Analysing domain suitability of a sentiment lexicon by identifying distributionally bipolar words

Lucie Flekova, Ubiquitous Knowledge Processing Lab (UKP, TU Darmstadt),
Daniel Preotiuc-Pietro (University of Pennsylvania) and Eugen Ruppert (LangTech, TU Darmstadt)

Lazy guy

Lazy sunday

Word polarity lexicons

- SemEval 2014, 2015
  - vast majority of systems still based on sentiment lexica + supervised cl.

Word polarity lexicons

- SemEval 2014, 2015
  - vast majority of systems still based on sentiment lexica + supervised cl.

  - **Cold**
  - **Dark**
  - **Limited**

  - **Wisdom**
  - **Sincere**

Word polarity lexicons

- SemEval 2014, 2015
  - vast majority of systems still based on sentiment lexica + supervised cl.

- **Cold**
- **Dark**
- **Limited**

- **Wisdom**
- **Sincere**

Word polarity lexicons

- SemEval 2014, 2015
  - vast majority of systems still based on sentiment lexica + supervised cl.

  - **Cold** :cold beer (+) or cold food (-),
  - **Dark:** dark chocolate (+) or dark soul (-).
  - **Limited:** Limited edition (+) or limited intellect (-)

  - **Wisdom:** wisdom tooth (-) or wisdom source (+)
  - **Sincere:** sincere condolences (-) or sincere love (+)

- Lexicon ambiguities at a contextual level
- Sense disambiguation does not help here

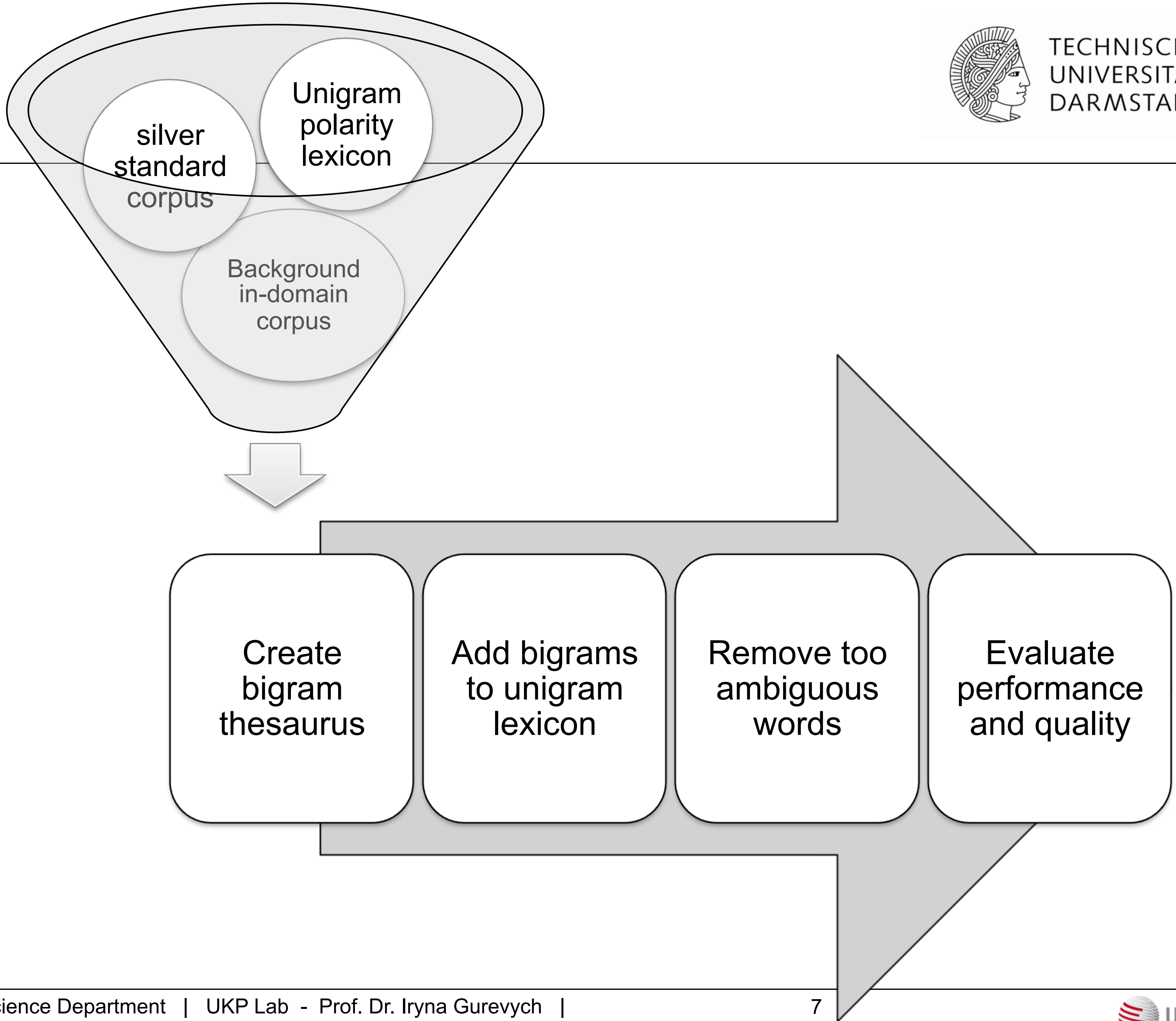
why limit
stress reliever
mission impossible
lazy sunday
desperate housewives
cold beer

