - Split up data so that no head in test data is seen in training
 - Balanced data set for two classes (by removing OBJ instances)
 - **Data used:** 1248 training, 132 test compounds

Results for Individual Features

- How well does each feature do on its own?

Random baseline: 50%

Feature	Accuracy
<i>suffix</i>	72.0%
<i>sg_outside_DC</i>	68.9%
<i>sg_inside_DC</i>	68.9%
<i>head_in_DC</i>	66.7%
<i>sum_adjectives</i>	61.4%
<i>of_outside_DC</i>	59.8%
<i>by_outside_DC</i>	56.0%
<i>by_inside_DC</i>	54.5%

- All features have predictive power above random; other features are better than *of*-phrases, which is most important for Grimshaw (1990)

Individual Features: OBJ/NOBJ Prediction

- Which DC readings do the features correlate with?

Feature	Accuracy	Direction of Prediction
<i>suffix</i>	72.0%	<i>ment, ance, al</i> : OBJ ; <i>ion, ing</i> : NOBJ
<i>sg_outside_DC</i>	68.9%	OBJ
<i>sg_inside_DC</i>	68.9%	NOBJ
<i>head_in_DC</i>	66.7%	OBJ
<i>sum_adjectives</i>	61.4%	NOBJ
<i>of_outside_DC</i>	59.8%	OBJ
<i>by_outside_DC</i>	56.0%	NOBJ
<i>by_inside_DC</i>	54.5%	OBJ

- sg_outside_DC*, *head_in_DC*, *of_outside_DC* and *by_inside_DC* correlate with OBJ-readings
- We will return to this aspect later to compare the features in more models

The Suffix

- The theoretical literature does not say much about the different suffixes
- Both Grimshaw (1990) and Borer (2013) predict that *-ing* should pattern with OBJ-DCs, but this is not confirmed (see later discussion)
- The direction of prediction for the suffixes is not stable across models
- Note that data is very sparse for this feature (the only categorical feature with 5 different classes)
- The suffix does not say anything about ASN-hood of DCs or event OBJ vs. NOBJ reading
 - It acts as a good predictor

Sg_outside_DC: Examples

Head noun	<i>Sg_outside_DC</i>	OBJ-reading
<i>dominance</i>	100%	61.5%
<i>mongering</i>	100%	100%
<i>maintenance</i>	99.99%	100%
<i>counseling</i>	99.97%	10%
<i>trafficking</i>	99.97%	100%
<i>ignorance</i>	99.96%	0%
<i>consumption</i>	99.96%	87.5%
<i>assurance</i>	32.71%	18.18%
<i>instruction</i>	28.08%	11.7%
<i>reinforcement</i>	25.69%	88.88%
<i>remittance</i>	12.37%	94.11%

Heads with (in)frequent singular form outside compounds; outliers in bold

Head_in_DC: Examples

Head noun	<i>Head_in_DC</i>	OBJ-reading
<i>laundering</i>	94.80%	95.45%
<i>mongering</i>	91.77%	100%
<i>growing</i>	68.68%	95.23%
<i>trafficking</i>	61.99%	100%
<i>enforcement</i>	53.68%	66.66%
<i>chasing</i>	44.74%	90%
<i>insurance</i>	43.73%	46.15%
<i>rental</i>	42.95%	87.5%
<i>acquittal</i>	1.80%	12.5%
<i>ignorance</i>	0.85%	0%
<i>refusal</i>	0.77%	43.75%
<i>anticipation</i>	0.70%	37.5%
<i>defiance</i>	0.64%	35.29%

