Diagnosing Word Order Errors

Adriane Boyd and Detmar Meurers
The Ohio State University and Universität Tübingen

Berlin. October 16, 2009

Background

Word order and Foreign Language Teaching

- It is hard to learn word order:
  - Language learners are known to produce a range of word order errors (cf., e.g., Odlin 1989).
  - Word order differs significantly across languages → transfer errors (cf., e.g., Selinker 1972; Odlin 2003)

- It is important to master word order, especially since word order errors can significantly complicate comprehension.
  - Example from Hiroshima English Learners’ Corpus:
    1. *He get to cleaned his son.*
    → *He get his son to cleaned.*
  - Exercise target:
    2. *He made his son clean the room.*

ICALL research has largely focused on diagnosing word-based learner errors (i.e., morpho-syntax).

Such approaches can rely on parsing algorithms to reign in the recursive potential of natural language.

How about word order mistakes, a type of error regularly produced by language learners?
Approaches to diagnosing word order errors

Deep processing: Basics

- Use grammars, which are compact representations of the wide range of lexical and word order possibilities.
- Efficient parsing algorithms are available to license a potentially infinite set of strings based on finite grammars.
- The additional erroneous word orders can be captured by:
  - extra phrase structure rules (so-called mal-rules, cf. e.g., Heiß 1998; Fortmann & Forst 2004)
  - manipulation of chart edges, the hypotheses introduced by phrase structure rules in a chart parser (Reuer 2003)

Our perspective and approach

- Word order errors are not uniform:
  - some involve lexical triggers (one of a finite set of words is known to occur) or indicative patterns,
  - others can only be spotted with deeper analysis.
- FLT activities are not uniform:
  - some can be set up to include specific lexical material or patterns,
  - in others it is hard to control lexical and structural variation.
⇒ Activity-based ICALL systems need a flexible approach to word order error detection and diagnosis.

Two types of word order errors

- We explore two aspects of English grammar with interesting word order properties:
  - phrasal verbs
  - adverbs
- For each, we describe
  - linguistic properties,
  - exercises supporting awareness of the relevant word order patterns, and
  - the processing needed for those exercises.
Phrasal verbs

- Separable phrasal verbs
  - Particles can precede or follow a full NP object.
    1. a. * wrote down the number
    2. b. wrote the number down
  - Particles must follow a pronominal NP object.
    3. a. * wrote down it
    4. b. wrote it down

- Inseparable phrasal verbs
  - Particles always precede any NP object.
    5. a. ran into {my neighbor, her}
    6. b. * ran {my neighbor, her} into

Phrasal verbs

Pedagogical relevance of particle placement

- English learners make errors in particle placement:
  7. a. * so they give up it
  8. b. * food which will build up him
  9. c. * rather than speed up it

Examples from the Chinese Learner English Corpus (CLEC 2004)

- Learners also avoid using phrasal verbs:
  - Liao & Fukuya (2002) show that Chinese learners of English avoid phrasal verbs; similar research for other L1.
  - We also found patterns of avoidance in the CLEC:
    - heavy use of pattern that is always grammatical
    - little use of patterns restricted to certain verb & object types

Phrasal verbs

Example exercise tasks

Part 1 of the exercise targets lexical particle choice:

Complete the following sentence:

Please turn the radio _____ a little. It's too loud.

Part 2 targets particle placement (and pronoun choice).

Now, replace the object with a pronoun:

Please turn down the radio a little. It's too loud.

→ Please ______________ a little. It's too loud.

Phrasal verbs

Processing the example exercises

- We target two possible error patterns:
  - separable-phrasal-verb < particle < pronoun
    8. a. * wrote down it
  - inseparable-phrasal-verb < NP < particle
    9. a. * ran my neighbor into
    10. b. * ran her into

- Regular expression matching with those patterns is sufficient to capture the targeted errors.
  - The relevant words (or strings) to be matched are specified in the activity model.
  - Desired error diagnosis and feedback is one-to-one with those patterns.

⇒ Particle placement is an example for a word order phenomenon which can adequately be diagnosed based on a shallow analysis.
Adverb placement in English

- English has many different adverbs, and the word order possibilities depend on adverb subclass distinctions.
- The rules governing adverb placement are difficult to articulate and master.
- Many adverb placements are not right or wrong, but more or less natural.
- Students frequently misplace adverbs

(10) a. they cannot already live without the dope.
   b. There have been already several campaigns held by `Outdoor`.
   c. while any covert action brings rarely such negative connotations.
   d. It seems that the Earth has still a lot to reveal . . .