guilty pleasure
bloody mary
wisdom tooth
gold digger
sincere condolences

Assessing lexicon suitability for new platform

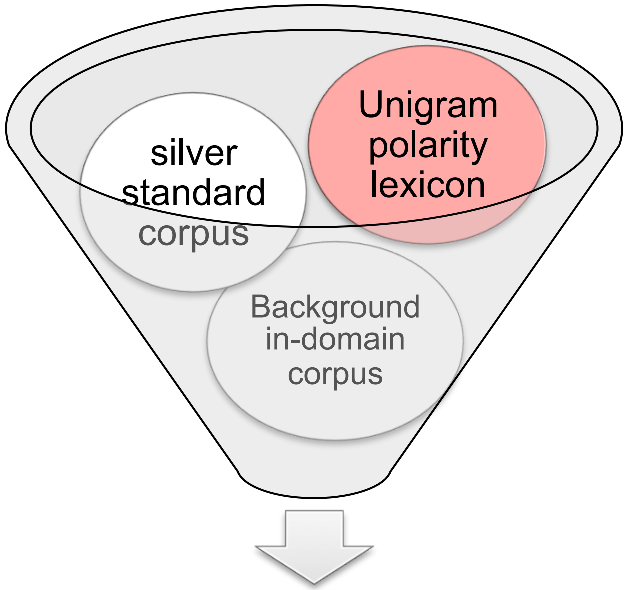
TECHNISCHE UNIVERSITÄT DARMSTADT

How do you quantify if a lexicon you use does more harm than help to the data you use, and how should you adapt it?



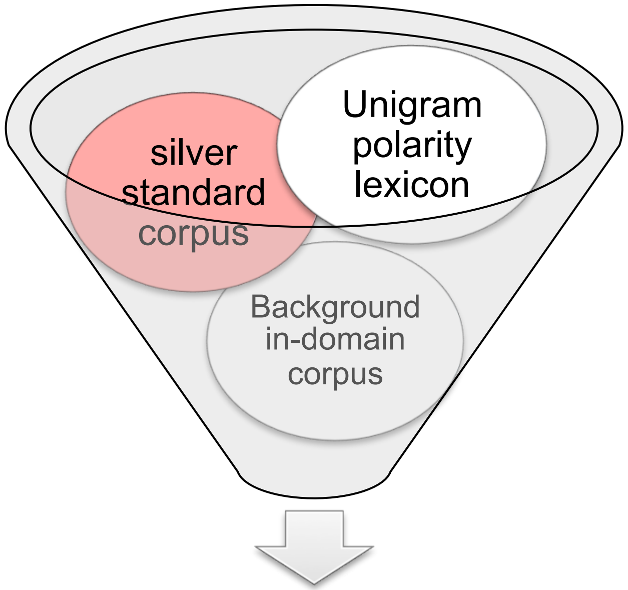
silver standard corpus
Unigram polarity lexicon
Background in-domain corpus
Create bigram thesaurus
Add bigrams to unigram lexicon
Remove too ambiguous words
Evaluate performance and quality

