Grammar Checking in the Arboretum: Finding and Curing Trees

Emily M. Bender
(with Dan Flickinger & Stephan Oepen)

Supported by Stanford University's MediaX Program on Learning and Training

Grammar checking for CALL: Requirements

- Robustness when faced with malformed input
- Detection of grammatical errors
- Localization of grammatical errors
- Description of grammatical errors
- Ability to suggest correct form
- Natural language understanding
- Reusability across different native languages

Applications

- Self-directed language learning exercises
    - Structured tasks
    - Open ended conversation, within some domain
- Evaluation tools for ESL teachers to measure student progress
- Automated testing of language skills, e.g improving TOEFL

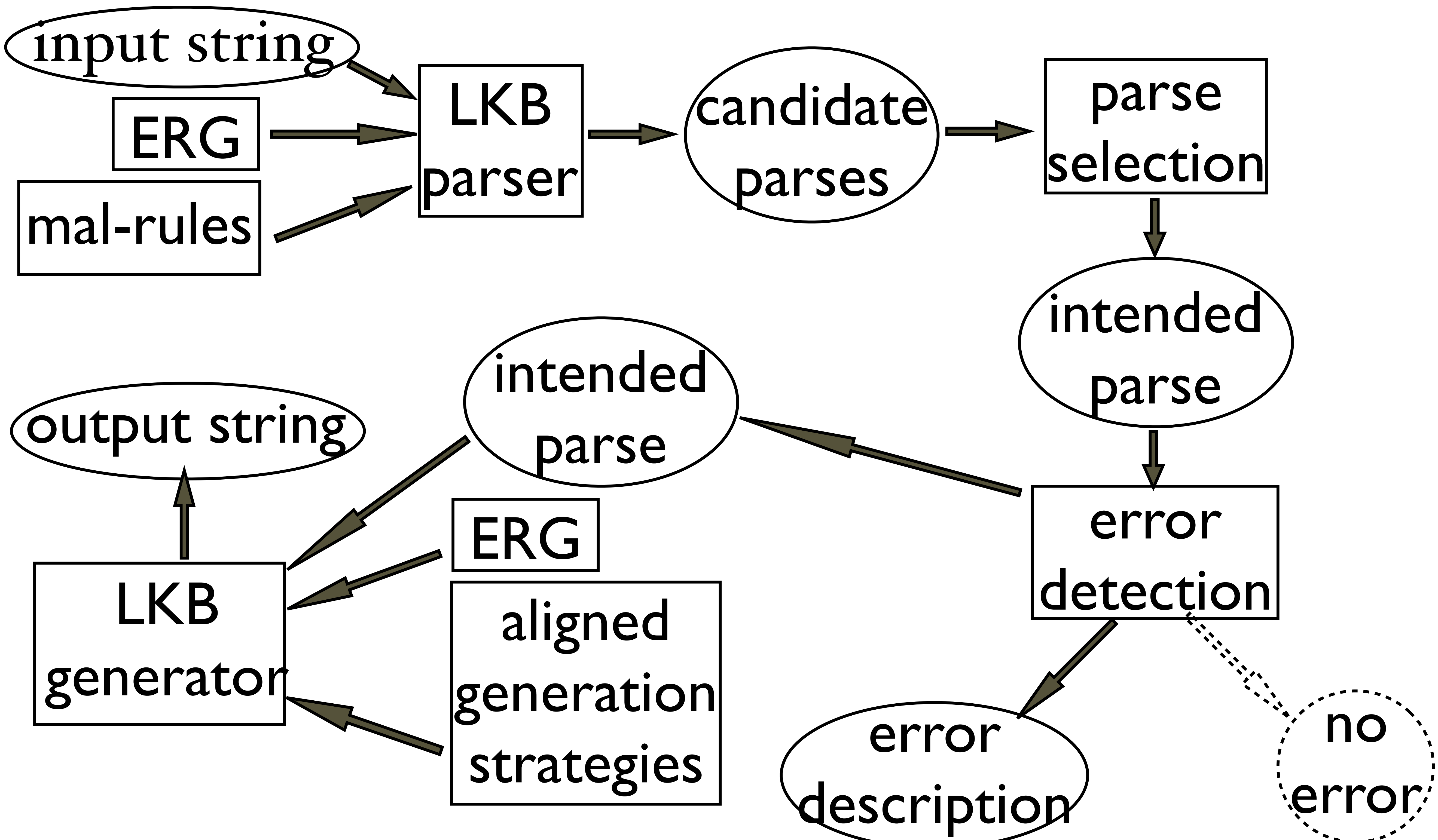
Resources

- LKB (Copestake 2002): A parser (Malouf et al 2000) and generator (Carroll et al 1999) for typed feature structure grammars.
- ERG (Flickinger 2000): A broad-coverage precision HPSG for English, suitable for parsing, generation, and natural language understanding.
- Redwoods (Toutanova et al 2002): Parse ranking techniques based on a rich, dynamic treebank.

Strategy

- Augment ERG with mal-rules, relating malformed input to well-formed semantics
- Parse with mal-rules, generate without.
- Treat correction as a simple kind of semantics-based machine translation task.
- Diagnose errors based on mal-rules used.
- Semantic representations potentially serve as input to a dialogue system.

