Phonological Constraints and Morphological Preprocessing for Grapheme-to-Phoneme Conversion

Vera Demberg[1], Helmut Schmid[2] and Gregor Möhler[3]

[1] School of Informatics, University of Edinburgh, UK
[2] Institut für Maschinelle Sprachverarbeitung (IMS), Universität Stuttgart, Germany
[3] IBM Research and Development GmbH, Böblingen, Germany

ACL 2007, Prague

Introduction

- Grapheme-to-Phoneme conversion (g2p):
  Sternanisöl → /ˈʃtɛrnʔaniːsʔøːl/ (Engl. ‘star anise oil’)
- Applications: component of TTS system
  e.g. in spoken dialogue systems, speech-to-speech translation
- For correct pronunciation we need:
  g2p, syllabification, stress assignment
- Question: Does morphology help g2p?
- Contributions of this paper:
  1. introduction of phonological constraints
     (for word stress and syllabification)
  2. evaluation of morphological preprocessing

Overview

1. Related Work
2. Method
   - Design
   - Evaluation
3. Phonological Constraints
   - Design
   - Evaluation
4. Morphological Preprocessing
   - Morphological Systems
   - Evaluation

Overview

1. Related Work
2. Method
   - Design
   - Evaluation
3. Phonological Constraints
   - Design
   - Evaluation
4. Morphological Preprocessing
   - Morphological Systems
   - Evaluation

Related Work

G2P conversion

- Decision Trees
  [Kienappel and Kneser, 2001, Black et al., 1998, van den Bosch et al., 1998]
- Pronunciation by Analogy [Marchand and Damper, 2000]
- HMMs [Taylor, 2005, Minker, 1996, Rentzepopoulos and Kokkinakis, 1991]
- Joint n-gram Models
  [Bisani and Ney, 2002, Galescu and Allen, 2001, Chen, 2003]

Relation to Syllabification and Stress Assignment

- (Perfect) syllabification helps g2p [Marchand and Damper, 2005]
- stress assignment and position of syllable [Müller, 2001]

Morphological Preprocessing

- claim: morphological information is important for g2p
  [Sproat, 1996, Möbius, 2001, Black et al., 1998, Taylor, 2005]
- but: never evaluated for German
- English: [van den Bosch, 1997]

Related Work

G2P conversion

- Decision Trees
  [Kienappel and Kneser, 2001, Black et al., 1998, van den Bosch et al., 1998]
- Pronunciation by Analogy [Marchand and Damper, 2000]
- HMMs [Taylor, 2005, Minker, 1996, Rentzepopoulos and Kokkinakis, 1991]
- Joint n-gram Models
  [Bisani and Ney, 2002, Galescu and Allen, 2001, Chen, 2003]

Relation to Syllabification and Stress Assignment

- (Perfect) syllabification helps g2p [Marchand and Damper, 2005]
- stress assignment and position of syllable [Müller, 2001]

Morphological Preprocessing

- claim: morphological information is important for g2p
  [Sproat, 1996, Möbius, 2001, Black et al., 1998, Taylor, 2005]
- but: never evaluated for German
- English: [van den Bosch, 1997]

Related Work

G2P conversion

- Decision Trees
  [Kienappel and Kneser, 2001, Black et al., 1998, van den Bosch et al., 1998]
- Pronunciation by Analogy [Marchand and Damper, 2000]
- HMMs [Taylor, 2005, Minker, 1996, Rentzepopoulos and Kokkinakis, 1991]
- Joint n-gram Models
  [Bisani and Ney, 2002, Galescu and Allen, 2001, Chen, 2003]

Relation to Syllabification and Stress Assignment

- (Perfect) syllabification helps g2p [Marchand and Damper, 2005]
- stress assignment and position of syllable [Müller, 2001]

Morphological Preprocessing

- claim: morphological information is important for g2p
  [Sproat, 1996, Möbius, 2001, Black et al., 1998, Taylor, 2005]
- but: never evaluated for German
- English: [van den Bosch, 1997]

Overview

1. Related Work
2. Method
   - Design
   - Evaluation
3. Phonological Constraints
   - Design
   - Evaluation
4. Morphological Preprocessing
   - Morphological Systems
   - Evaluation

