Bootstrapping pronunciation models: a South African case study

Presented at the

CSIR Research and Innovation Conference

Marelie Davel & Etienne Barnard

27 February 2006

our future through science

Agenda

- Background
  - HLT in the developing world
  - Why pronunciation models?
- Bootstrapping & pronunciation modeling
- A Bootstrapping framework
  - Components
  - Efficiency
- Experimental approach
- Results
- Conclusions

© CSIR 2006 www.csir.co.za
our future through science

Background:
Human Language Technologies

- Speech processing:
  - Speech recognition, speech synthesis
  - Spoken dialogue systems, telephony systems
- Text-based language processing:
  - Search, information analysis, machine translation
- Human Factors in language-based systems:
  - System usability, culturally appropriate interfaces
  - System localisation

© CSIR 2006

Background:
HLT in the developing world

- Free and natural access
    - To information
    - To technology
- Reducing barriers
    - Literacy
    - Fluency in English
    - Technological literacy
    - Various types of disabilities
- Support for language diversity
- Support for service delivery

© CSIR 2006

Background:
HLT in the developing world

HLT requires extensive language resources:

- Electronic resources for local languages scarce
- Linguistic diversity high
- Skilled computational linguists scarce
- Language resource collection expensive

© CSIR 2006

Background:
Pronunciation modeling

- Ability to predict pronunciation based on written form of word

- Core component in speech processing systems:
  - Automatic speech recognition
  - Text-to-speech technology

- Example:
  - bright: b r ay t
  - girth: g er th

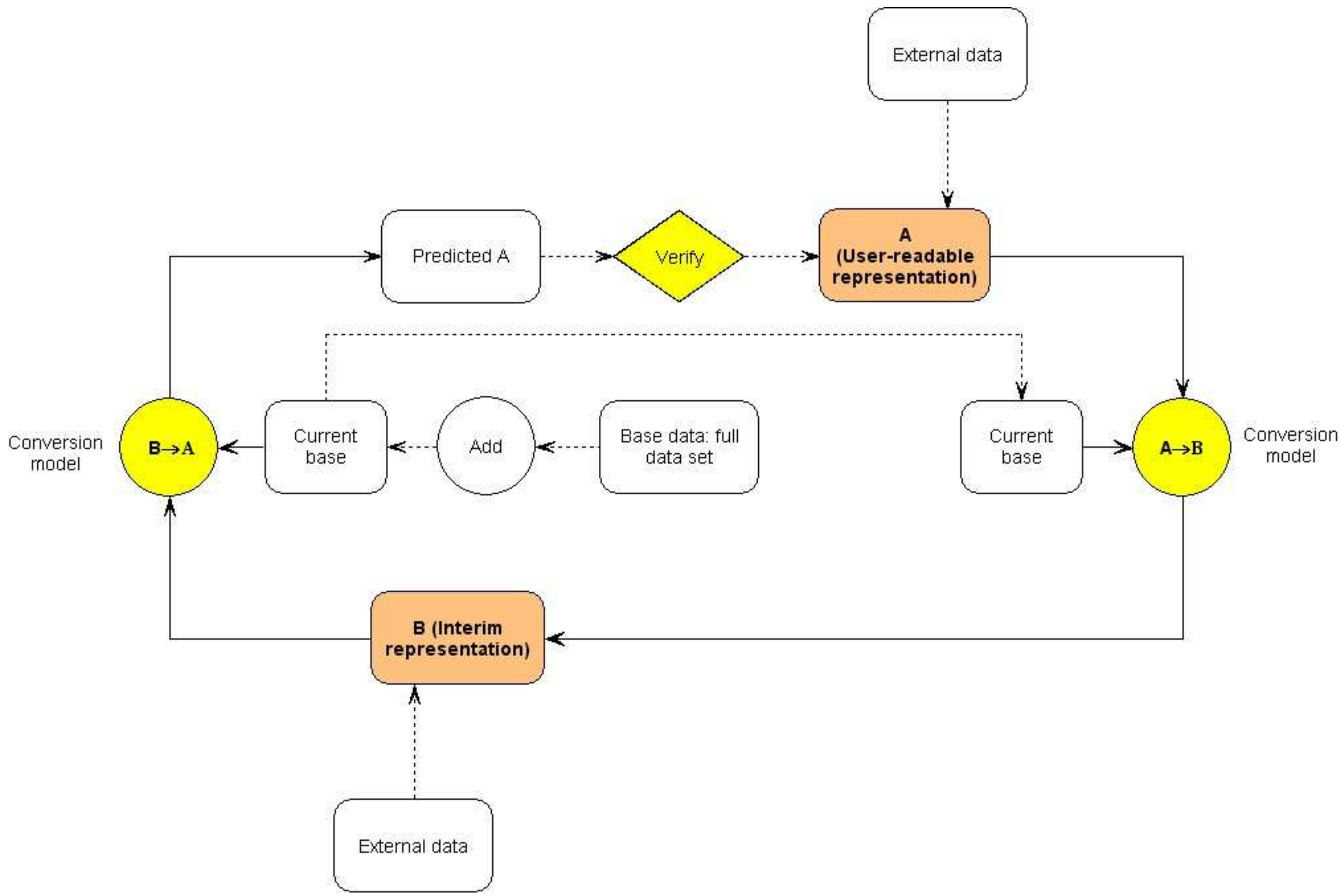
- Modeling pronunciations
  - Language-specific
  - Can use large pronunciation lexicons
  - Can learn from data

