Speech Recognition

Lecture 6: Language Modeling Software Libray

Mehryar Mohri
Courant Institute and Google Research
mohri@cims.nyu.com

Software Library

- GRM Library: Grammar Library. General software collection for constructing and modifying weighted automata and transducers representing grammars and statistical language models (Allauzen, MM, and Roark, 2005).

  http://www.research.att.com/projects/mohri/grm

This Lecture

- Counting
- Model creation, shrinking, and conversion
- Class-based models

Overview

- Generality: to support the representation and use of the various grammars in dynamic speech recognition.
- Efficiency: to support competitive large-vocabulary dynamic recognition using automata of several hundred million states and transitions.
- Reliability: to serve as a solid foundation for research in statistical language modeling.

Content

- Statistical Language Models: creating, shrinking, and converting language models.
- Grammar compilation:
  - weighted context-dependent rules, weighted context-free grammars (CFGs).
  - regular approximations of CFGs.
- Text and grammar processing utilities:
  - local grammars, suffix automata.
  - local determinization.
  - counting and merging.

Language Modeling Tools

- Counts: automata (strings or lattices), merging.
- Models:
  - Backoff or deleted interpolation smoothing.
  - Katz or absolute discounting.
  - Knesser-Ney models.
- Shrinking: weighted difference or relative entropy.
- Class-based modeling: straightforward.

Corpus

- Input:

Corpus	Labels	
hello.	<s>	1
bye.	</s>	2
hello.	<unknown>	3
bye bye.	hello	4
	bye	5

- Program:

```
    farcompilestrings -i labels corpus.txt > foo.far
or  cat lattice1.fsm ... latticeN.fsm > foo.far
```

Counting

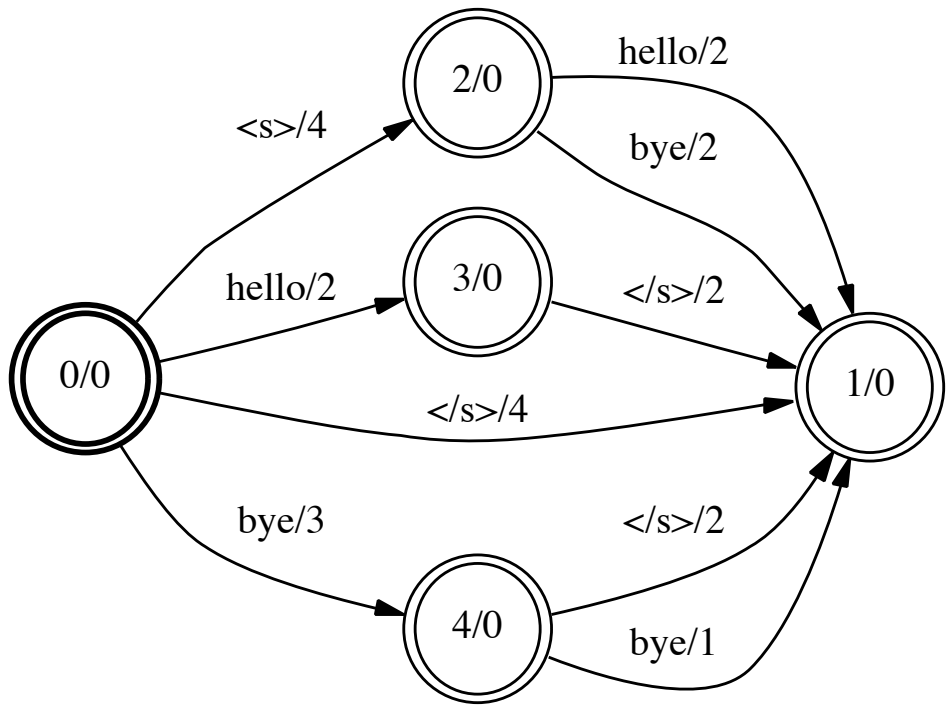
- Weights:
  - use `fsmpush` to remove initial weight and create a probabilistic automaton.
  - counting from far files.
  - counts produced in log semiring.
- Algorithm:
  - applies to all probabilistic automata.
  - In particular, no cycle with weight zero or less.