Heads with most/least frequent occurrence in compounds; outliers in bold

Of_outside_DC: Examples

Head noun	Of-phrases	OBJ-reading
<i>creation</i>	80.51%	72.72%
<i>avoidance</i>	70.40%	100%
<i>obstruction</i>	65.25%	90.47%
<i>removal</i>	63.53%	92%
<i>breaking</i>	58.83%	94.11%
<i>abandonment</i>	55.90%	90%
<i>assassination</i>	52.27%	11.76%
<i>preservation</i>	52.14%	100%
<i>education</i>	1.81%	30%
<i>proposal</i>	1.08%	76.19%
<i>counseling</i>	0.53%	10%
<i>insurance</i>	0.42%	46.15%
<i>mongering</i>	0%	100%

Heads with (in)frequent *of*-phrases outside compounds; outliers in bold

Results for Ablation Tests: All Corpus Features

- A model with all features together; cf. models that lack one feature

Features	Accuracy
All corpus-based features	72.0%
All corpus-based features, except <i>sg_outside_DC</i>	72.0%
All corpus-based features, except <i>sum_adjectives</i>	72.0%
All corpus-based features, except <i>sg_inside_DC</i>	72.0%
All corpus-based features, except <i>by_inside_DC</i>	72.0%
All corpus-based features, except <i>head_in_DC</i>	68.2%
All corpus-based features, except <i>suffix</i>	66.7%†
All corpus-based features, except <i>by_outside_DC</i>	59.1%†
All corpus-based features, except <i>of_outside_DC</i>	54.5%†
<i>of_outside_DC</i> , <i>by_outside_DC</i> , and <i>suffix</i> combined	72.7%

†The dagger indicates a significantly lower value than for all features

- *Of*-phrases represent the strongest feature here, followed by *by*-phrases outside DCs and the suffix

Ablation with Grimshaw's Features Only

Features	Accuracy	<i>cf. Individual features</i>	
All features	63.6%	<i>suffix</i>	72.0%
All features, except <i>sg_outside_DC</i>	63.6%	<i>sg_outside_DC</i>	68.9%
All features, except <i>sum_adjectives</i>	63.6%	<i>sg_inside_DC</i>	68.9%
All features, except <i>sg_inside DC</i>	60.6%	<i>head_in_DC</i>	66.7%
All features, except <i>by_inside_DC</i>	59.8%	<i>sum_adjectives</i>	61.4%
All features, except <i>of_outside_DC</i>	58.3%	<i>of_outside_DC</i>	59.8%
All features, except <i>by_outside_DC</i>	56.8%	<i>by_outside_DC</i>	56.0%
		<i>by_inside_DC</i>	54.5%

- The model is weaker than with *suffix* and *head_in_DC* (72%) and weaker than some individual features on their own (see above)
- The features overlap in contribution

Interim Summary on Corpus-based Features

1) Are Grimshaw's ASN-features predictive?

- Each of them scores above the random baseline of 50%;
- Yet, they overlap
- The suffix is the best corpus feature (as good as the full model)

2) Which features are strongest?

<i>suffix</i>	72.0%	All corpus-based features	72.0%
<i>sg_outside_DC</i>	68.9%	All corpus-based features, except <i>sg_outside_DC</i>	72.0%
<i>sg_inside_DC</i>	68.9%	All corpus-based features, except <i>sum_adjectives</i>	72.0%
<i>head_in_DC</i>	66.7%	All corpus-based features, except <i>sg_inside_DC</i>	72.0%
<i>sum_adjectives</i>	61.4%	All corpus-based features, except <i>by_inside_DC</i>	72.0%
<i>of_outside_DC</i>	59.8%	All corpus-based features, except <i>head_in_DC</i>	68.2%
<i>by_outside_DC</i>	56.0%	All corpus-based features, except <i>suffix</i>	66.7%†
<i>by_inside_DC</i>	54.5%	All corpus-based features, except <i>by_outside_DC</i>	59.1%†
		All corpus-based features, except <i>of_outside_DC</i>	54.5%†

- **Of- and by-phrases** have high contribution combined with others
- Grimshaw (1990) used these properties in combination