© CSIR 2006 www.csir.co.za

our future through science

Bootstrapping & Pronunciation modeling

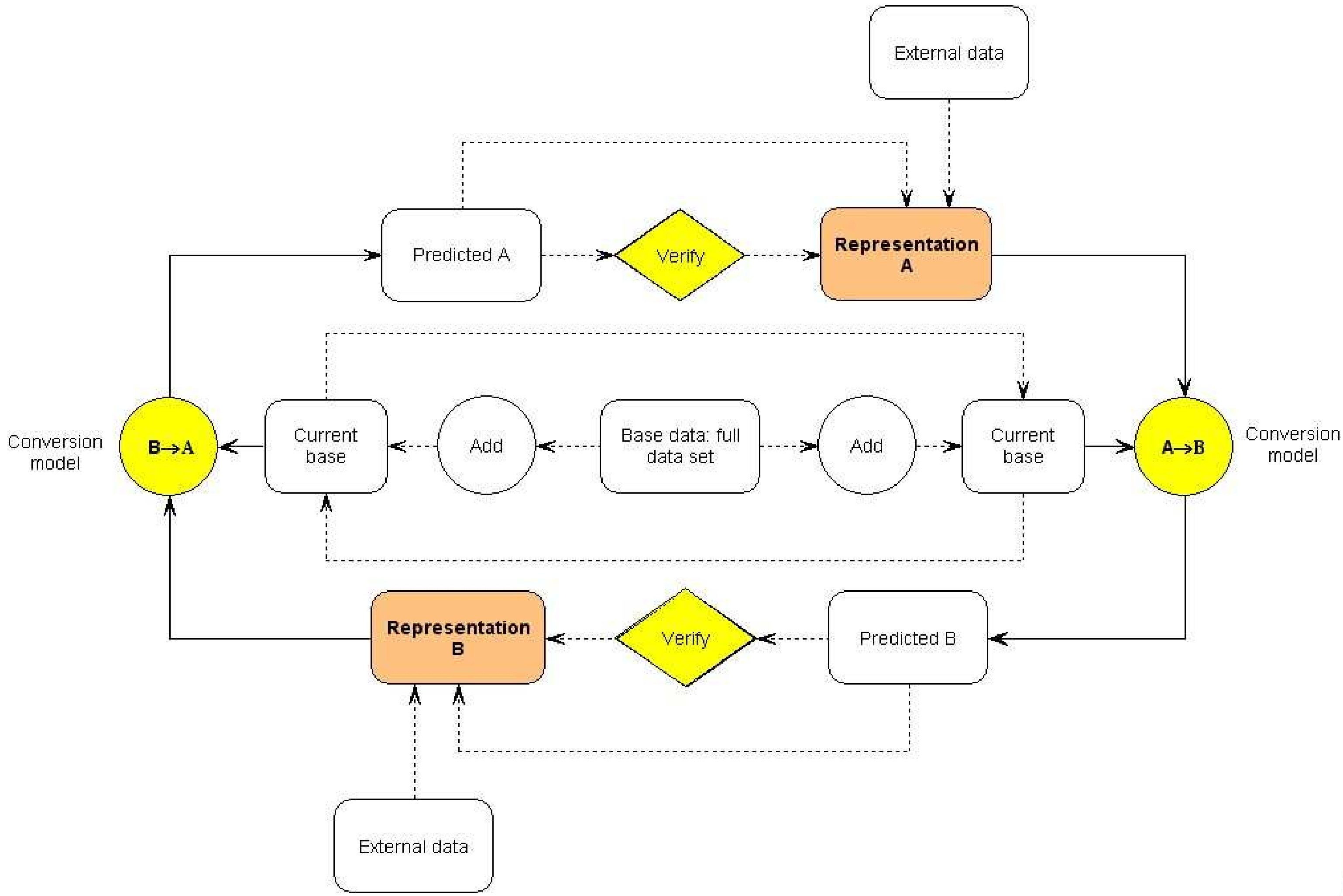
- Bootstrapping:
  Model improved iteratively
  Via a controlled series of increments
  Previous model utilised to generate next