Counting

- Program:

```
grmcount -n 2 -s 1 -f 2 foo.far > foo.2.counts.fsm
grmmerge foo.counts.fsm bar.counts.fsm > foobar.counts.fsm
```

- Graphical representation:



This Lecture

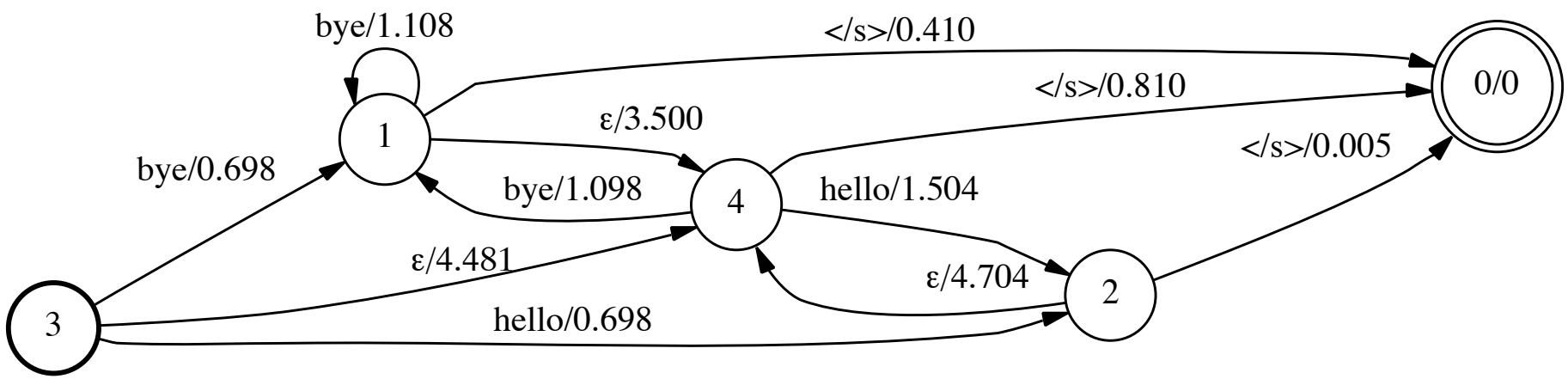
- Counting
- Model creation, shrinking, and conversion
- Class-based models

Creating Back-off Model

- Program:

  grmmake foo.2.counts.fsm > foo.2.lm.fsm

- Graphical representation:

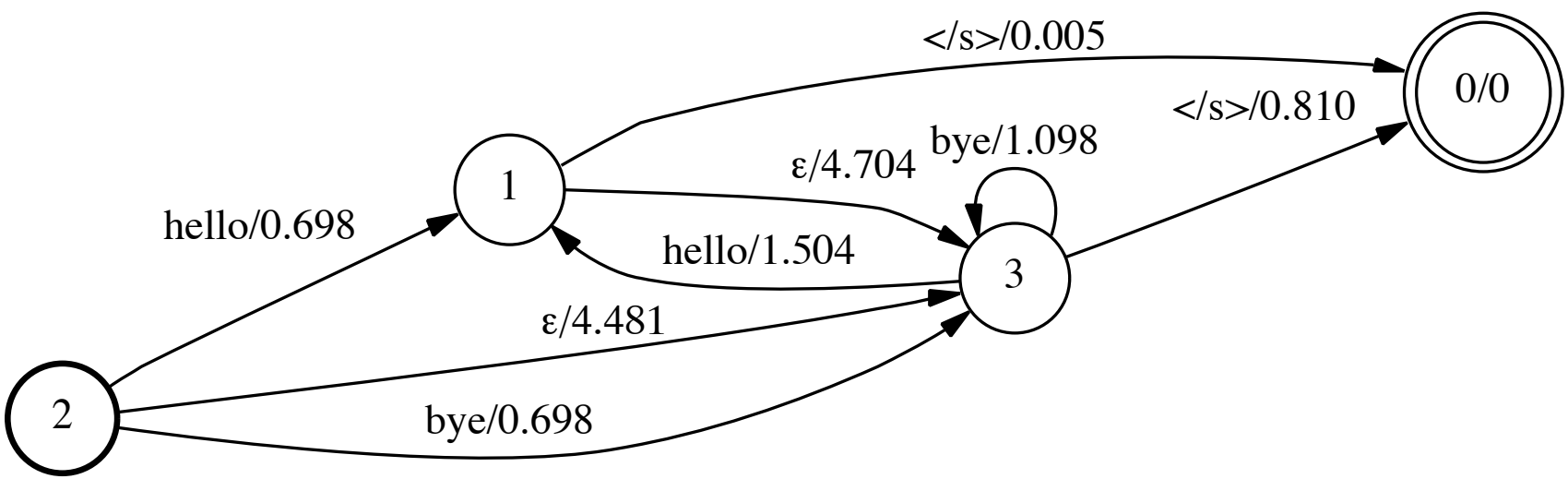


Shrinking Back-off Model

- Program:

  grmshrink -c 4 foo.2.lm.fsm > foo.2.s4.lm.fsm

- Graphical representation:



Back-off Smoothing

- Definition: for a bigram model,

$$\Pr[w_i|w_{i-1}] = \begin{cases} \dfrac{d_{c(w_{i-1}w_i)}\, c(w_{i-1}w_i)}{c(w_{i-1})} & \text{if } k > 0; \\ \alpha \Pr[w_i] & \text{otherwise;} \end{cases}$$

where

$$d_k = \begin{cases} 1 & \text{if } k > 5; \\ \approx \dfrac{(k+1)n_{k+1}}{k n_k} & \text{otherwise.} \end{cases}$$

Conversion of Back-off Model

- Interpolated model:

  grmconvert foo.lm.fsm -m interpolated > foo.int.lm.fsm

- Failure function model - failure class:
  - efficient representation of backoff models.
  - requires on-the-fly composition for decoding.

  grmconvert foo.lm.fsm -e failure_transitions > foo.flm.fsm

- Failure function model - basic class:
  - state-splitting: correct in tropical semiring.
  - pre-optimization: on-the-fly composition not required.

  grmconvert foo.lm.fsm -e state_splitting > foo.elm.fsm

This Lecture

- Counting
- Model creation, shrinking, and conversion
- Class-based models

Class-Based Models

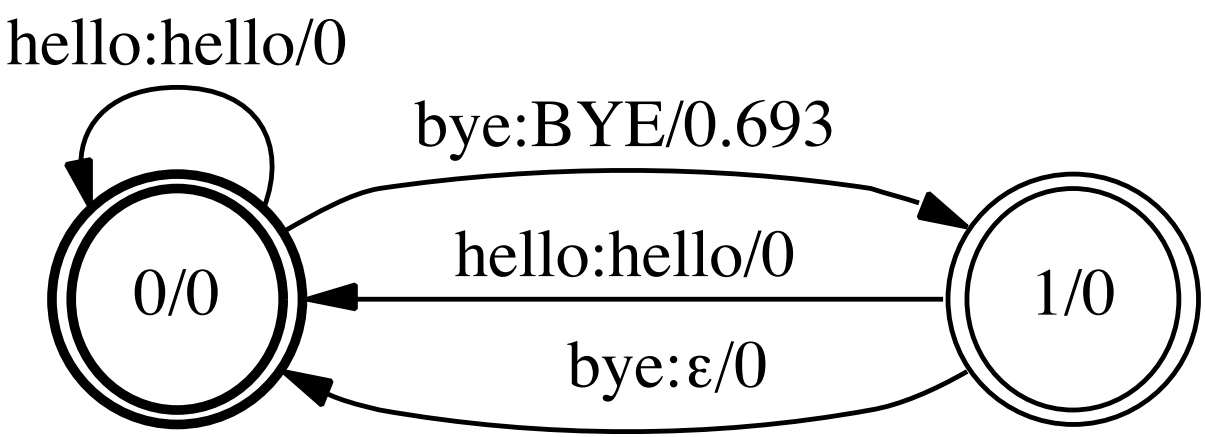
- Simple class-based models:

