

# Automatically Identifying and Correcting Errors in Learner Writing using a Word Cluster Approach

Laurence Anthony

Center for English Language Education in Science and Engineering (CELESE)

School of Science and Engineering  
Waseda University

<http://www.antlab.sci.waseda.ac.jp/>

1

## Outline

- Background
  - Usefulness of error correction
  - Types of errors in learner writing
  - Automatic error identification and correction methods
- AntError: A novel, automatic error identification and correction tool
  - Design and Methodology
  - Results
- Summary & Future Work

2

## Background

- Is error correction useful for learners?
  - Error correction does not lead to improved writing
    - Truscott, J (1996); Loewen, S (1998)
  - Error correction is harmful
    - Truscott, J (1996)

3

## Background

- Is error correction useful for learners?
  - Teachers are not consistent or systematic in giving feedback on errors
    - Zamel, V. (1985)
  - Teachers may comment on problems that are not the actual problems of the learners.
    - Cohen, A.D. & Cavalranti, M.C.(1990)

4

## Background

- Is error correction useful for learners?
  - Error correction for second language students is useful
    - Aljaarich, A. & Lantolf, J. (1994); Myles, J. (2002)
  - Error feedback helps to reduce errors at a faster rate than no feedback
    - Chen, F. J. (1997)
  - Grammar correction 'raises consciousness' towards errors
    - Healey, D. (1992)

5

## Background

- Types of errors found in student writing
  - lexical (e.g. What do you lik?)
  - syntactic (e.g. What does you like?)
  - pragmatic (e.g. What do you like sports?)
- Sentence level errors (Beare, K, 2004)
  - tense
  - punctuation
  - word order
  - wrong word
  - missing word
  - spelling
  - grammar

6

## Background

- Automatic Error Identification and Correction
  - Spelling and grammar checkers
    - Liou, H. (1991; 1992; 1993; 1994)
    - Hong, W, & Davies, G. (1997)
    - Yao, Y. & Warden, C. A. C. (1996)
  - Grammar checkers unable to deal with complex grammatical and stylistic problems
    - Hong, W, & Davies, G. (1997)
  - Most programs use 'drill & kill' exercises
    - Chen, J. F. & Warden, C. A. C. (1997)

7

## Background

- Automatic Error Identification and Correction
  - Software for checking key phrases used in writing
    - e.g. Jamieson, et al. (1993)
  - Problems:
    - All key phrases have to be added to the system by hand
    - Domain specific

8

## Background

- Automatic Error Identification and Correction
  - EDEN: Sentence level error correction using a database of sentences with error tags
    - Saiga et al. (2003)
  - Problems:
    - Requires list of error tags
    - Requires error tagging of the database
    - Domain specific

9

## “AntError”

- A novel, automatic error identification and correction tool
  - Works at the sentence level
  - Does not require a pre-defined set of key-phrases or error tags
  - Does not require error tagged data
  - Domain independent

10

## “AntError”

- Design
  - Redefine the error identification problem
    - Do not try to identify all possible errors in an learner's sentence
    - Identify differences (errors) between a learner's sentence and candidate sentences from a set of template “correct” sentences
  - Disadvantages:
    - Not suitable for “free writing” exercises, where template “correct” sentences are unavailable
      - essay writing, diaries, unrestricted QA exercises

11

## “AntError”

- Design
  - Redefine the error identification problem
    - Do not try to identify all possible errors in an learner's sentence
    - Identify differences (errors) between a learner's sentence and candidate sentences from a set of template “correct” sentences
  - Advantages:
    - Very suitable for most CALL systems
      - multiple choice quizzes, restricted QA quizzes, simple translation exercises
    - Fast
      - less than 2 sec for over 20 sentences

12

## “AntError”

- Methodology
  - Create a set of template “correct” answers
  - Compare learner’s sentence with each template sentence
    - If an exact match is found ⇒ no error
    - If no exact match ⇒ error
  - If no exact match
    - Find the ‘most similar’ template sentence
      - using Levenshtein distance measure (<http://www.merriampark.com/ld.htm>)
      - using phrase distance measure (Dree, M., 2002)

13

## “AntError”

- Methodology
  - If no exact match
    - Find largest clusters present in learner’s sentence and ‘most similar’ template sentence
    - Find redundant or missing clusters in learner’s sentence
    - Give instructions (corrections) for re-arranging, deleting or adding clusters to learner’s sentence until it matches the template sentence
    - Compare the POS tags of non-matching cluster words
    - Give instructions at POS level for correcting non-matching cluster words