Bootstrapping in action (Demonstration)

© CSIR 2006

Bootstrapping framework: Components



© CSIR 2006

our future through science

Bootstrapping framework: Efficiency

- Combine machine learning and human intervention, in order to minimise the amount of *human effort* required.

- Machine learning factors
  - Accuracy of representation
  - Conversion accuracy
  - Set sampling ability
  - System continuity
  - Robustness to human error
  - On-line conversion speed
  - Quality and cost of automated verification mechanisms
  - Validity of base data
  - Effect of incorporating additional resources

- Human Factors
  - Required user expertise
  - User learning curve
  - Cost of intervention
  - Task difficulty
  - Quality and cost of user verification mechanisms
  - Difficulty of manual task
  - Initial set-up cost

© CSIR 2006

Bootstrapping framework

- Prior work:
  - Demonstrated efficiency for small lexicons [1,2]
  - Developed new algorithms for efficient rule extraction [3,4]
  - Verified the human factors involved, including linguistic sophistication of user and implications of audio assistance [5]
  - Developed additional tools to support process, including automated error detection [6]

- This experiment:
  - Evaluate efficiency for a medium-sized lexicon: large enough for practical use

© CSIR 2006

Experimental approach

- Combine all prior results (each 1000 to 2000 words) to obtain a single 5000-word lexicon
- Bootstrap from 5000 to 8000 words, measuring actual effort
- Bootstrap parameters:
  - Linguistically sophisticated user
  - Incremental Default&Refine (synchronised every 50 words)
  - Automated error detection performed at end of cycle
  - Audio assistance optional