$$\Pr[w_i|h] = \Pr[w_i|C_i] \Pr[C_i|h].$$

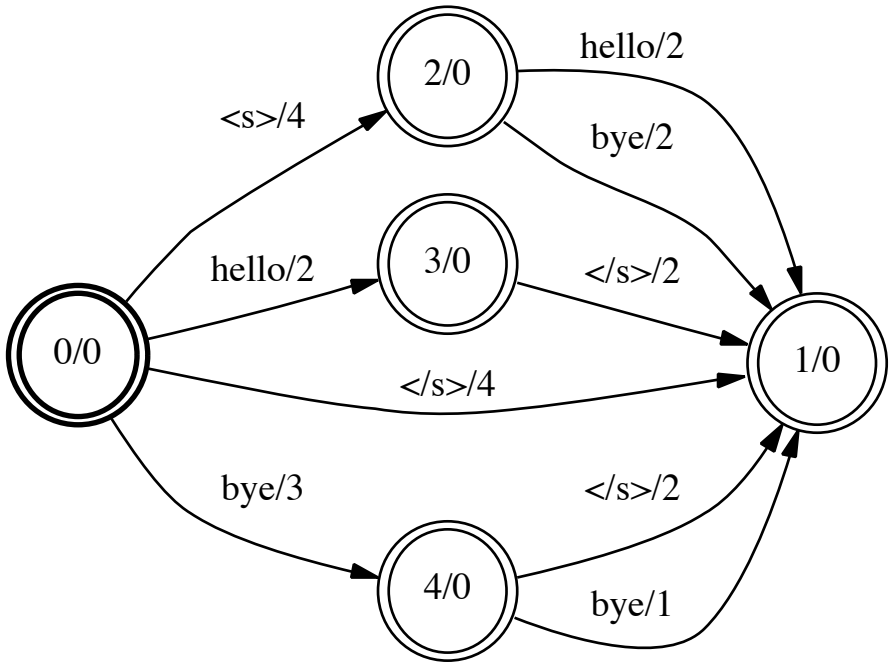
- Methods in GRM: no special utility needed.
  - create transducer mapping strings to classes.
  - use `fsmcompose` to map from word corpus to classes.
  - build and make model over classes.
  - use `fsmcompose` to map from classes to words.
- Generality: classes defined by weighted automata.

Class-Based Model - Example

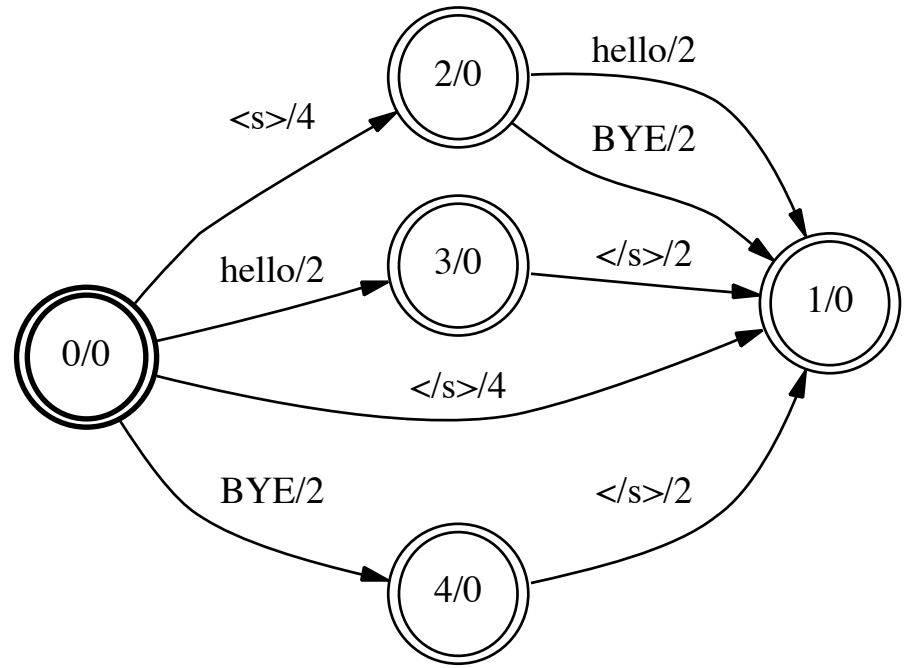
- Example: BYE = {bye, bye bye}.
- Graphical representation: mapping from strings to classes.



Class-Based Model - Counts

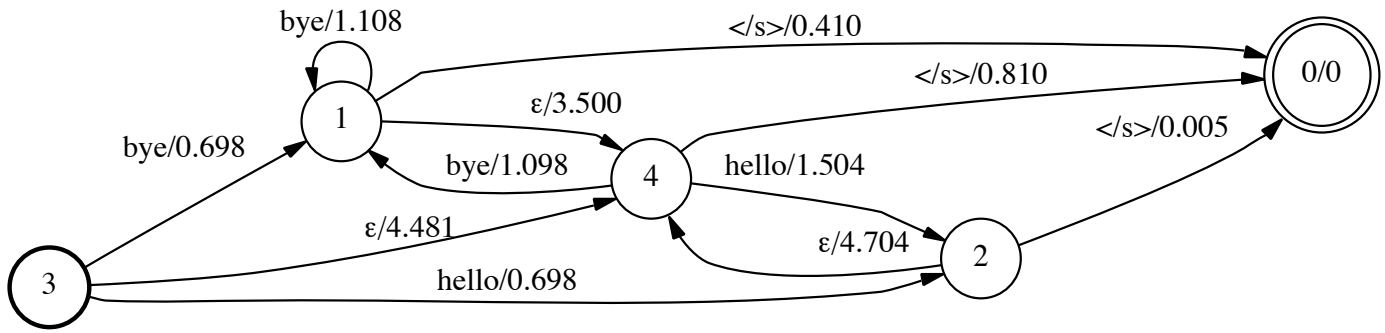


Original counts.