Outline

- 1) Our Methodology: Data Extraction and Annotation
- 2) Case study 1: Corpus-based Morphosyntactic Features
- 3) Case study 2: Human Annotation of Result-vs-Process in DC**
- 4) Interpretation of Results
- 5) Conclusion and Future Plans

Capturing Grimshaw's ASN-hood

- Grimshaw (1990): ASNs express processes/events like verbs
- Her tests aim to **identify the process (vs. result) interpretation** of a derived nominal, which should associate with **OBJ in DCs**
- Possible problems with corpus data:
 - Crucial data is sparse (*sum_adjectives* at most 1%, *by_phrases*: < 6.2%)
 - Combining features even leads to more sparsity
 - Automatic processing is not perfect
- Find a better way to get at Grimshaw's **process-vs-result** contrast in derived nominals: **human judgments!**
- If she is right, high *processhood* in DCs should indicate OBJ
- **Human annotation of process-vs-result** reading in DCs: on a scale from 1 (result entity) to 5 (full process)

Process-vs-Result in DCs

- New annotation task: native speaker intuitions with respect to whether a DC's head is more like an ASN or a RN
- **Instructions:** ASN vs RN *assignment*:
- “*The teacher's **assignment** of tasks* expresses a process in which the teacher assigns tasks. However, in *this long **assignment** took several hours to complete*, the noun *assignment* is interpreted as a result of the process of assigning something – namely, the task itself.”
 - *task assignment* closer to 5 than *Math assignment*; *truck driving* ~ 5
 - *apartment building*: close to 5 if < 'to build apartments';
 - close to 1 if < 'a building with apartments'
- Inter-annotator agreement: **0.46** (Spearman: moderate)
- Feature value = average between the three annotators: [1-5]*20 => 20 to 100 values

Individual Features Models

- P-R significantly improves the model (cf. 72%) and is the best feature

Features	Accuracy
All features	78.8% ←
<i>process-vs-result</i>	76.5% ←
<i>suffix</i>	72.0%†
<i>sg_outside_DC</i>	68.9%†
<i>sg_inside_DC</i>	68.9%†
<i>head_in_DC</i>	66.7%†
<i>sum_adjectives</i>	61.4%†
<i>of_outside_DC</i>	59.8%†
<i>by_outside_DC</i>	56.0%†
<i>by_inside_DC</i>	54.5%†
<i>process-vs-result</i> and <i>suffix</i> combined	78.0%
Random baseline	50.0%

†The dagger indicates a significantly lower value by comparison to all features

Contrasting P-R Values on Similar Heads

- Usually, the OBJ-reading has a higher P-R, but we also find OBJ-DCs with low P-R values, and NOBJ-DCs with high P-R values:

High $P-R > 60\%$			Low $P-R < 60\%$		
DC	P-R	Reading	DC	P-R	Reading
⇒ home building	100%	OBJ	police building	20.0%	NOBJ
book reading	100%	OBJ	temperature reading	40.0%	OBJ
ship breaking	93.3%	OBJ	record breaking	40.0%	OBJ
⇒ science teaching	93.3%	OBJ	church teaching	46.7%	NOBJ
career counseling	93.3%	NOBJ	telephone counseling	53.3%	NOBJ
⇒ slum clearance	80.0%	OBJ	safety clearance	20.0%	NOBJ
⇒ body movement	80.0%	OBJ	student movement	33.3%	NOBJ
nicotine withdrawal	80.0%	NOBJ	summer withdrawal	33.3%	NOBJ
refuse disposal	80.0%	OBJ	garbage disposal	46.7%	OBJ
temperature tolerance	73.3%	OBJ	alcohol tolerance	20.0%	OBJ
⇒ cancer treatment	73.3%	OBJ	spa treatment	46.7%	NOBJ

Ablation Experiments

- When *P-R* is present, the **removal of *of_outside_DC*** worsens the model **just as much as the removal of *P-R*** (72% vs. 78.8%)