Joint n-gram Model

$$\langle p;\hat{b};a\rangle_1^n = \arg\max_{\langle p;b;a\rangle_1^n} \prod_{i=1}^{n+1} P(\langle l;p;b;a\rangle_i \,|\, \langle l;p;b;a\rangle_{i-k}^{i-1})$$

l	letter
p	phoneme-sequence
b	syllable boundary
a	stress marker
k	context size

- **Goal**
  compute the most probable pronunciation $\langle p;\hat{b};a\rangle_1^n$ of a word given the word's orthographic form l_1^n
- **Alignment**
  1 letter $\rightarrow$ 0 - 2 phonemes, 1 syllable boundary flag, 1 stress marker

	R	ö	s	c	h	e	n	
/	r	ˈœː	s.	ç		ə	n.	/

- **Joint States**
  each state is a tuple $\langle l;p;b;a\rangle_i$
- Viterbi algorithm

Efficiency

State space very large:

- Each letter maps onto 12 different phonemes on average
- Working with 5-grams
- 12^5 = 250 k possible state sequences
- Smoothing with variant of Modified Kneser-Ney Smoothing

Peaked distribution:

- Pruning – consider only most probable states
- Threshold $t = 15$ best state sequences at a time (experiments: $5 < t < 35$)
- No significant difference in quality with respect to full state space
- $\approx$ 120 wds / min on 1.5 GHz machine

Results for Joint n-gram Model

- Joint n-gram model is competitive: similar to Pronunciation by Analogy (PbA), much better than decision trees
- Evaluation on phonemes only (stress / syllables not evaluated here)

language	corpus	# words	joint n-gram	PbA	decision tree
German	CELEX	230k	7.5%		15.0%
English	Nettalk	20k	35.4%	34.7%	
	a) auto. syll		35.3%	35.2%	
	b) man. syll		29.4%	28.3%	
English	TWB	18k	28.5%	28.2%	
English	beep	200k	14.3%	13.3%	
English	CELEX	100k	23.7%		31.7%
French	Brulex	27k	10.9%		

Table: G2P word error rates for different g2p conversion algorithms.

Overview

1. Related Work
2. Method
   - Design
   - Evaluation
3. Phonological Constraints
   - Design
   - Evaluation
4. Morphological Preprocessing
   - Morphological Systems
   - Evaluation

Phonological Constraints

Model $$\langle p;\hat{b};a\rangle_1^n = \arg\max_{\langle p;b;a\rangle_1^n} \prod_{i=1}^{n+1} P(\langle l;p;b;a\rangle_i \,|\, \langle l;p;b;a\rangle_{i-k}^{i-1})$$

Motivation (from conversions in German)

- many errors due to incorrect syllabification and stress assignment:
  - no syllable nucleus, or more than one (e.g. /ap.faːr.t/)
  - up to 20% words stressed incorrectly:
    (27% no stress, 37% $>$ 1 main stresses, 36% stress in wrong position)
- problems due to lack of context (just 5 letters seen at any time)

Introduce constraints

1. One nucleus per syllable
2. One (main) stress per word

Phonological Constraints

Model $$\langle p;\hat{b};a\rangle_1^n = \arg\max_{\langle p;b;a\rangle_1^n} \prod_{i=1}^{n+1} P(\langle l;p;b;a\rangle_i \mid \langle l;p;b;a\rangle_{i-k}^{i-1})$$

Motivation (from conversions in German)

- many errors due to incorrect syllabification and stress assignment:
  - no syllable nucleus, or more than one (e.g. /ap.faːr.t/)
  - up to 20% words stressed incorrectly: (27% no stress, 37% $>$ 1 main stresses, 36% stress in wrong position)
- problems due to lack of context (just 5 letters seen at any time)

Introduce constraints

1. One nucleus per syllable
2. One (main) stress per word

Implementation of Phonological Constraints

Goal: Find most probable phonemization that does not violate constraints.