Ingredient 1: Unigram polarity lexicon



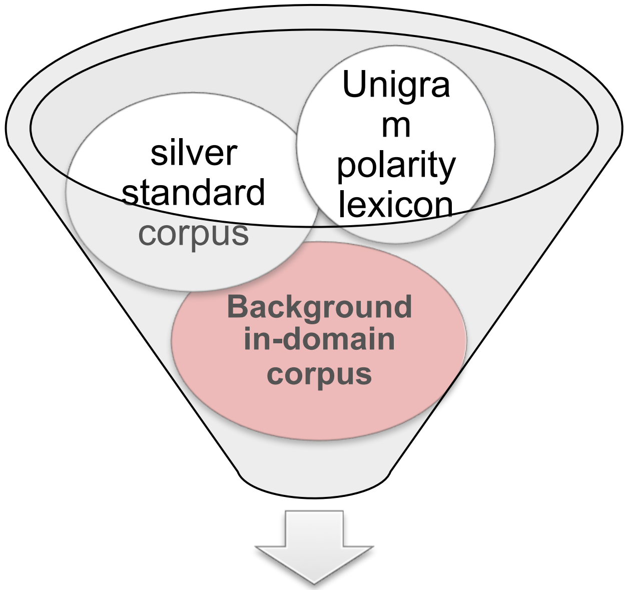
- We demonstrate our approach on two polarity lexicons consisting of single words:
- the lexicon of Hu and Liu (Hu and Liu, 2004)
- the MPQA lexicon (Wilson et al., 2005).