Features	Accuracy
All features	78.8%
All features, except <i>sg_inside_DC</i>	80.3%
All features, except <i>head_in_DC</i>	79.5%
All features, except <i>sg_outside_DC</i>	78.8%
All features, except <i>by_inside_DC</i>	78.8%
All features, except <i>sum_adjectives</i>	78.8%
All features, except <i>by_outside_DC</i>	75.0%
All features, except <i>suffix</i>	73.5%†
All features, except <i>of_outside_DC</i>	72.0%† ←
All features, except <i>P-R</i>	72.0%† ←
<i>P-R, of_outside_DC, suffix, by_outside_DC</i> combined	80.3%

†The dagger indicates a significantly lower value by comparison to all features

Outline

- 1) Our Methodology: Data Extraction and Annotation
- 2) Case study 1: Corpus-based Morphosyntactic Features
- 3) Case study 2: Human Annotation of Result-vs-Process in DC
- 4) Interpretation of Results**
- 5) Conclusion and Future Plans

Summary of Results

Ablation Experiments

All corpus-based features	72.0%
All corpus-based features, except <i>sg_outside_DC</i>	72.0%
All corpus-based features, except <i>sum_adjectives</i>	72.0%
All corpus-based features, except <i>sg_inside_DC</i>	72.0%
All corpus-based features, except <i>by_inside_DC</i>	72.0%
All corpus-based features, except <i>head_in_DC</i>	68.2%
All corpus-based features, except <i>suffix</i>	66.7%†
All corpus-based features, except <i>by_outside_DC</i>	59.1%†
All corpus-based features, except <i>of_outside_DC</i>	54.5%†
<i>of_outside_DC</i> , <i>by_outside_DC</i> , and <i>suffix</i> combined	72.7%

All features	78.8%
All features, except <i>sg_inside_DC</i>	80.3%
All features, except <i>head_in_DC</i>	79.5%
All features, except <i>sg_outside_DC</i>	78.8%
All features, except <i>by_inside_DC</i>	78.8%
All features, except <i>sum_adjectives</i>	78.8%
All features, except <i>by_outside_DC</i>	75.0%
All features, except <i>suffix</i>	73.5%†
All features, except <i>of_outside_DC</i>	72.0%†
All features, except <i>P-R</i>	72.0%†

Individual Features

All features	78.8%
<i>process-vs-result</i>	76.5%
<i>suffix</i>	72.0%†
<i>sg_outside_DC</i>	58.9%†
<i>sg_inside_DC</i>	58.9%†
<i>head_in_DC</i>	56.7%†
<i>sum_adjectives</i>	61.4%†
<i>of_outside_DC</i>	59.8%†
<i>by_outside_DC</i>	56.0%†
<i>by_inside_DC</i>	54.5%†



Importance of features

1. ***P-R*** is the **best feature** – based on **human annotation**: it supports our hypothesis about the importance of ASN-hood in interpreting DCs and Grimshaw's intuition about the ASN vs. RN contrast
 2. ***Suffix*** is the **next best feature**
 3. ***Of_outside_DC***: **most important** for Grimshaw; **relies a lot on other features** (cf. ablation tests vs. individual contribution), but P-R cannot compensate for its absence
 4. ***By_outside_DC***: when absent, P-R can compensate for it
 5. ***Singular*** marking outside/inside: good individually but, when absent, other features can compensate (the model stays just as good)
- **Do we find a correlation between ASN-hood and OBJ-DCs?**