Method 1:

- add flags *A* (accent precedes) and *N* (syllable contains nucleus) for current state
- splits each state into 4 new states
- probability 0 if e.g. *A* flag is set and a_i indicates 'stress'

$$P(\langle l;p;b;a\rangle_i \,|\, \langle l;p;b;a\rangle_{i-k}^{i-1}, A, N)$$

Method 2:

- enforce constraints by eliminating invalid transitions (modification of Viterbi algorithm)
- reduces data sparseness problem
- use transitional probabilities from old model without flags

Benefit of Integrating Constraints

The introduction of constraints decreases word error rates consistently and significantly.

		word error rates (WER)	
language	condition	no constraints	with constraint(s)
German	syllab.+stress+g2p	21.5%	13.7%
German	syllab. on letters	3.5%	3.1%
German	syllab. on phonemes	1.8%	1.5%
German	stress assignm. on letters	30.9%	9.9%
English	syllab.+g2p	40.5%	37.5%
English	syllab. on phonemes	12.7%	8.8%

Table: The table shows word error rates for German CELEX and English NetTalk.

Overview

1. Related Work
2. Method
   - Design
   - Evaluation
3. Phonological Constraints
   - Design
   - Evaluation
4. Morphological Preprocessing
   - Morphological Systems
   - Evaluation

Morphological Preprocessing

Pronunciation often depends on morphology:

- **Compounding**

  loophole: /ˈluːfəʊl/ vs. /ˈluːpˌhəʊl/ `1loop1hole`

- **Derivation**

  Röschen: /rœʃən/ vs. /rœːsçən/ `1Rös3chen`

- **Affixation**

  vertikal vs. *vertickern*: /v/ vs. /f/ `1vertikal, 4ver1tick3er2n`

  Weihungen vs. *Gen*: /ə/ vs. /eː/ `1Weih3ung2en, 1Gen`

Morphological Preprocessing

Pronunciation often depends on morphology:

- **Compounding**

  loophole: /ˈluːfəʊl/ vs. /ˈluːpˌhəʊl/ `1loop1hole`

- **Derivation**

  Röschen: /rœʃən/ vs. /rœːsçən/ `1Rös3chen`

- **Affixation**

  vertikal vs. *vertickern*: /v/ vs. /f/ `1vertikal, 4ver1tick3er2n`

  Weihungen vs. *Gen*: /ə/ vs. /eː/ `1Weih3ung2en, 1Gen`

Morphological Preprocessing

Pronunciation often depends on morphology:

- **Compounding**

  loophole: /ˈluːfəʊl/ vs. /ˈluːpˌhəʊl/ `1loop1hole`

- **Derivation**

  Röschen: /rœʃən/ vs. /rœːsçən/ `1Rös3chen`

- **Affixation**

  vertikal vs. *vertickern*: /v/ vs. /f/ `1vertikal, 4ver1tick3er2n`

  Weihungen vs. *Gen*: /ə/ vs. /eː/ `1Weih3ung2en, 1Gen`

Morphological Preprocessing

Pronunciation often depends on morphology:

- **Compounding**

  loophole: /ˈluːfəʊl/ vs. /ˈluːpˌhəʊl/ `1loop1hole`

- **Derivation**

  Röschen: /rœʃən/ vs. /rœːsçən/ `1Rös3chen`

- **Affixation**

  vertikal vs. *vertickern*: /v/ vs. /f/ `1vertikal, 4ver1tick3er2n`

  Weihungen vs. *Gen*: /ə/ vs. /eː/ `1Weih3ung2en, 1Gen`

Background: Methods for Morphological Segmentation

Morphological segmentation for German:

- Manual annotation
  - CELEX [Guide, 1995]
- Rule-based systems
  - SMOR [Schmid et al., 2004]
  - ETI[a]
- Unsupervised systems
  - [Demberg, 2007]
  - [Bernhard, 2006]
  - [Bordag, 2005]
  - [Keshava and Pitler, 2006]
  - [Creutz and Lagus, 2006]

System	F-Measure
CELEX	100%
SMOR	83.6%
ETI[a]	79.5%
Demberg	68.8%
Bernhard	63.5%
Bordag	61.4%
Keshava and Pitler	59.2%
Creutz and Lagus[b]	52.6%

[a] morphological component of TTS system from Eloquent Technology, Inc.

