Automatically Creating a Lexicon of Verbal Polarity Shifters: Mono- and Cross-lingual Methods for German

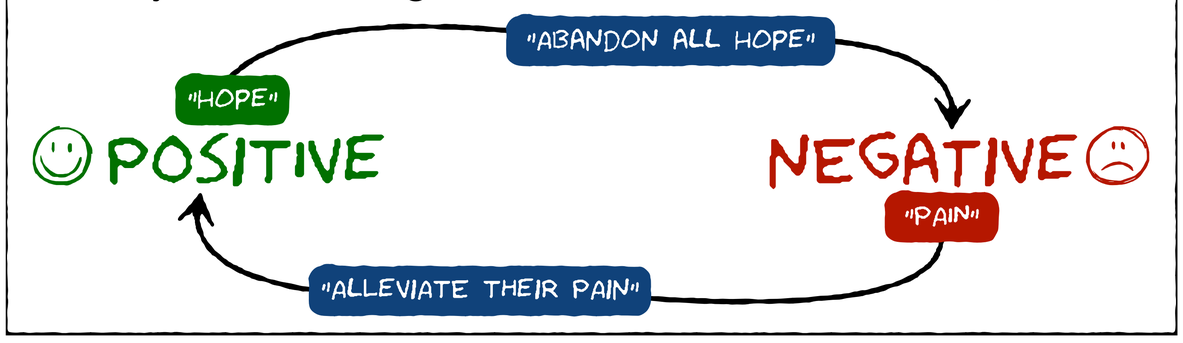
Marc Schulder, Michael Wiegand
Spoken Language Systems
Saarland University, Germany

Josef Ruppenhofer
Institute for German Language
Mannheim, Germany

Introduction

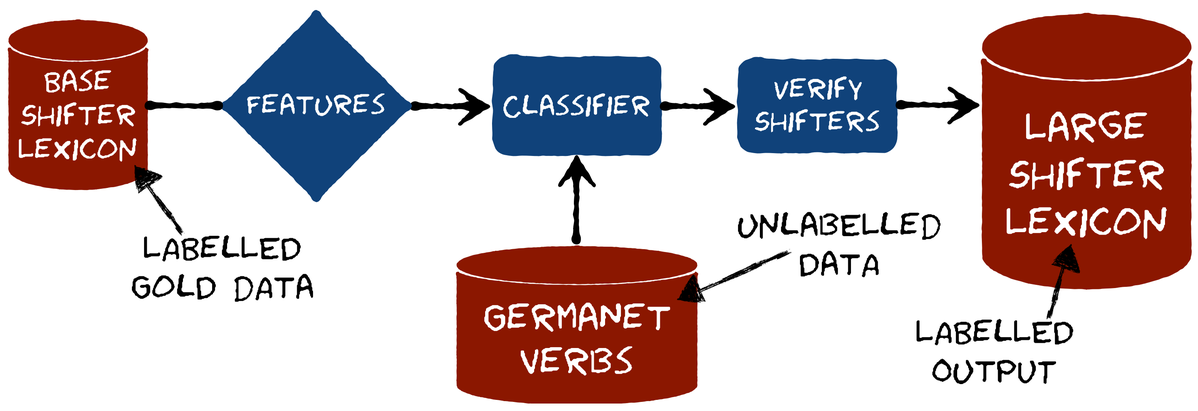
Polarity Shifters

Shifters are content words that affect polarities in ways that are very similar to negation words like *no*, *not*, *neither*, etc.



Goal: A list of German verbs that are polarity shifters.
Uses: NLP tasks affected by negation, e.g. polarity classification.

Bootstrapping Approach



Task: Binary classification: Shifter vs Non-shifter
Gold Data: 2000 German verbs (11% shifters).
Unlabelled Data: 7262 German verbs.

