Implementing a Syntax-Morphology Interface for Athabaskan

Emily M. Bender and Jeff Good
ebender@u.washington.edu and good@eva.mpg.de
University of Washington and MPI Leipzig
LSA, Oakland CA, 1/7/05

Overview

- *Introduction/goals*
- *Terminological distinctions*
- *Reject two possible interfaces*
- *Proposed interface design: run-time and development*
- *Conclusions*

... Illustrated with examples from Slave (Rice 1989)

Introduction: Montage

- *Suite of tools to assist in the documentation of underdescribed languages* (Bender et al 2004)
- *Focus on grammar (especially morphology and morphosyntax)*
- *Integrate with other initiatives building tools for transcribed texts and lexicons (e.g., ELAN, FIELD, AGTK)*

Some terminology

- *Morphophonology:*
  - *Morphotactics (e.g., position classes)*
  - *Morph.-conditioned phonological rules*
  - *General phonological rules*
  - *Mapping to abstract morphemes*
- *Morphosyntax:*
  - *Syntactic-semantic representations built from analysis of strings of abstract morphemes*

Introduction: Montage

- *Overarching goal: Allow the "ordinary working linguist" to make use of sophisticated grammar engineering tools without being grammar engineers themselves*
- *This talk: the Montage model for morphological analysis, and the morphology-syntax interface*

Possible interfaces

- *Morphophonology in morphosyntax*
- *Morphosyntax in morphophonology*
- *Independent morphophonology and morphosyntax*

Morphophonology in Morphosyntax

- *Morphosyntactic rules associated with morphophonological effects*
- *Standard in HPSG, perhaps most thoroughly worked out in Orgun 1996*
- *Assumed in current version of the LKB (Copestake 2002)*

Morphosyntax in Morphophonology

- *Interpret abstract morphemes as actual feature bundles*
- *Output of morphophonology is a lexical edge which can be used directly by the morphosyntactic parser*
- *Doesn't generalize to morphosyntactically complicated cases*

Morphophonology in Morphosyntax

- *Hard to reuse morphophonological work in morphosyntax*
- *Hard to push all morphophonology into one efficient machine*
- *Particularly awkward for strictly phonological effects*

Slave Morphological Causatives

hedenétį	hedné**h**tį
's/he fell asleep'	*'s/he put him/her to sleep'*

(Rice 1989:454)

- Syntactically and semantically, the causative form cannot be produced merely by adding features to the intransitive form.

Epenthesis in Slave

- *An epenthetic "peg element" is inserted before verb stems if they would not otherwise be preceded by some syllable in their word (Rice 1989:133)*
- **he**hjį *'I sing' vs.* nejį *'you sing'*

Theoretical conclusion

- *A computational system should allow morphophonology and morphosyntax to be modeled as independent, articulated systems*
- *The point of interface is the abstract morpheme*

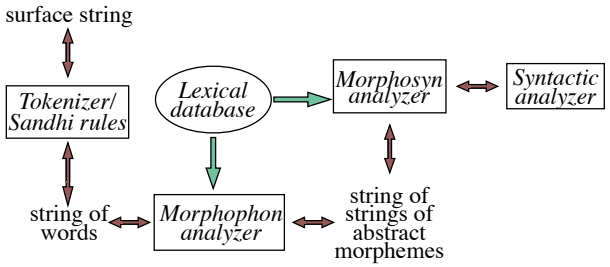
Independent Morphophonology and Morphosyntax

- *Morphophonology: maps surface forms to strings of abstract morphemes*
- *Morphosyntax: maps strings of abstract morphemes to syntactic/semantic information (feature structures)*

Development interface

- *Spell each underlying stem only once*
- *Define default morphotactic/syntactic pairings*
- *Allow multiple continuation classes for the same word sense and vice versa*

Runtime Interface

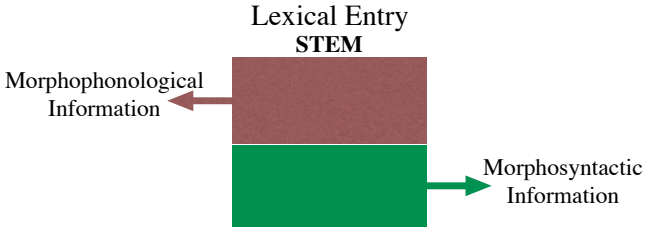


cf. Kaplan et al 2004, Siegel and Bender 2002

Slave Verb Classifiers

- *Verb classes based on the "classifiers" a verb theme contains* (Rice 1989:439–470)
  - *Ø-classifier,* **Ø**-ʔáh *'eat, chew'*
  - *h-classifier,* **h**-t'ó *'suck'*
  - *d-classifier,* **d**-shin *'sing'*
  - *l-classifier,* ná-**l**-séh *'hunt'*