Ingredient 2: Silver standard sentiment corpus

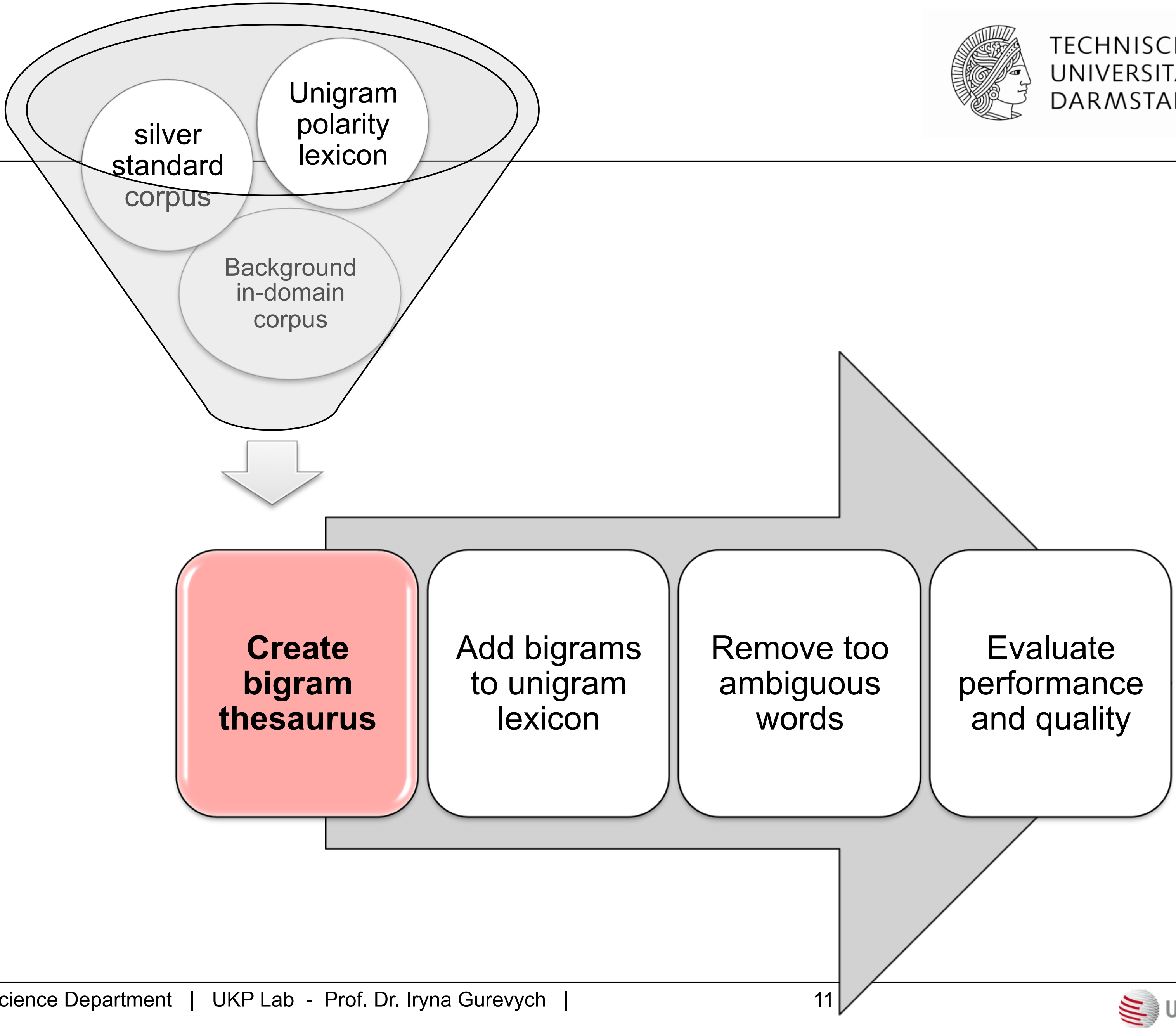


- 1.6 million tweets from the Sentiment140 data set (Go et al., 2009)
- collected by searching for positive and negative emoticons

Ingredient 3: Twitter corpus (unlabeled data)



- Twitter corpus of 1 % of all English tweets from the year 2013 = **460 million tweets**



silver standard corpus
Unigram polarity lexicon
Background in-domain corpus
Create bigram thesaurus
Add bigrams to unigram lexicon
Remove too ambiguous words
Evaluate performance and quality

Creating Twitter Bigram Thesaurus

- Using not PMI, but its adaptation Lexicographer's Mutual Information (LMI)

$$\mathrm{PMI}(w,c) = \log_2\left(\frac{f(w,c)}{f(w)\cdot f(c)}\right)$$

$$\mathrm{LMI}(w,c) = \mathrm{PMI}(w,c)\cdot f(w,c)$$

Distributional Sentiment:

- LMI computed separately on positive and negative tweets from Sentiment140 (Go et al., 2009, 1.6m tweets)

- Bigram LMI over a corpus of positive, resp. negative tweets

- For comparability of LMI_pos and LMI_neg, bigrams weighted by their relative frequency in POS and NEG data

$$\mathrm{LMI}_{neg_{REL}}(w,c) = \mathrm{LMI}_{neg}(w,c)\cdot \frac{f_{neg}(w,c)}{f_{neg}(w,c)+f_{pos}(w,c)}$$

Creating Twitter Bigram Thesaurus

TECHNISCHE UNIVERSITÄT DARMSTADT

Distributional Sentiment Silver:

- LMI computed separately on positive and negative tweets from Sentiment140 (Go et al., 2009, 1.6m tw.)

+

Distributional Thesaurus:

- computed on 80 million English Tweets based on left and right neighbor bigrams

- limited size of silver standard data = not the most reliable scores

-> further boost of LMI by incorporating scores from a background corpus (LMIGLOB)

LMI_neg_glob(word, context) = LMI_neg(word, context) x LMI_glob(word, context)
LMI_pos_glob(word, context) = LMI_pos(word, context) x LMI_glob(word, context)