© CSIR 2006 www.csir.co.za
our future through science

Results

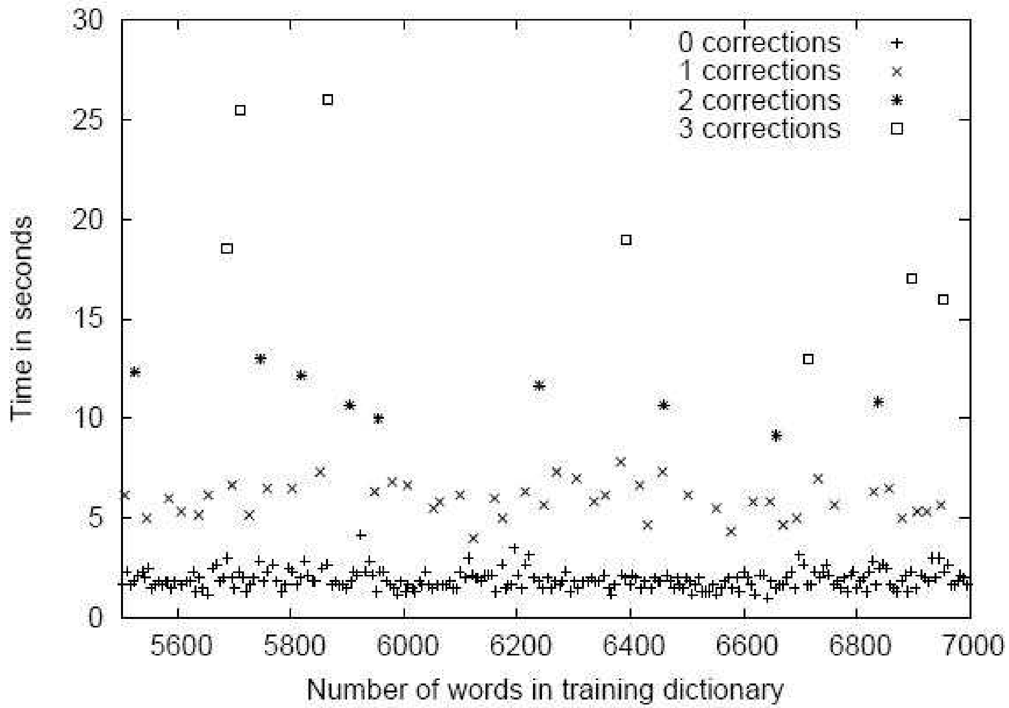


Figure 6.10: *Time taken to verify words requiring zero, one, two or three corrections, as a function of the number of words verified. For the first three measures, the averages were computed for blocks of 5 words each.*

© CSIR 2006

Results

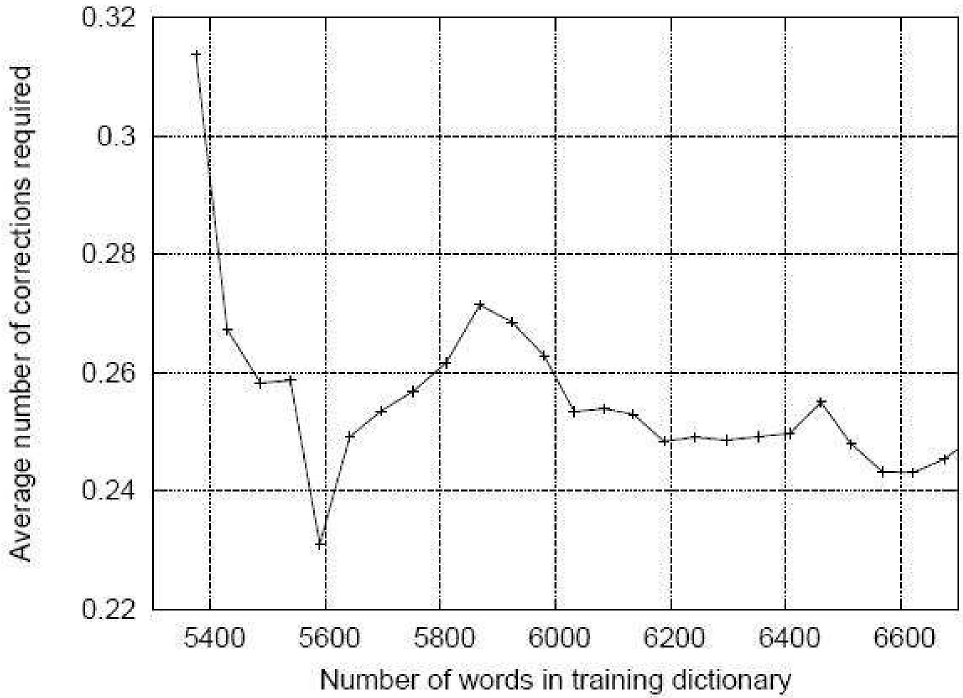


Figure 6.11: *The average number of corrections required as a function of the number of words verified. Averages were computed for blocks of 50 words each.*