Direction of Prediction for Features

• **Feature** **Individual** **Ablation** **Corpus-Ablation**

P-R OBJ OBJ N/A

suffix=ment OBJ OBJ OBJ

suffix=ance OBJ NOBJ NOBJ

suffix=ion NOBJ OBJ OBJ

suffix=al OBJ NOBJ OBJ

suffix=ing NOBJ OBJ NOBJ

sg_inside_DC NOBJ OBJ NOBJ

by inside DC OBJ NOBJ NOBJ

sg_outside_DC OBJ OBJ OBJ

head in DC OBJ OBJ OBJ

sum-adjectives NOBJ OBJ OBJ

of_outside_DC OBJ OBJ OBJ

by_outside_DC NOBJ NOBJ NOBJ

Prediction of OBJ-DCs

- Prediction of **OBJ-DCs**:
 - **P-R**: as predicted
 - **of_outside_DC**: as predicted
 - **sg_outside_DC**: as predicted
 - **head_in_DC**: see next slide
 - **suffix -ment**: most stable (vs. *-ing*)

• Feature	Individual	Ablation	Corpus-Ablation
<i>P-R</i>	OBJ	OBJ	N/A
<i>suffix=ment</i>	OBJ	OBJ	OBJ
<i>suffix=ance</i>	OBJ	NOBJ	NOBJ
<i>suffix=ion</i>	NOBJ	OBJ	OBJ
<i>suffix=al</i>	OBJ	NOBJ	OBJ
<i>suffix=ing</i>	NOBJ	OBJ	NOBJ
<i>sg_inside_DC</i>	NOBJ	OBJ	NOBJ
<i>by_inside_DC</i>	OBJ	NOBJ	NOBJ
<i>sg_outside_DC</i>	OBJ	OBJ	OBJ
<i>head_in_DC</i>	OBJ	OBJ	OBJ
<i>sum-adjectives</i>	NOBJ	OBJ	OBJ
<i>of_outside_DC</i>	OBJ	OBJ	OBJ
<i>by_outside_DC</i>	NOBJ	NOBJ	NOBJ

- Possible explanations for the unstable predictions:
 - **sg_inside_DC**: OBJ-DCs appear more often in the plural than ASNs
 - **sum_adjectives, by-phrases**: data sparsity: max. 1.8%; 4.4%; 6.2%
 - *by*-phrases are not always external arguments: cf. *book by Chomsky*; the extraction data is also less reliable
- *by_outside_DC*: NOBJ (stable): we need to check if we have a correlation with SUBJ-DCs (there seems to be a SUBJ-DC correlation with verbs that frequently realize subjects)

Discussion on *Head_in_DC*

- High percentage of occurrences of a head inside compounds
- It indicates an **OBJ interpretation**
- High frequency of a head noun in compounds indicates its specialization for compounds
- The fact that this correlates with an OBJ reading shows us that **if a deverbal noun typically forms a compound with one of its arguments, then this argument will be the object**
- This supports Grimshaw's claim that **DCs embed event structure with OBJ/internal arguments**

Outline

- 1) Our Methodology: Data Extraction and Annotation
- 2) Case study 1: Corpus-based Morphosyntactic Features
- 3) Case study 2: Human Annotation of Result-vs-Process in DC
- 4) Interpretation of Results
- 5) Conclusion and Future Plans**

Conclusions

- Heads of DCs are ambiguous between ASNs and RNs and this affects the compositional interpretation of DCs
- We used corpus data on the **morphosyntax of the head noun** and human judgments on the **P-vs-R distinction** between ASNs and RNs
- Most results support Grimshaw's thesis that **DCs include event structure with internal/OBJ arguments**

Corpus features, however, do **worse than the P-R annotation**

- Data sparsity for crucial features (e.g., aspectual adverbials, adjectives), even more so for combinations
- Overlap between the different morphosyntactic properties
- Corpus evidence is noisier than human judgments and therefore weaker; however, human judgments are expensive and corpus is cheap
- Need for even larger corpora and better tools