14

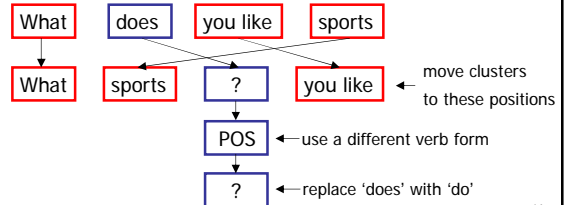
## “AntError”

- Methodology (Example)
  - Template “correct” sentences
    - What sports do you like?
    - Which sports do you like?
    - What sport do you like?
    - Which sport do you like?
  - Learner’s sentence
    - What does you like sports?
  - ‘Most similar’ template sentence
    - Exact match ⇒ no error
    - No exact match ⇒ error ⇒ What sports do you like?

15

## “AntError”

- Methodology (Example)
  - Find largest clusters appearing in learner’s sentence and template sentence



16

## “AntError”

- Results (1) – **これはいくらか?**
  - Template “correct” sentences:
    - How much is this? How much does this cost?
    - How much is it? How much does it cost?
    - How much money does this cost? ...
  - Learner sentences:
    - Total: 26 (with errors: 13, without errors: 13)
    - Error types: 13
  - Examples
    - How cost dose it?
    - How many costs is this?
    - How much will you cost this?

17

## “AntError”

- Results (1) – **これはいくらか?**

	“Word” Spell/Grammar	“AntError”
Error sent. marked with error	5/13 (38.5 %)	13/13 (100%)
Error sent. “corrected”	0/13 (0%)	13/13 (100%)
Non-error sent. marked with “error”	0	0
Non-error sent. “corrected”	0	0
Time to complete	20 sec.	0.7 sec.

18

## “AntError”

### ■ Results (1) “AntError” correction

- Learner sentence:
  - How much will you cost this ?
- ‘Closest’ template sentence:
  - How much does it cost ?
- Chunks: “How much”, “cost”
- Error correction hints:
  - ‘How much’ (okay)
  - Change ‘will’ to a ‘3rd person singular form of verb’ (does)
  - Change ‘you’ to a different ‘personal pronoun’ (it)
  - ‘cost’ (okay)
  - ‘this’ (delete)

19

## “AntError”

### ■ Results (1) “AntError” marking

- How much does it cost? SCORE: 0
- How much does this cost? SCORE: 0
- How much is it? SCORE: 0
- How much is this? SCORE: 0
- How much it is? SCORE: 2
- How long is this? SCORE: 4
- How much will it cost? SCORE: 4
- How does it cost? SCORE: 7

20

## “AntError”

### ■ Results (1) “AntError” marking

- How many costs does this? SCORE: 8
- How many is cost this? SCORE: 8
- How cost dose it? SCORE: 10
- How many take it? SCORE: 10
- How many costs is this? SCORE: 11
- What is cost this one? SCORE: 11
- How much will you cost this? SCORE: 14

21

## “AntError”

### ■ Results (2) – あなたはどれくらい岡山で勉強しているか、

- Template “correct” sentences:
  - How long have you studied in Okayama?
  - How long have you been studying in Okayama?
  - How many years have you studied in Okayama? ...
- Learner sentences:
  - Total: 25 (with errors: 19, without errors: 6)
  - Error types: 14
- Examples
  - What year have you been studying in the Okayama?
  - What years do you study in Okayama?
  - How long were studying in Okayama?

22

## “AntError”

### ■ Results (2) – あなたはどれくらい岡山で勉強しているか、

	“Word” Spell/Grammar	“AntError”
Error sent. marked with error	2/19 (10.5 %)	19/19 (100%)
Error sent. “corrected”	1/19 (5.3 %)	19/19 (100%)
Non-error sent. marked with “error”	0	0
Non-error sent. “corrected”	0	0
Time to complete	20 sec.	1.8 sec.

