Matrix Developers Group: UW (and beyond) Site Update

DELPH-IN, Barcelona
July 20-24, 2009

Scott Drellishak

Added to the Matrix customization system:

- A library for case.
- A library for agreement (person, number, gender).
- A library for direct-inverse languages.
- Test cases for the above, including fragments of German, Hindi, Tagalog, Dyirbal, Cree, Fore, and Sahaptin.
- Enhancements to the questionnaire.

Presenting on Wednesday.

Laurie Poulson

Added to Matrix customization system an implementation of tense and aspect that:

- Provides a platform for defining tense and aspect elements
- Assumes a bipartite division of aspect but does not require it
- Assumes that tense and aspect are INDEX features of the verb's event variable

Working on improving lexical type hierarchy definition through the questionnaire.

- Current version requires a pain-staking hand definition of every node and leaf.
- New version will:
  - Create types based on features defined on the questionnaire.
  - Cross-classify appropriate types automatically.
  - Integrate with Scott's type inference code.

Antske: Word Order Library

Added to the Matrix customization system:

- Verb second main clauses
- V-complements for auxiliaries now works
- Verbal clusters (where all verbs must be adjacent) are formed for:
  - Verb-initial and verb-final languages
  - (Optionally) for VSO, OSV and free word order languages

Future work:

- Other auxiliary variations:
  - Auxiliaries that may precede or follow the verb
  - Allowing languages to have both auxiliaries that precede the verb, and auxiliaries that follow it:
- Adpositions
- Adjectives
- Ditransitive verbs

Presenting on Tuesday.

Adding an argument optionality library to the Matrix customization system. The library currently accounts for:

- Lexically licensed dropping.
- Argument affixes on verbs.
- Co-occurrence restrictions with an overt argument (all logical possibilities).
- Context restrictions (e.g. dropping only allowed in certain tenses/aspects/moods or person/number/gender)

Future work:

- Clitic argument markers, e.g., second place clitics as in Chemehuevi.
- Word order restrictions on affixes, e.g., mandatory absence of (some) affixes depending on the placement of an overt argument as in Abkhaz and Nkore-Kiga.

KEN

Working on Matrix customization validation. Building the Matrix Testing Database (MatrixTDB) for use with [incr tsdb()]. Tasks include:

- Functionalizing MatrixTDB.
- Improving performance of MatrixTDB.
- Filter writing for MatrixTDB.
- Modifying MatrixTDB to handle newer customization system features.
- Documenting the system and the code.

Mike Goodman

DELPH-IN related work at NICT and UW:

- Updated the Tanaka Corpus (with user contributed corrections).
- Created Egad (Erroneous Generation Analysis and Detection): using generation for error detection.
- Used Egad to fix bugs and inadequacies in Jacy bringing generation coverage up nearly 20%.

Presenting today.

Future work on the customization system:

- Improving the speed of the unit tests.
- Re-factoring the lexical rules code.
- Bug fixes and cross-library improvements.

Joshua Hou

- Created web interface for lexical acquisition:
  - User enters a word.
  - Word is presented to user in several templatic (fill in the blank) sentences.
  - User selects the appropriate context for the word (template).
  - The appropriate lexical type is inferred for the lexeme based on the context selected.
- Lexical acquisition for increased coverage of small Lillooet and English grammars.
- Acquisition of both lexemes and lexical types.
- Required hand-constructed template sentences for the elicitation of lexical type.
- Designed for use by non-linguist native speakers.
- (Not yet publicly accessible.)

Bill McNeill

Working on: The effect of feature selection on HPSG parse ranking across languages.

- Using the logon batch experiment system.
- Running parse ranking experiments for Japanese and English.
- Measuring the effect of changes in maximum entropy feature sets on accuracy.
- Looking to see whether different sets of features work better with the different languages, or if the same set of features works for both.

Presenting on Tuesday.

Emily Bender: Wambaya

- Exploring replacing argument composition analysis with set of auxiliary + verb cluster constructions.
- Maintaining two separate branches of the grammar, and striving for equivalent competence (nearly there).
- Verb cluster approach requires four additional constructions
- Interesting interactions with subordinate clauses, reflexive/reciprocal construction, and a few others.

Wambaya

Haven't tried generation yet, but massive speed up in parsing:

Aggregate	(g)old			new			reduction		
	tasks ϕ	time ϕ (s)	space ϕ (kb)	tasks ϕ	time ϕ (s)	space ϕ (kb)	tasks %	time %	space %
$10 \leq$ *i-length* < 15	157029	2.66	48107	34548	3.57	65829	78.0	-34.2	-36.8
$5 \leq$ *i-length* < 10	27709	0.78	13456	3331	0.03	6718	88.0	95.7	50.1
$0 \leq$ *i-length* < 5	2167	0.02	4924	920	0.01	4602	57.5	61.6	6.5
Total	**6772**	**0.15**	**6451**	**1423**	**0.02**	**5124**	**79.0**	**86.3**	**20.6**

(generated by [incr tsdb()] at 26-jun-2009 (11:41 h))

Paper available upon request!

Cyberling Workshop 2009

- Working towards bringing computational methods to mainstream linguistics.
  - Improve cyber-infrastructure
  - Promote culture of data sharing
  - Promote culture of data using (ahem)
- Follow on to LSA symposium in January.
- Workshop at Berkeley last weekend, to bring together people from diverse subfields and viewpoints interested in these issues.
- The collaboration continues online:
  `http://cyberling.elanguage.net`

Please Cite This Paper: The Grammar Matrix Standard Project Reference

- Writing up current state of the Grammar Matrix customization system
- Evaluation over a sample of 7 typologically and geographically diverse languages
- Reflecting on what we've learned about grammar customization