© CSIR 2006

our future through science

Results

Table 6.3: *Typical observed values for various bootstrapping parameters.*

Bootstrapping parameter		Estimated value
Training cost	t_{train}	< 120 min
Verification cost for single words, with x corrections required for a word in state s:	$t_{verify(single,s)}$	$(2 + 4.5x)$ sec
Verification cost during error detection (per 1000 words):	$t_{verify(error-det)}$	< 10 min
Verification cost during error detection (per 400 words):	$t_{verify(error-det)}$	< 3 min
Task difficulty - bootstrapping, no error detection	$error\_rate_{bootstrap}$	$0\% - 1\%$
Task difficulty - bootstrapping, error detection	$error\_rate_{bootstrap}$	$0\% - 0.5\%$
Task difficulty - manual	$error\_rate_{manual}$	$0 - 0.5\%$
Manual development speed	$t_{develop}$	$19.2 - 30$ sec
Initial set-up cost	$t_{setup\_bootstrap}$ - t_{setup_manual}	< 60 min

© CSIR 2006 www.csir.co.za
our future through science

Results

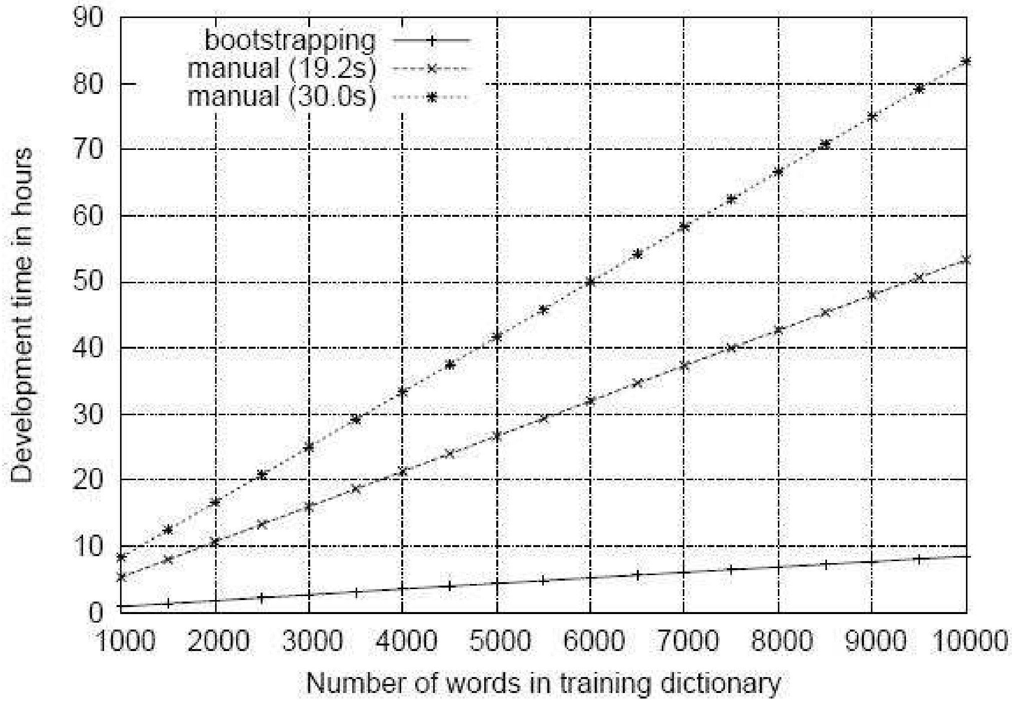


Figure 6.12: *Time estimates for creating different sized dictionaries. Manual development is illustrated for values of* $t_{develop}(1)$ *of* 19.2 *and* 30 *seconds, respectively.*