Source	Features
Data	Similarity to Negation, Polarity Clash, Verb Particles, Co-occurrence with *any*, Anti-Shifter
Resource	GermaNet, Salsa FrameNet, EffektGermaNet
Xling	Bilingual Dictionary, Cross-lingual Embedding

Will this work? It does for English (Schulder et al., IJCNLP 2017).
What's new? Applied to German, added cross-lingual features.
Why not deep learning? Not enough data!

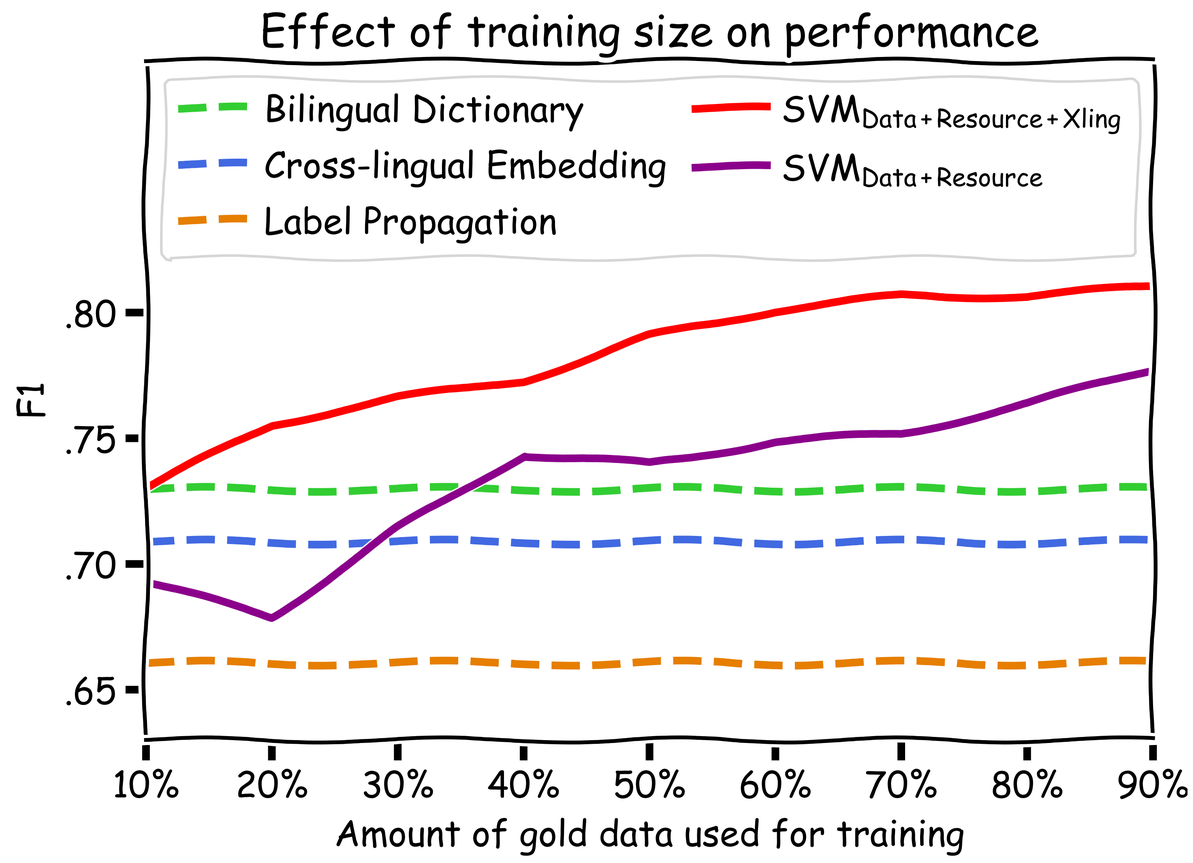
Results

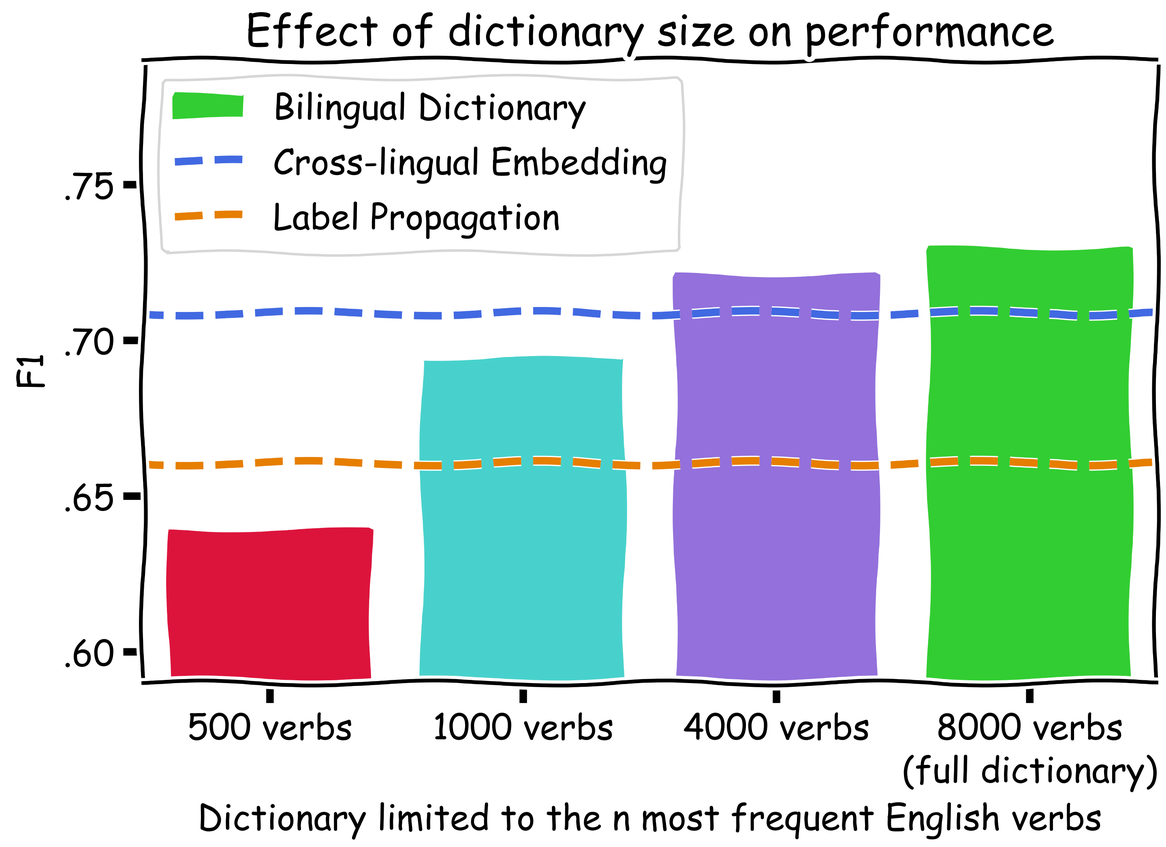
Output: Lexicon of 677 German verbal shifters.
Data: `https://github.com/uds-lsv/coling2018`
Lessons learned:

- Classification of shifters works for German, too.
- Cross-lingual features are strong feature addition.
- Dictionary and embeddings are both useful.

Evaluation

	Classifier	Prec	Rec	F1
Base	Majority Label	44.4	50.0	47.0
	Label Propagation	67.2	65.0	66.1
Xling	Cross-lingual Embedding	67.6	74.6	70.9
	Bilingual Dictionary	69.2	77.3	73.0
SVM	Data	60.8	72.6	66.2
	Resource	79.4	73.9	76.4
	Data+Resource	79.0	76.7	77.7
	Data+Resource+Xling	**80.3**	**82.0**	**81.0**





Future Extension to Other Languages

Sounds great!
What approach should I use for a particular language?