Future Plans

- We extracted the **base verbs and their objects/subjects** – check whether the high frequency of a direct object with a verb correlates with an OBJ reading of the corresponding DCs
- We will **check if verb features improve the corpus-based results**
- Replicate the study with **DCs that have zero-derived heads** (e.g., *age limit, budget cut, tree climb*) within a new project
 - Grimshaw (1990), Borer (2013) claim that zero-derived nominals cannot be ASNs
 - Yet, OBJ-DCs seem to prevail also with DCs headed by zero-nouns (cf. Abrosimova 2017)
 - Could it be that OBJ is just a conceptually more salient reading and the OBJ reading does not come from hidden verbal properties?
- By investigating the properties of zero-nouns (as ASNs or RNs) and replicating the study with DCs, we can test if OBJ non-heads really come from some VP within DCs or are just conceptually more salient

Acknowledgments

- Annotators: Katherine Fraser, Whitney Frazier Peterson, Bethany Lochbihler
- Technical support from the SFB 732 INF-project – thanks to Kerstin Eckart
- Alla Abrosimova helped with other technical details
- Research funded by the DFG for the projects B1 – *The form and interpretation of derived nominals* – and D11 – *A Crosslingual Approach to the Analysis of Compound Nouns* – as part of the SFB 732 at the University of Stuttgart
- Current funding: DFG-Project IO 91/1-1: *Zero-derived Nouns and Deverbal Nominalization*

References

- Alla Abrosimoba. 2017. ‘Object’ and ‘Subject’ Readings in English Deverbal Compounds: Corpus-driven Insights from Verb Subcategorization Frames. MA thesis, University of Stuttgart.
- Artemis Alexiadou and Jane Grimshaw. 2008. Verbs, nouns, and affixation. In Florian Schäfer, editor, *Working Papers of the SFB 732 Incremental Specification in Context*, volume 1, pages 1–16. Universität Stuttgart.
- Hagit Borer. 2013. *Taking Form*. Oxford University Press, Oxford.
- Dan Gillick. 2009. Sentence boundary detection and the problem with the U.S. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, NAACL-Short ’09, pages 241–244, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Jane Grimshaw. 1990. *Argument Structure*. MIT Press, Cambridge, MA.
- Zhongqiang Huang, Mary Harper, and Slav Petrov. 2010. Self-training with products of latent variable grammars. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP ’10, pages 12–22, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Mirella Lapata. 2002. The disambiguation of nominalizations. *Journal of Computational Linguistics*, 28:3:357–388.
- Courtney Napoles, Matthew Gormley, and Benjamin Van Durme. 2012. Annotated Gigaword. In *Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction*, AKBCWEKEX ’12, pages 95–100, Stroudsburg, PA, USA. Association for Computational Linguistics.

Appendix-1: ASNs vs. RNs (Grimshaw 1990)

- Arguments are introduced by verbs via their event structure (aspectual properties, argument licensing, verbal properties)
 - ASNs preserve event structure & arguments from verbs; RNs do not
 - ASN: obligatory internal arguments (vs. RNs) (Grimshaw 1990: 50-52)
- (7) a. The **assignment** is to be avoided. (RN)
b. *The **constant assignment** is to be avoided. (ASN-RN)
c. The **constant assignment of unsolvable problems** is to be avoided. (ASN)
- *Constant* and *frequent* are aspectual modifiers when they appear with a singular noun => they require event structure (7b, c); if the noun is plural, it can be a RN:
- (9) The **constant assignments** were avoided by the students. (RN)

Appendix-1: ASNs vs. RNs (Grimshaw 1990)

- *Intentional, deliberate, careful* are agent-oriented modifiers and only appear with event structure => ASNs but not RNs
- (11) a. *The instructor's **intentional examination** took a long time.
b. The instructor's **intentional examination of the papers** took a long time.
- ASNs reject plural (not nominal enough) vs. RNs (Grimshaw 1990: 54)
- (18) a. The **assignments** were long. (RN)
b. *The **assignments of the problems** took a long time. (ASN)