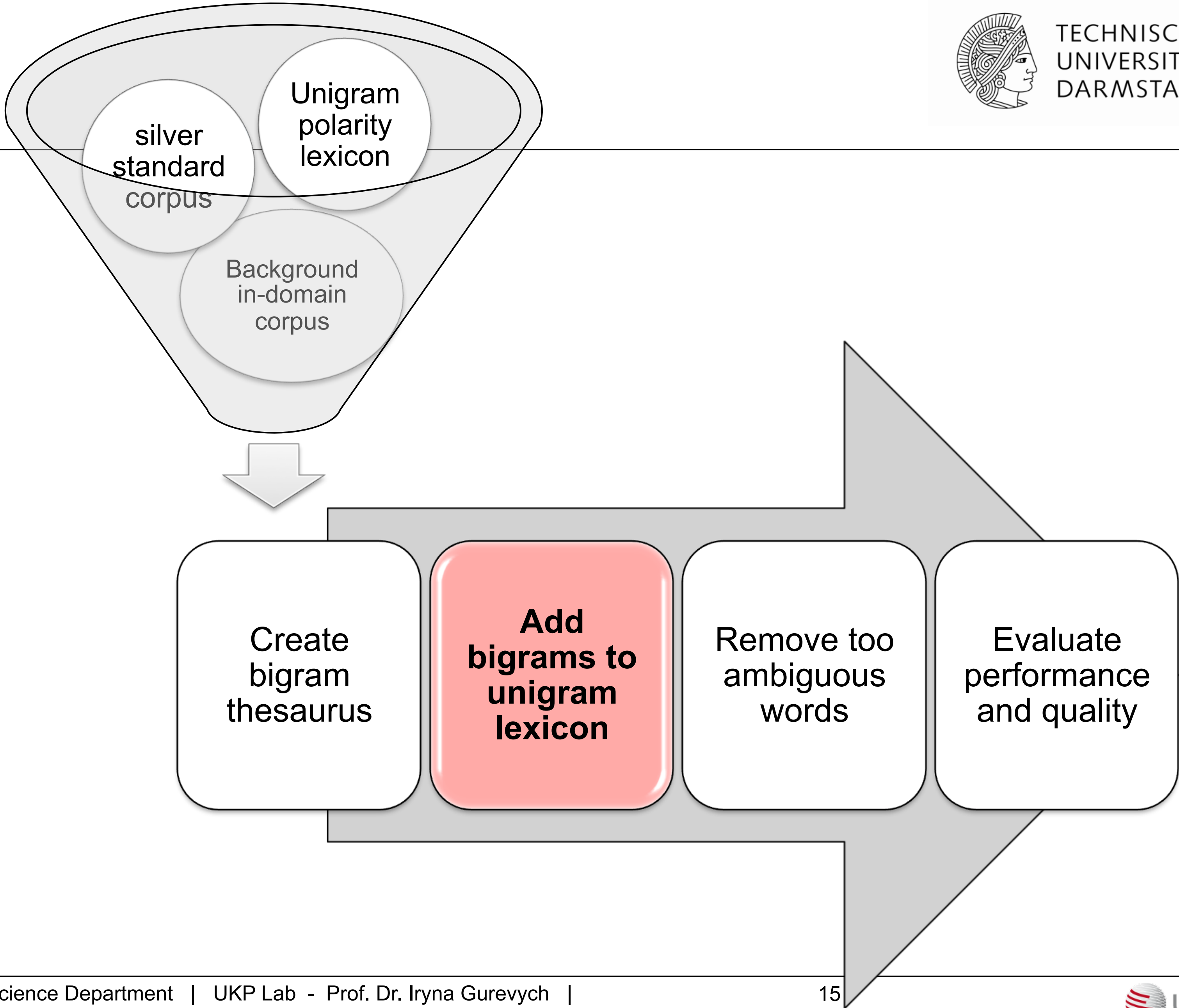
- Emphasizes frequent & informative bigrams, even when their score in one polarity data set is low

Creating Twitter Bigram Thesaurus

TECHNISCHE UNIVERSITÄT DARMSTADT

global LMI semantic orientation = LMI_pos_glob – LMI_neg_glob

dark_past = -128.14,
dark_chocolate=+1558.96,
...


silver standard corpus
Unigram polarity lexicon
Background in-domain corpus
Create bigram thesaurus
Add bigrams to unigram lexicon
Remove too ambiguous words
Evaluate performance and quality

Twitter Bigram Thesaurus: invert polar bigrams

TECHNISCHE UNIVERSITÄT DARMSTADT

DARK: dark_past = -128.14, dark_chocolate=+1558.96, ...