[b] Morfessor version 1.0

Benefit from Morphological Preprocessing

type	method	F-Measure wrt. CELEX	WER g2p	WER syllabification
unsuperv.	Keshava & Pitler	59.2%	15.1%	4.95%
–	no morphology		13.7%	3.10%
rule-based	ETI (rule-based)	79.5%	13.6%	2.63%
manual	CELEX	100%	13.2%	1.91%

- Insufficient quality of **unsupervised methods**: introduces additional errors instead of improving quality
- Morphological segmentations from **rule-based** system marginally improve g2p conversion with joint n-gram, and significantly improve syllabification
- **Perfect** morphological segmentation significantly improves both g2p conversion and syllabification

Morphology and g2p conversion algorithms

- What if we are using another g2p method? E.g. a decision tree?
- Effect of morphological preprocessing depends on g2p algorithm
- When an algorithm is used that performs less well (e.g. a decision tree), morphological segmentation has a larger positive effect
- Only one of the unsupervised algorithms improves performance of decision tree

		decision tree		joint n-gram	
type	morphology	PER	WER-ss	PER	WER-ss^{+}
unsuperv.	Keshava & Pitler	3.83%	28.3%		15.1%
–	**no morph.**	**3.63%**	**26.59%**	**2.52%**	**13.7%**
unsuperv.	Demberg	3.45%	26.09%		
rule-based	SMOR	3.00%	23.76%		
rule-based	ETI	2.8%	21.13%	2.53%	13.6%
manual	CELEX	2.64%	21.64%	2.36%	13.2%

Morphology and g2p conversion algorithms

- What if we are using another g2p method? E.g. a decision tree?
- Effect of morphological preprocessing depends on g2p algorithm
- When an algorithm is used that performs less well (e.g. a decision tree), morphological segmentation has a larger positive effect
- Only one of the unsupervised algorithms improves performance of decision tree

		decision tree		joint n-gram	
type	morphology	PER	WER-ss	PER	WER-ss$^+$
unsuperv.	Keshava & Pitler	3.83%	28.3%		15.1%
–	**no morph.**	**3.63%**	**26.59%**	**2.52%**	**13.7%**
unsuperv.	Demberg	3.45%	26.09%		
rule-based	SMOR	3.00%	23.76%		
rule-based	ETI	2.8%	21.13%	2.53%	13.6%
manual	CELEX	2.64%	21.64%	2.36%	13.2%

Morphology and g2p conversion algorithms

- What if we are using another g2p method? E.g. a decision tree?
- Effect of morphological preprocessing depends on g2p algorithm
- When an algorithm is used that performs less well (e.g. a decision tree), morphological segmentation has a larger positive effect
- Only one of the unsupervised algorithms improves performance of decision tree

		decision tree		joint n-gram	
type	morphology	PER	WER-ss	PER	WER-ss^{+}
unsuperv.	Keshava & Pitler	3.83%	28.3%		15.1%
–	**no morph.**	**3.63%**	**26.59%**	**2.52%**	**13.7%**
unsuperv.	Demberg	3.45%	26.09%		
rule-based	SMOR	3.00%	23.76%		
rule-based	ETI	2.8%	21.13%	2.53%	13.6%
manual	CELEX	2.64%	21.64%	2.36%	13.2%

Morphology and g2p conversion algorithms

- What if we are using another g2p method? E.g. a decision tree?
- Effect of morphological preprocessing depends on g2p algorithm
- When an algorithm is used that performs less well (e.g. a decision tree), morphological segmentation has a larger positive effect
- Only one of the unsupervised algorithms improves performance of decision tree

		decision tree		joint n-gram	
type	morphology	PER	WER-ss	PER	WER-ss^{+}
unsuperv.	Keshava & Pitler	3.83%	28.3%		15.1%
–	**no morph.**	**3.63%**	**26.59%**	**2.52%**	**13.7%**
unsuperv.	Demberg	3.45%	26.09%		
rule-based	SMOR	3.00%	23.76%		
rule-based	ETI	2.8%	21.13%	2.53%	13.6%
manual	CELEX	2.64%	21.64%	2.36%	13.2%

Morphology and g2p conversion algorithms

Interested in descriptions and results for unsupervised systems?