Appendix 2: Corpus and Tools

- The Annotated Gigaword Corpus (Napoles et al. 2012) – LDC Catalog No. LDC2012T21
- 10-million documents from 7 news outlets: more than **4-billion words**
- Automatic processing and annotation:
 - 1. Segmentation (using Splitta - Gillick, 2009) and tokenization (using Stanford's CoreNLP pipeline)
 - 2. Lemmatization and POS tags (Stanford's CoreNLP pipeline)
 - 3. Treebank-style constituent parse trees (Huang et al. 2010, Avg. F score = 91.4 on WSJ sec 22)
 - 4. Syntactic dependency trees (using Stanford's CoreNLP pipeline for the conversion from constituency to dependency trees)
- Removed within-file (1010 files) duplicate sentences (170 >143 GB)
- Reliability: *of*-phrases (P: 91, R: 90.1); *by*-phrases (P: 85, R: 73.8)

Appendix-3: Selection of Target Head Nouns

- For each suffix, we selected 25 nouns derived from transitive verbs, which head NN compounds (no N before or after) in Gigaword;
- *Arrival* – the only unaccusative verb

Appendix-4: Post-processing of Annotations

- Initial database of 3111 compounds
- Select the reading that all 3 annotators agreed on (no matter how preferred): SUBJ, OBJ or OTHER
- Remove 271 errors, 771 disagreements
- Remove true ambiguous cases (2): *police killing*, *doctor referral*
- DCs headed by *arrival*: SUBJ > OBJ (we also manually checked other alternating verbs)
- Conflate OTHER and SUBJ to NOBJ (=> binary classification)

Appendix-5: Suffixes

- **Grimshaw (1990):** *ing*-nominals are always ASNs => **OBJ**
- **Borer (2013):** *ing* introduces the Originator (ext. arg.) itself and biases the DC towards an **OBJ reading**
 - Both theories predict a correlation between *ing* and OBJ, which we did not find, although OBJ is dominant with *-ing* in the full database:

ion:	OBJ: 71.3%	vs. NOBJ: 28.7%
ment:	OBJ: 65.5%	vs. NOBJ: 34.5%
ing:	OBJ: 64%	vs. NOBJ: 36%
al:	OBJ: 61.4%	vs. NOBJ: 38.6%
ance:	OBJ: 48.8%	vs. NOBJ: 51.2%
- Latinate suffixes (*-ion*, *-ment*, *-al*, *-ance*) are taken to behave similarly in theory, but we find a bias for OBJ in *-ion*, *-ment* and *-al*, and for NOBJ in *-ance*
- Further research is needed: both cleaner data on our side and linguistic research on the selectional preferences of suffixes

Appendix 6: *Process-vs-Result*: Examples

- 173 DCs with process-vs-result **rating 5** => OBJ-reading **91.32%**
- 268 DCs have process-vs-result **rating 1** => OBJ-reading **31.71%**

<i>process-vs-result 5</i>		<i>process-vs-result 1</i>	
baking	growing	acquaintance	burial
breeding	laundering	acquittal	education
building	measuring	allowance	ignorance
chasing	mongering	announcement	inheritance
cultivation	planting	approval	insurance
construction	trafficking	assassination	payment
consumption	testing	assignment	proposal
evacuation	training	assistance	referral
exploration	weighting	building	writing

Appendix-7: Comparison to NLP Studies

- Our best performances:
 - 72.7% vs 50% => **22.7%** improvement (corpus data)
 - 80.3% vs 50% => **30.3%** improvement (corpus & judgments)
- Previous work in the NLP literature targets state-of-the-art performance in prediction with methods different from ours
- Our purpose was to start from linguistic theory and test linguistic hypotheses
- These studies include more suffixes (-**er**, -**ee**) and zero-derived nouns; -**er** and -**ee** are biased, so they are more predictive;
- We had only 'event'-denoting suffixes, where SUBJ/OBJ are similarly conceivable
- Lapata (2002): 86.1% vs. 61.5% => **24.6%** above the baseline