Negative word to positive bigram:

Hu&Liu		MPQA	
why	**limit**	**vice**	versa
sneak	peek	**stress**	reliever
mission	**impossible**	calmed	**down**
lazy	sunday	**deep**	breath
desperate	housewives	**long**	awaited
cold	beer	**cloud**	computing
guilty	pleasure	**dark**	haired
belated	birthday	**bloody**	mary

Positive word to negative bigram:

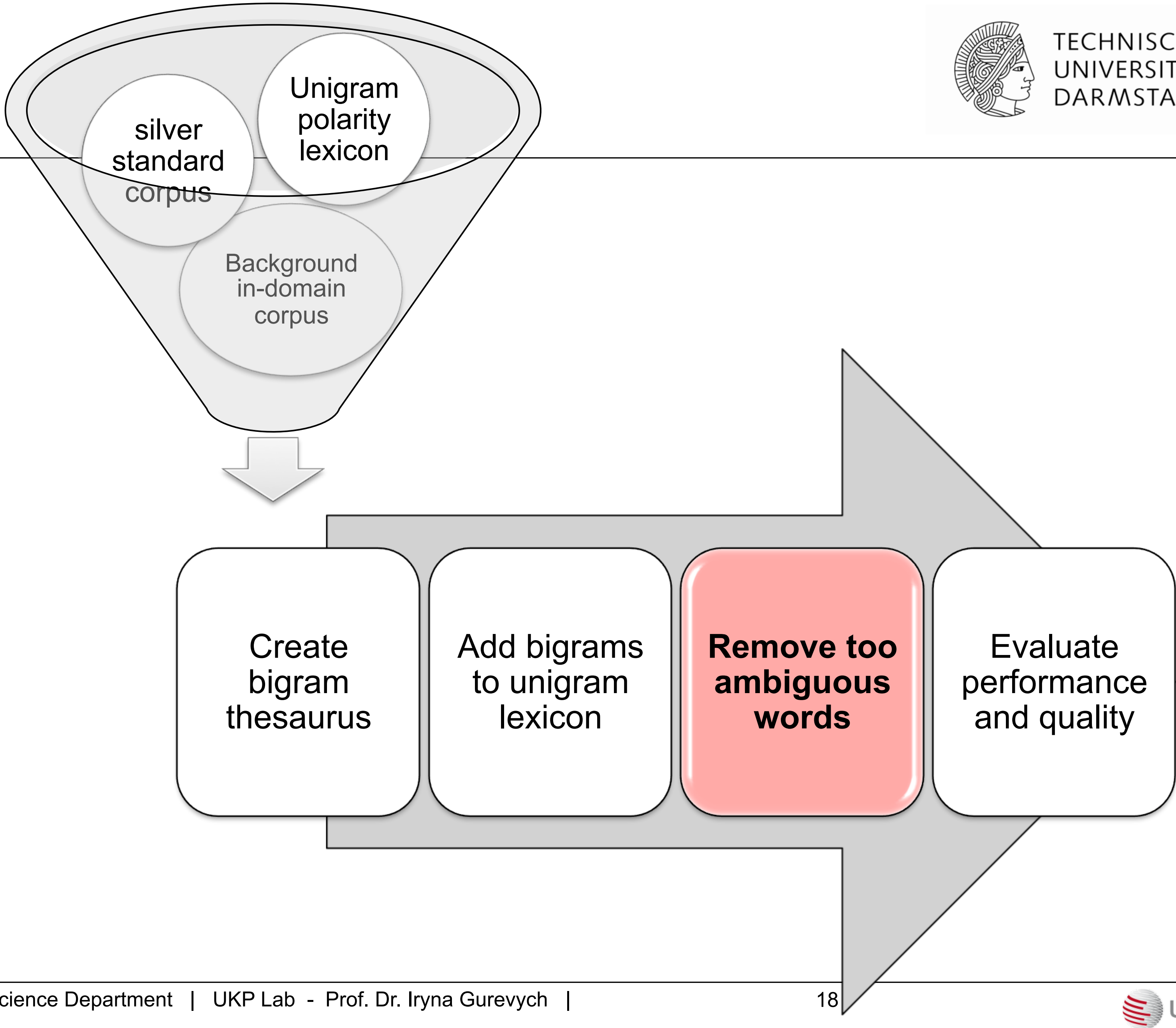
Hu&Liu		MPQA	
good	luck	**super**	duper
wisdom	tooth	**happy**	camper
oh	**well**	**just**	puked
gotta	**work**	**heart**	breaker
hot	outside	**gold**	digger
feels	**better**	**light**	bulbs
super	tired	**sincere**	condolendes
enough	money	**frank**	iero

https://www.ukp.tu-darmstadt.de/data/sentiment-analysis/inverted-polarity-bigrams/