Wednesday, 13:30
Hall III (same room)

A Language-Independent Unsupervised Model for Morphological Segmentation

		decision tree		joint n-gram	
type	morphology	PER	WER-ss	PER	WER-ss$^+$
unsuperv.	Keshava & Pitler	3.83%	28.3%		15.1%
–	**no morph.**	**3.63%**	**26.59%**	**2.52%**	**13.7%**
unsuperv.	Demberg	3.45%	26.09%		
rule-based	SMOR	3.00%	23.76%		
rule-based	ETI	2.8%	21.13%	2.53%	13.6%
manual	CELEX	2.64%	21.64%	2.36%	13.2%

Other Results

Summary of other results from our work (refer to paper for more detail):

- **Data Sparseness**
  Morphology is more beneficial with little training data
- **Modularity**
  Better to do all steps in one model than separate models for g2p, syllabification and stress
- **Other Languages**
  Morphology not beneficial for English

Conclusions

- Integration of phonological constraints significantly improves grapheme-to-phoneme conversion
- Morphological segmentation can help g2p conversion and syllabification in German
- Whether it is worth to do morphological preprocessing depends on
  - g2p algorithm used
  - training set size
  - quality of morphological system (unsupervised systems not good enough)
  - language
- Best to do g2p conversion, syllabification and stress assignment in one module

Acknowledgments

Thank you:

- Hinrich Schütze
- Frank Keller
- reviewers

... and thanks to you for your attention!

Bernhard, D. (2006).
Unsupervised morphological segmentation based on segment predictability and word segments alignment.
In *Proceedings of 2nd Pascal Challenges Workshop*, pages 19–24, Venice, Italy.

Bisani, M. and Ney, H. (2002).
Investigations on joint multigram models for grapheme-to-phoneme conversion.
In *ICSLP*, pages 105–108.

Black, A., Lenzo, K., and Pagel, V. (1998).
Issues in building general letter to sound rules.
In *Third ESCA on Speech Synthesis*.

Bordag, S. (2005).
Unsupervised knowledge-free morpheme boundary detection.
In *Proceedings of RANLP 05*.

Chen, S. F. (2003).
Conditional and joint models for grapheme-to-phoneme conversion.
In *Eurospeech*.

Creutz, M. and Lagus, K. (2006).
Unsupervised models for morpheme segmentation and morphology learning.
In *ACM Transaction on Speech and Language Processing*.

Demberg, V. (2007).
A language-independent unsupervised model for morphological segmentation.
In *Proc. of ACL-2007*.

Galescu, L. and Allen, J. (2001).

Bi-directional conversion between graphemes and phonemes using a joint n-gram model.
In *Proc. of the 4th ISCA Workshop on Speech Synthesis*.

Guide, C. L. U. (1995).
Center for Lexical Information.
Max-Planck-Institut for Psycholinguistics, Nijmegen.

Keshava, S. and Pitler, E. (2006).
A simpler, intuitive approach to morpheme induction.
In *Proceedings of 2nd Pascal Challenges Workshop*, pages 31–35, Venice, Italy.

Kienappel, A. K. and Kneser, R. (2001).
Designing very compact decision trees for grapheme-to-phoneme transcription.
In *Eurospeech*, Scandinavia.

Marchand, Y. and Damper, R. I. (2000).
A multi-strategy approach to improving pronunciation by analogy.
In *Computational Linguistics*, volume 26, pages 195–219.

Marchand, Y. and Damper, R. I. (2005).
Can syllabification improve pronunciation by analogy of English?
Natural Language Engineering.

Minker, W. (1996).
Grapheme-to-phoneme conversion - an approach based on hidden markov models.

Möbius, B. (2001).
German and Multilingual Speech Synthesis.
phonetic AIMS, Arbeitspapiere des Instituts für Maschinelle Spachverarbeitung.

Müller, K. (2001).
Automatic detection of syllable boundaries combining the advantages of treebank and bracketed corpora training.
In *Proceedings of ACL*, pages 402–409.

Rentzepopoulos, P. and Kokkinakis, G. (1991).
Phoneme to grapheme conversion using HMM.
In *Eurospeech*, pages 797–800.

Schmid, H., Fitschen, A., and Heid, U. (2004).
SMOR: A German computational morphology covering derivation, composition and inflection.
In *Proc. of LREC*.

Sproat, R. (1996).
Multilingual text analysis for text-to-speech synthesis.
In *Proc. ICSLP '96*, Philadelphia, PA.

Taylor, P. (2005).
Hidden Markov models for grapheme to phoneme conversion.
In *INTERSPEECH*, pages 1973–1976, Lisbon, Portugal.

van den Bosch, A., Weijters, T., and Daelemans, W. (1998).
Modularity in inductive-learned word pronunciation systems.
In *Proceedings NeMLaP3/CoNNL98*, page 185194, Sydney.