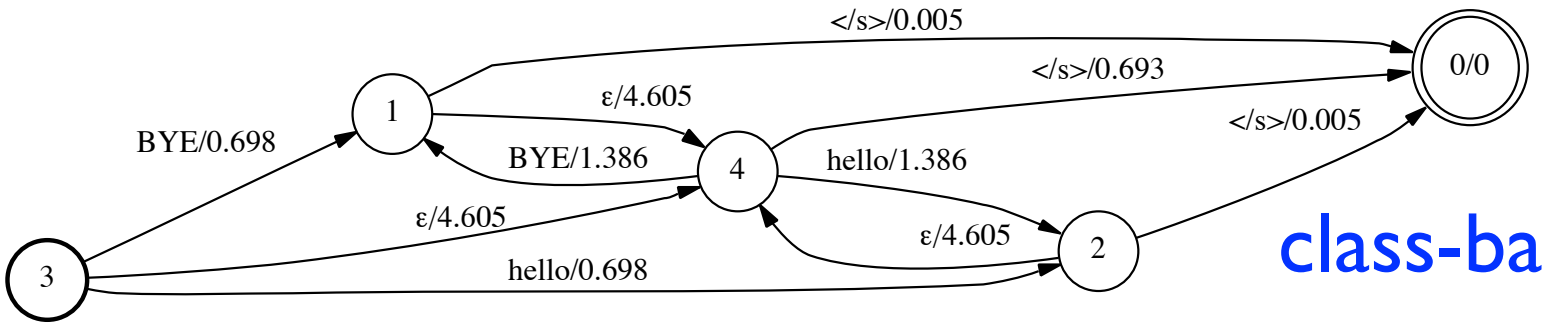


Class-based counts.