Twitter Bigram Thesaurus: observations

Polarity shifting occurs in a broad range of situations, e.g.:

- polar word as an intensity expression:
  - **super** tired
- polar word in names:
  - **desperate** housewives,
    frank iero
- multiword expressions, idioms and collocations
  - **cloud** computing,
    sincere condolences,
    light bulbs
- polar nominal context
  - **cold** beer/person,
    dark chocolate/thoughts,
    stress reliever/management,
    guilty pleasure/feeling

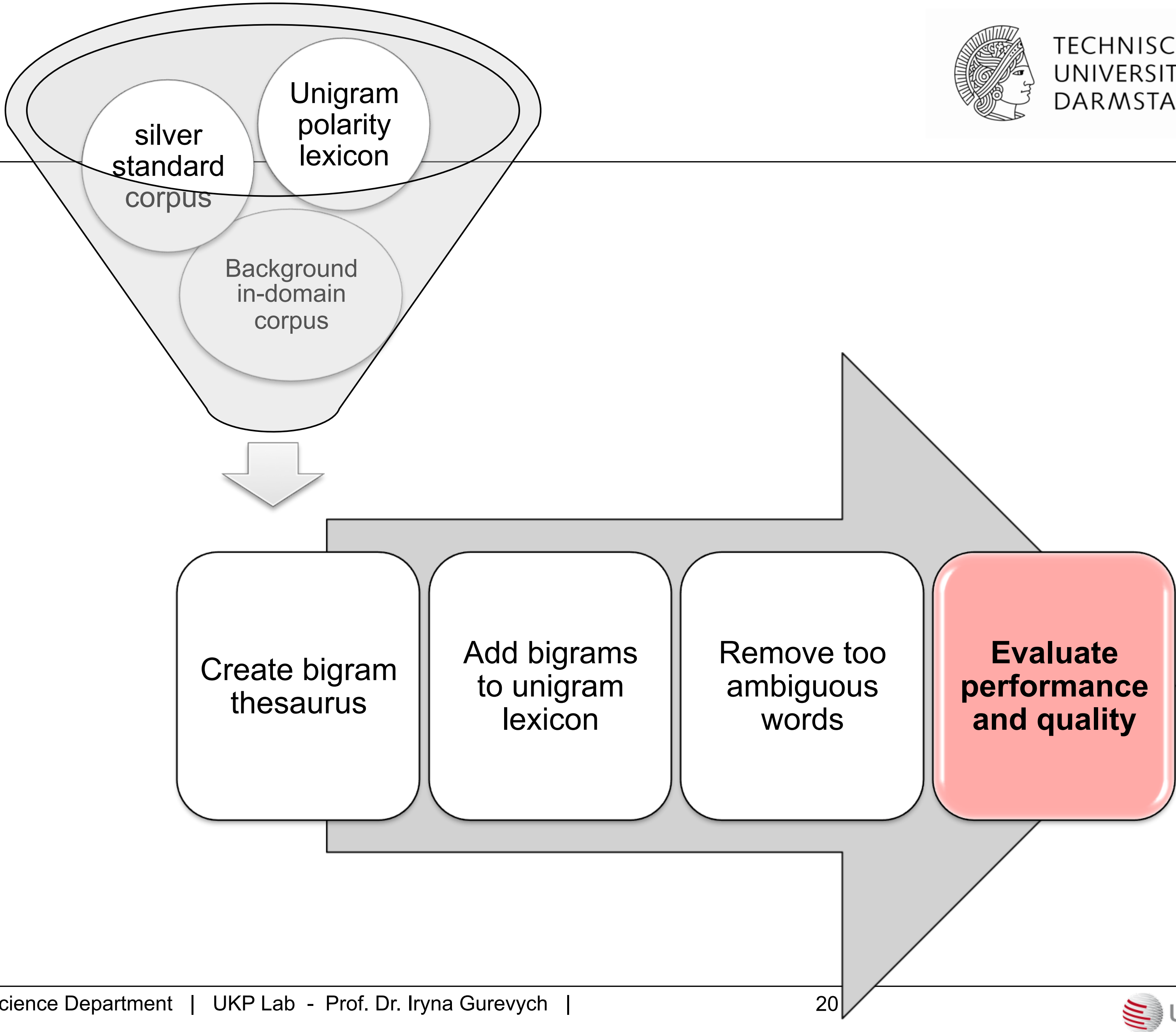

silver standard corpus
Unigram polarity lexicon
Background in-domain corpus
Create bigram thesaurus
Add bigrams to unigram lexicon
Remove too ambiguous words
Evaluate performance and quality