Examples from Polish part of Int. Corpus of Leaner English (PICLE 2004)

Adverb placement

- Processing the example exercises
  - Instance-based matching is inadequate:
    - Many placements throughout a sentence are possible.
    - Targeted errors are predictable, but numerous.
    - Generalizations about the many adverbs of English and the subclasses they form are lost.
    - Reference to syntactic structure is needed for
      - identification of possible placements,
      - error diagnosis, and
      - content of feedback.
  - Deep processing
    - Parsing can identify the necessary sentence structure.
    - The lexicon of a grammar supports modeling adverb classes.

- Deep processing
  - Parsing can identify the necessary sentence structure.
  - The lexicon of a grammar supports modeling adverb classes.

Example exercise tasks

Task 1:

Find and move any misplaced adverbs:

(11) She has finished almost her breakfast.

Task 2:

Add the given adverb to the sentence:

Adverb: slowly

(12) Taking his visitor by the arm, he walked her along the corridor.

(Example taken from British National Corpus)

Adverb placement

Combining native and interlanguage patterns

- We need to model a learner grammar which combines
  - native English patterns with
  - anticipated interlanguage patterns.
- Word orders not licensed by the space between native and interlanguage patterns should be excluded, to support efficient processing.
- The combination of native and interlanguage patterns should not result in spurious ambiguities (i.e., same word order licensed by different structures).
Adverb placement

Targeted word orders

> Adverb placement can be described in terms of linear order with respect to constituents.

(13) Sid 2 might 3 be 4 taking 5 his mother 6 to the store 7.
1. clause-initial
2. preceding a finite auxiliary
3. preceding a nonfinite auxiliary
4. preceding a main verb
5. preceding an NP complement
6. preceding a PP complement
7. following the VP

> This is the basic picture; the situation is more complex in the presence of negative auxiliaries or passive sentences.

> For each adverb subclass, we rate the positions in terms of acceptability (good, bad, marked).

Adverb placement encoding in the prototype

The lexical principle constraining and recording adverb position

(word, synsem:head: (adv, mod:synsem))

fun adv_placement(+,+,-).
adv_placement(@clause, minus, pre_clause) if true.
adv_placement(@fin_aux, minus, pre_finite_aux) if true.
adv_placement(@nfin_aux, minus, pre_nonfinite_aux) if true.
adv_placement(@main_vp, minus, pre_main_verb) if true.
adv_placement(@np_comp, plus, pre_np_comp) if true.
adv_placement(@pp_comp, plus, pre_pp_comp) if true.
adv_placement(@fin_vp, plus, post_finite vp) if true.

Deep processing in prototype

> In the implemented prototype, we parse sentences with all envisaged adverb placements, using an HPSG grammar implemented in the TRALE system (MILCA environment; Meurers, Penn & Richter 2002).

> We encode the actual adverb position through the value of two features in the lexical entry of the adverb:
  - MOD: what category the adverb combines with
  - POSTHEAD: whether the adverb occurs before/after the head

> The lexical subclass of the adverb and its position is passed up and encoded as part of the overall structure, where it can inform negative or positive feedback.

Adverb placement and beyond

> Adverb position is constrained and recorded using a lexical principle, i.e., not in terms of a local tree.
  - Such lexicalization is appropriate for words which are fixed by the activity model.
  - Phrases (e.g., NPs) not targeted by an activity can be pre-processed by a chunker/supertagger to keep a limited lexicon across a range of contextualized activities.

> Argument reordering encoded parallel to optional complement selection in MERGE (Meurers et al. 2003).

> Outlook:
  - For local tree-based word order phenomena (e.g., SOV \(\rightarrow\) VOS) mal-rules can be used.
  - For other word order phenomena, a formalism that supports word order domains beyond local trees (e.g., GIDLP, Daniels & Meurers 2004) can be used.
Summary

- When to use instance-based matching:
  - lexical material and erroneous placements are predictable and listable
  - there is no grammatical variation
  - error patterns correspond directly to intended feedback
- When deep processing is preferable:
  - possible correct answers are predictable but not (conveniently) listable for a given activity
  - predictable erroneous placements occur throughout a recursively built structure
  - feedback is desired which requires linguistic information about the learner input that can only be obtained through deep analysis
- Lexicalization of word order options can be an attractive, modular alternative to mal-rule based encodings.

References