System overview



Sample mal-rule

mal_bare_np_sg_phrase := *generic_bare_np_phrase* &

$$\begin{bmatrix} \text{C-CONT.RELS} & \langle ! \begin{bmatrix} \text{PRED} & \textit{mal\_bare\_div\_q\_rel} \end{bmatrix} ! \rangle \\ \text{ARGS} & \langle \begin{bmatrix} \text{SYNSEM.LOCAL.AGR.DIVISIBLE} & - \end{bmatrix} \rangle \\ \text{ROBUST} & + \end{bmatrix}$$

NP

N′[DIVISIBLE -]

dog

mal_bare_div_q_rel(x) dog(x)

cf:

_the_q_rel(x) dog(x)

More mal-rules

- Assign 3sg agreement to non3sg forms (e.g., *talk*) and vice versa
- Allow main verbs to invert with their subjects
- Allow main verbs to precede sentence-negating *not*

Mal lexical types and Mal lexical entries

- Verbs like *allow* with infinitival rather than gerund complements:
  - *We allow to sleep* (cf. *We allow sleeping*)
- Verbs like *want* with bare infinitival complements:
  - *We want run* (cf. *We want to run*)

Error detection & diagnosis

- Assume we can find the preferred parse
- Check whether the highest node in the preferred parse is [ROBUST +]
- Extract rules/lexical entries licensing each node and collect those that are [ROBUST +]
- Look up error description corresponding to each robust rule or lexical entry

Ambiguity in generation

- Multiple output strings for one input semantics
- Due to semantically vacuous syntactic choices (topicalization, *that*-deletion, *do*-insertion, ...)

Aligned generation

- Corpus-based string selection is inappropriate.
- Instead, align output of generator to choices made in the input parse
- Best-first generation with a quasi-stochastic ranking strategy, using the single input parse as the sole source of evidence
- Give priority to the creation of specific edges in a bottom-up chart generator

Aligned generation strategies

- When considering adding an edge (local subtree) to the agenda:
  - Is same configuration of rules (within subtree) found in the input parse? → 100
  - Is the same rule with the same lexical yield found in the input parse? → 80
  - Is the same rule found anywhere in the input parse? → 60

Evaluation

- Over a small test suite of well-formed input (107 sentences), the configuration strategy always returned the input parse, when supported by the grammar (87/107).
- However, the configuration strategy alone is insufficient with malformed input:
  - *The dog chase the cat > The cat the dog chases*
  - *Dog wants to know cat arrive > A/the cat a/the dog wants to know arrives.*

Evaluation

- Adding either of the remaining two strategies will help some, but neither alone is sufficient:
  - *We want know cat chase dog* >
    - *A/the dog we want to know a/the cat chases*
    - *We want to know that a/the cat chases a/the dog*
- All three together will get the right result:
  - *We want to know a/the cat chases a/the dog*

Evaluation

- Some cases are beyond even all three strategies:
  - *Green blue red cat slept > A/the blue red green cat slept.*
- Suggests the need for a configuration + yield strategy as the first test.
- More systematic evaluation awaits a broader range of mal-rules and a test suite based on naturally occurring data.

Future work: Building out the mal-rules

- Prioritize based on a corpus of learner English, preferably error-tagged.
- Evaluate degree to which mal-rules developed for one L1 group apply to another.
- Explore the extent of constraints imposed by strict semantic compositionality.

Future work: Parse selection

- The mal-rules will normalize malformed input to well-formed semantics.
- Try ranking based on dependencies (derived from semantic representations), and training on a well-formed corpus.
- Success will depend on how close the semantic representations produced match those derived from analogous correct sentences.

Related work

- ICICLE: Interactive Computer Identification and Correction of Language Errors (Michaud & McCoy 2003, www.eeics.udel.edu/research/icicle/)
- Menzel & Schröder 1999: 'Error diagnosis for language learning systems'

Conclusion

- Precise, deep NLP using grammars like the ERG raises the possibility of automated language tutors that can both keep a conversation going and correct errors.
- Aligned generation may have applications beyond the current CALL project.

References

- Carroll, J., A. Copestake, D. Flickinger, & V. Poznanski. 1999. 'An Efficient Chart Generator for (Semi-)Lexicalist Grammars. *Proceedings of the 7th European Workshop on Natural Language Generation (EWNLG'99)*, Toulouse: 86-95.

- Copestake, A. 2002. *Implementing Typed Feature Structure Grammars.* Stanford: CSLI Publications.

- Flickinger, D. 2000. 'On Building a More Efficient Grammar Using Types.' *Natural Language Engineering* 6(1):15-28.

- Malouf, R., J. Carroll, & A. Copestake. 2000. 'Efficient Features Structure Operations without Compilation.' *Natural Language Engineering* 6(1):29-46.

References

- Menzel, W. & I. Schröder. 1999. 'Error Diagnosis for Language Learning Systems.' *ReCALL* May 1999:20-30.
- Michaud, L. & K. McCoy. 2003. 'Evaluating a Model to Disambiguate Natural Language Parses on the Basis of User Language Proficiency.' *Proceedings of the 9th International Conference on User Modeling.*
- Toutanova, K., C. Manning & S. Oepen. 'Parse Ranking for a Rich HPSG Grammar.' *Proceedings of the First Workshop on Treebanks and Linguistic Theories (TLT2002)*, Sozopol, Bulgaria.