Finding the most ambiguous unigrams

Some words occur in many contexts with both original and switched polarity = harmful in either of the polarity sides = better to remove it

Word ambiguity = (#positive contexts - #negative contexts) / #contexts

Hu&Liu		MPQA	
hot	.022	just	-.002
support	.022	less	.009
important	-.023	sound	-.011
super	-.043	real	.027
crazy	-.045	little	.032
right	-.065	help	-.037
proper	-.093	back	-.046
worked	-.111	mean	.090
top	.113	down	-.216
enough	-.114	too	-.239


silver standard corpus
Unigram polarity lexicon
Background in-domain corpus
Create bigram thesaurus
Add bigrams to unigram lexicon
Remove too ambiguous words
Evaluate performance and quality

Test corpus

- Facebook posts rated for affect by two psychology experts on scale 1 – 9 (1 = strong negative, 9 = strong positive sentiment)
  - normal distribution of ratings
  - inter-annotator agreement: weighted Cohen's κ = 0.61 on exact score

- Neutral posts for our task removed, posts containing no lexicon word removed (20%) => left with:
  - 1,601 posts for MPQA
  - 1,526 posts for Hu & Liu.

Sentiment polarity prediction results

	Features	Acc. HL	Acc. MPQA
Baseline	Unigrams	.7070	.6608
Add bigrams to unigram lexicon	Uni+bigrams	.7215	.6633
	Uni+bigramsPos	.7123	.6621
	Uni+bigramsNeg	.7163	.6621
Remove too ambiguous words	Pruned	.7228	.6627
	Pruneg+bigrams	**.7333**	**.6646**
	Pruned+bigramsPos	.7150	.6633
	Pruned+BigramsNeg	.7287	.6640
	All in-domain bigrams	.6907	**.7008**

Error sources

- Remaining ambiguity due to more complex phrase structure.
  - Helpful:
    '*holy shit, tech support...*',
    holy (+1), support (+1) X holy shit (-0.35), tech support (-0.85)
  - Not helpful:
    holy shit (-) in
    'holy shit monday night was amazing'.
  - work ahead (-) in
    '*New house....yeah!! lots of work ahead of us!!!*'

- Longer negation window
  - *feeling sick* (-) in
    'Isn't feeling sick woohoo!'

- Positive bigrams which have learnt negativity from a broader context:
  - Not helpful: *looking good (-)*, *happy camper (-) in*
    'someone is a happy camper!', 'It is looking good!'

Intrinsic qualitative evaluation

- Raters saw a list of 100 bigrams of each lexicon:

- *"Which polarity does this word pair have:"*

WISDOM TOOTH
( ) positive, ( ) negative, ( ) neutral

- Each bigram is rated by three annotators and the majority vote is selected.

Intrinsic qualitative evaluation

- Cohen’s Kappa = 0.55

Hu & Liu	Pos	Neu	Neg
Pos	**30**	10	9
Neg	11	10	**30**

MPQA	Pos	Neu	Neg
Pos	**21**	24	3
Neg	5	18	**25**

- Some of the bigrams, especially for MPQA, assessed as objective
- confusion between negatively and positively labeled bigrams very low

Conclusions

1. Our method helps to determine how much and why a given general-purpose lexicon is useful in a specific target domain or platform
2. Technique to identify frequent bigrams of inverted polarity (domain shift), and to identify unigrams with high bipolarity (likely neutral)
3. LMI scores capture human perception of polarity and improve performance on our task

- Our bigram lexicon extension of Hu and Liu available at:

https://www.ukp.tu-darmstadt.de/data/sentiment-analysis/inverted-polarity-bigrams/

TECHNISCHE UNIVERSITÄT DARMSTADT

Thank you for your attention!

flekova@ukp.informatik.tu-darmstadt.de