Models

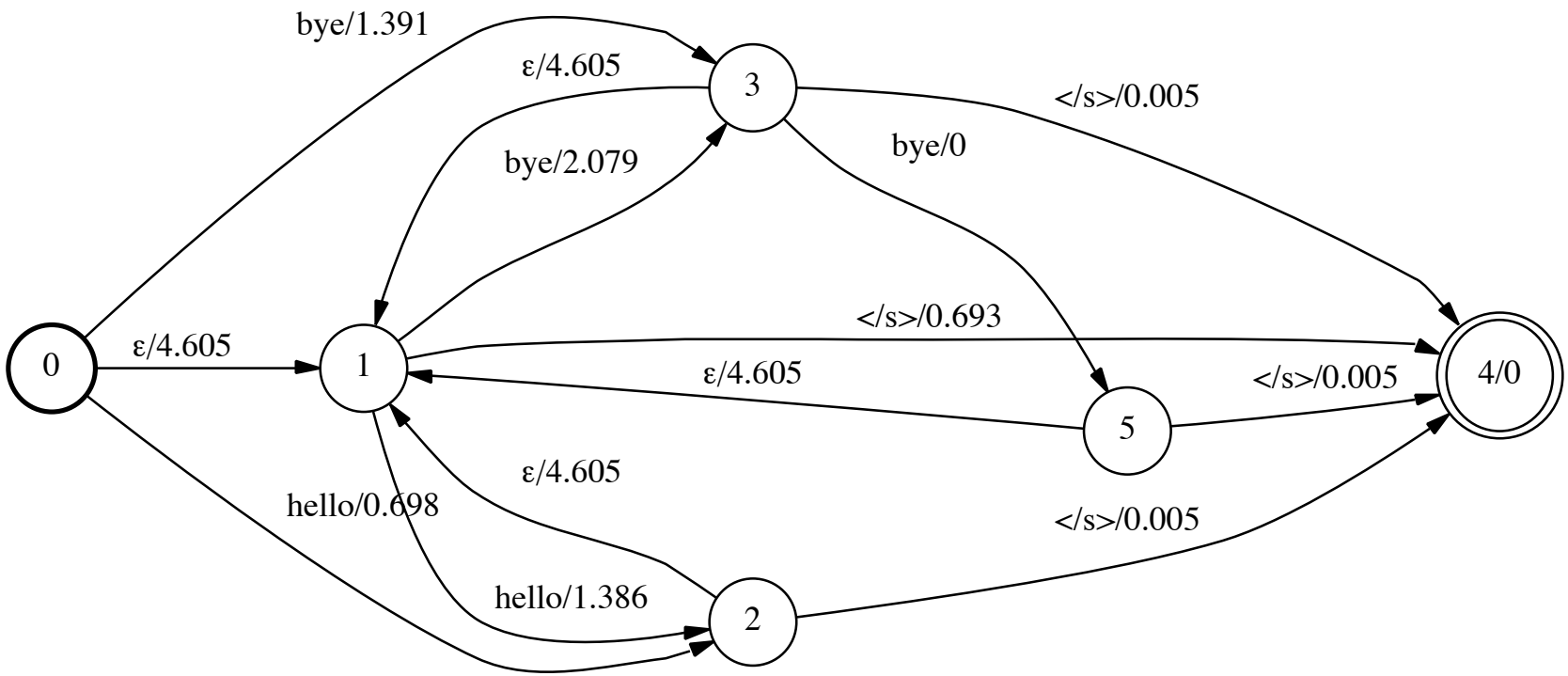


original model.



class-based model.

Final Class-Based Model



Conclusion

- GRM Library: general utilities for text and grammar processing.
  - generality: e.g., counts from arbitrary automata or a class-based-model.
  - efficiency: practical, e.g., counting lattices in 1/50*th* real-time.
  - testing: statistical tests for reliability.

References

- Cyril Allauzen, Mehryar Mohri, and Brian Roark. Generalized Algorithms for Constructing Statistical Language Models. In 41st *Meeting of the Association for Computational Linguistics (ACL 2003)*, Proceedings of the Conference, Sapporo, Japan. July 2003.

- Cyril Allauzen, Mehryar Mohri, and Brian Roark. The Design Principles and Algorithms of a Weighted Grammar Library. *International Journal of Foundations of Computer Science*, 16(3): 403-421, 2005.

- Peter F. Brown, Vincent J. Della Pietra, Peter V. deSouza, Jennifer C. Lai, and Robert L. Mercer. 1992. Class-based n-gram models of natural language. *Computational Linguistics*, 18 (4):467-479.

- Stanley Chen and Joshua Goodman. An empirical study of smoothing techniques for language modeling. Technical Report, TR-10-98, Harvard University. 1998.

- William Gale and Kenneth W. Church. What's wrong with adding one? In N. Oostdijk and P. de Hann, editors, *Corpus-Based Research into Language*. Rodolpi, Amsterdam.

- Good, I. The population frequencies of species and the estimation of population parameters, *Biometrica*, 40, 237-264, 1953.

References

- Frederick Jelinek and Robert L. Mercer. 1980. Interpolated estimation of Markov source parameters from sparse data. In *Proceedings of the Workshop on Pattern Recognition in Practice*, pages 381-397.

- Slava Katz . Estimation of probabilities from sparse data for the language model component of a speech recognizer, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35, 400-401, 1987.

- Reinhard Kneser and Hermann Ney. Improved backing-off for m-gram language modeling. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 1, pages 181-184, 1995.

- David A. McAllester, Robert E. Schapire: On the Convergence Rate of Good-Turing Estimators. *Proceedings of Conference on Learning Theory (COLT)* 2000: 1-6.

- Mehryar Mohri. Weighted Grammar Tools: the GRM Library. In *Robustness in Language and Speech Technology.* pages 165-186. Kluwer Academic Publishers, The Netherlands, 2001.

- Hermann Ney, Ute Essen, and Reinhard Kneser. 1994. On structuring probabilistic dependences in stochastic language modeling. *Computer Speech and Language*, 8:1-38.

References

- Kristie Seymore and Ronald Rosenfeld. Scalable backoff language models. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 1996.

- Andreas Stolcke. 1998. Entropy-based pruning of back-off language models. In *Proceedings of DARPA Broadcast News Transcription and Understanding Workshop*, pages 270-274.

- Ian H. Witten and Timothy C. Bell. The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression, *IEEE Transactions on Information Theory*, 37(4):1085-1094, 1991.