© CSIR 2006

www.csir.co.za

our future through science

Results

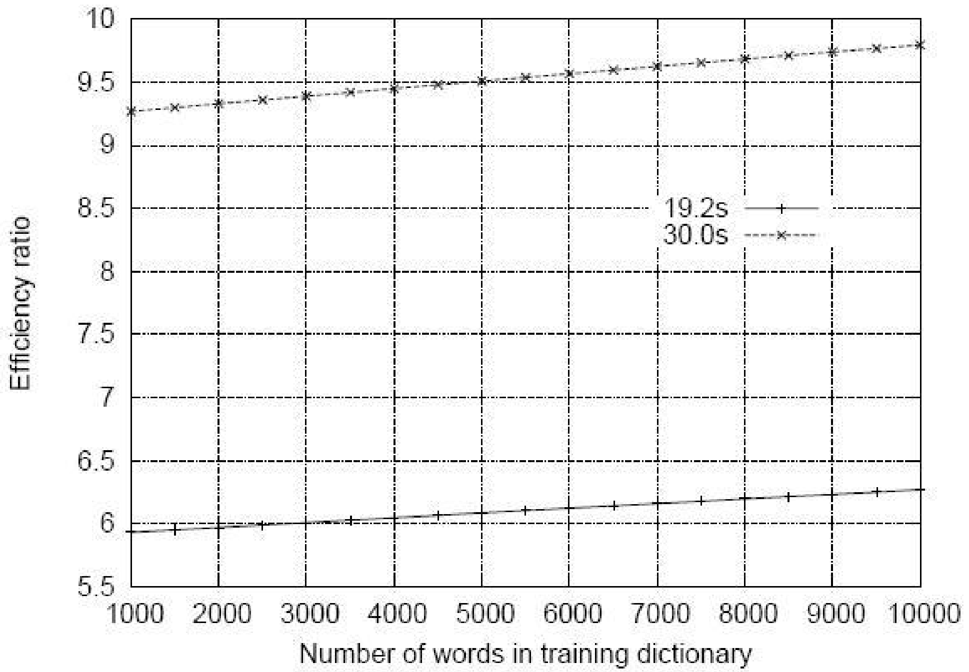


Figure 6.13: *Estimates of the efficiency of bootstrapping, as compared with manual development for values of* $t_{develop}(1)$ *of* 19.2 *and* 30 *seconds, respectively.*

© CSIR 2006

our future through science

Conclusions

- Dictionaries developed usable in practice
  - Afrikaans: general-purpose Text-to-Speech developed
  - isiZulu: general-purpose Text-to-Speech developed
  - Sepedi: automatic speech recognition system developed
- Approach practical and efficient
- Future work:
  - Open Source release imminent
  - Apply approach to all 11 official languages
  - Expand meta-information to be bootstrapped (including tone, stress)
  - Further algorithmic improvements
  - Evaluate implications of framework for additional resources

© CSIR 2006

our future through science

References

[1] M. Davel and E. Barnard, “Bootstrapping for language resource generation,” in *Proceedings of the Symposium of the Pattern Recognition Association of South Africa*, South Africa, 2003, pp. 97–100

[2] S. Maskey, L. Tomokiyo, and A.Black, “Bootstrapping phonetic lexicons for new languages,” in *Proceedings of Interspeech*, Jeju, Korea, October 2004, pp. 69–72.

[3] M. Davel and E. Barnard, “The efficient creation of pronunication dictionaries: machine learning factors in bootstrapping,” in *Proceedings of Interspeech*, Jeju, Korea, October 2004, pp. 2781–2784.

[4] M. Davel and E.Barnard, “A default-and-refinement approach to pronunciation prediction,” in *Proceedings of the Symposium of the Pattern Recognition Association of South Africa*, South Africa, November 2004, pp. 119–123.

[5] M. Davel and E. Barnard, “The efficient creation of pronunication dictionaries: human factors in bootstrapping,” in *Proceedings of Interspeech*, Jeju, Korea, October 2004, pp. 2797–2800.

[6] M. Davel and E. Barnard, “Bootstrapping pronunciation dictionaries: practical issues,” in *Proceedings of Interspeech*, Lisboa, Portugal, September 2005.

© CSIR 2006