Bipartite lexical database design



Build on Copestake et al 2004, FIELD

One morphosyntactic entry::many morphophonological entries

- *"Each classifier has a basic function, although they must be considered as part of the verb theme since this semantic content is not always clear."* (Rice 1989:453)
- *In some cases, verbs can alternate in their choice of classifier, with no non-morphophonological consequences* (Rice 1989:449–50)

One morphophonological entry::many morphosyntactic entries

- *Homophony within the same morphological class*
- *Multiple valence patterns, not predicted by a productive valence alternation*

Correlated morphophonological and morphosyntactic choices

- tę *'ice' (as a noun),* -tę*'freeze' (as a verb)* *(Rice 1989:161)*
- ká**tę**diįtse *'s/he broke through the ice'(incorporated noun)* *(Rice 1989:653)*
- *These stems will be associated with multiple morphophonological and morphosyntactic classes*
- *Handle correlation explicitly or implicitly*

Conclusions

- *Morphology and syntax are best treated as independent of one another*
- *Point of interface is abstract morphemes*
- *Two interfaces are required: run time and development*
- *Morphologically exuberant languages like Athabaskan are informative*

Acknowledgments

Thanks for helpful discussion to: Duane Blanchard, Anya Dormer, Scott Drellishak, Ann Gaponoff, David Goss-Grubbs, Jeremy Kahn, Bill McNeill, Matty Noble, Laurie Poulson

References

AGTK: Annotation Graph Toolkit. http://www.ldc.upenn.edu/Projects/AG/
ELAN: EUDICO Linguistic Annotator. http://www.mpi.nl/tools/elan.html
FIELD: Field Input Environment for Linguistic Data. http://emeld.org/tools/fieldinput.cfm
Grammar Matrix: Precision Grammar Starter Kit. http://www.delph-in.net/matrix/
LKB: LKB Grammar Development Environment. http://www.delph-in.net/lkb/
QLDB: Querying Linguistic Datbases. http://www.ldc.upenn.edu/Projects/QLDB/
XFST: Xerox Finite State Transducer. http://www.fsmbook.com/

Beesley, Kenneth R. and Lauri Karttunen. 2003. *Finite State Morphology*. Stanford: CSLI.

Bender, Emily M., Dan Flickinger, Jeff Good and Ivan A. Sag. 2004. Montage: Leveraging Advances in Grammar Engineering, Linguistic Ontologies, and Mark-up for the Documentation of Underdescribed Languages. *Proceedings of the Workshop on First Steps for the Documentation of Minority Languages: Computational Linguistic Tools for Morphology, Lexicon and Corpus Compilation, LREC 2004*, Lisbon, Portugal.

Bender, Emily M., Dan Flickinger and Stephan Oepen. 2002. The Grammar Matrix: An Open-Source Starter-Kit for the Rapid Development of Cross-Linguistically Consistent Broad-Coverage Precision Grammars. *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*. Taipei, Taiwan. pp. 8-14.

Copestake, Ann. 2002. *Implementing Typed Feature Structure Grammars*. Stanford: CSLI.

Copestake, Ann, Fabre Lambeau, Benjamin Waldron, Francis Bond, Dan Flickinger, and Stephan Oepen. 2004. A Lexicon Module for a Grammar Development Environment. *Proceedings of the 4th International Conference of Language Resources and Evaluation*. Lisbon, Portugal. pp. 1111-1114.

Good, Jeff. The descriptive grammar as a (meta)database. *Proceedings of the E-MELD Workshop 2004: Linguistic Databases and Best Practice*, July 15–18 2004, Detroit, Michigan.

Kaplan, Ronald M., John T. Maxwell III, Tracy Holloway King and Richard Crouch. 2004. Integrating Finite-state Technology with Deep LFG Grammars. *Proceedings of the Workshop on Combining Shallow and Deep Processing for NLP* (ESSLLI 2004).

Orgun, C. Orhan. Sign-based Morphology. UC Berkeley, Ph.D. Dissertation. Available as: ROA 171-0197.

Rice, Keren. 1989. *A grammar of Slave*. Berlin: Mouton.

Siegel, Melanie and Emily M. Bender. 2002. Efficient Deep Processing of Japanese. In *Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization at the 19th International Conference on Computational Linguistics*. Taipei, Taiwan