23

## “AntError”

### ■ Results (2) “AntError” correction

- Learner sentence:
  - How long were studying in Okayama?
- ‘Closest’ template sentence:
  - How long have you studied in Okayama ?
- Chunks: “How long”, “in Okayama”
- Error correction hints:
  - ‘How long’ (okay)
  - Change ‘were’ to a ‘singular present form of verb’ (have)
  - Change ‘studying’ to a ‘personal pronoun’ (you)
  - Add a ‘past tense of verb’ (studied)
  - ‘in Okayama’ (okay)

24

## “AntError”

- Results (3) – 彼は図書館で働いている。
  - Template “correct” sentences:
    - He works in a library. He is working at a library.
    - He is working in a library. He works in the library.
    - He works at a library. He works at the library.
  - Learner sentences:
    - Total: 25 (with errors: 20, without errors: 5)
    - Error types: 20
  - Examples
    - I am warking bookstation.
    - He works in lidurey.
    - She works a library.

25

## “AntError”

- Results (3) – 彼は図書館で働いている。

	“Word” Spell/Grammar	“AntError”
Error sent. marked with error	12/20 (60 %)	20/20 (100%)
Error sent. “corrected”	3/20 (15 %)	20/20 (100%)
Non-error sent. marked with “error”	0	0
Non-error sent. “corrected”	0	0
Time to complete	72 sec.	1.6 sec.

26

## “AntError”

- Results (3) “AntError” correction
  - Learner sentence:
    - She works a library.
  - ‘Closest’ template sentence:
    - He works at a library .
  - Chunks: “a library”, “works”
  - Error correction hints:
    - Change ‘She’ to a different ‘personal pronoun’ (He)
    - works (okay)
    - Add a ‘preposition/subordinating conjunction’ (at) (at position 3)
    - ‘a library’ (okay)

27

## “AntError”

- Results (4) – あなたの体重はいくらですか？
  - Template “correct” sentences:
    - How much do you weigh? What is your weight?
    - How much is your weight? What do you weigh?
  - Learner sentences:
    - Total: 25 (with errors: 19, without errors: 6)
    - Error types: 19
  - Examples
    - How much do you have a body?
    - What are you fat?
    - How many do you weigh?

28

## “AntError”

- Results (4) – あなたの体重はいくらですか？

	“Word” Spell/Grammar	“AntError”
Error sent. marked with error	5/19 (26.3 %)	19/19 (100%)
Error sent. “corrected”	1/19 (5.3 %)	19/19 (100%)
Non-error sent. marked with “error”	0	0
Non-error sent. “corrected”	0	0
Time to complete	38 sec.	0.6 sec.

29

## “AntError”

- Results (4) “AntError” correction

- Learner sentence:
  - How many do you weigh?
- ‘Closest’ template sentence:
  - How much do you weigh ?
- Chunks: “do you weigh”, “how”
- Error correction hints:
  - ‘How’ (okay)
  - Change ‘many’ to a different ‘adjective’ (much)
  - ‘do you weigh’ (okay)

30

## “AntError”

- Results (5) – 「2年後」に... をする。
  - Template “correct” sentences:
    - in ten years
    - in ten years time
  - Learner sentences:
    - Total: 25 (with errors: 15, without errors: 10)
    - Error types: 9
  - Examples
    - ten years after
    - ten years later
    - in the ten years
    - for ten years

31

## “AntError”

- Results (5) – 「2年後」に... をする。

	“Word” Spell/Grammar	“AntError”
Error sent. marked with error	0/15 (0 %)	15/15 (100%)
Error sent. “corrected”	0/15 (0 %)	15/15 (100%)
Non-error sent. marked with “error”	0	0
Non-error sent. “corrected”	0	0
Time to complete	28 sec.	0.1 sec.

32

## “AntError”

- Results (5) “AntError” correction
  - Learner sentence:
    - for ten years
  - ‘Closest’ template sentence:
    - in ten years
  - Chunks: “ten years”
  - Error correction hints:
    - Change ‘for’ to a different ‘preposition/subordinating conjunction’ (in)
    - ‘ten years’ (okay)

33

## Summary & Future Work

- Developed an automatic learner error identification and correction tool
  - “AntError”
- 100% accuracy when identifying errors
- 100% accuracy when “correcting” errors
- Fast (< 0.1 sec. per learner sentence)
- Flexible (can be adapted to any problem with template “correct” sentences)
- Offers increasingly detailed corrections
- Offers possibility for automatic grading

34

## Summary & Future Work

- Increase detail of error corrections
- Improve “AntError” to be more useful with semi-free learner sentences
  - e.g. allowing wildcard template sentence words and clusters
- Develop an intuitive graphical user interface
  - Similar to other “Ant” tools, such as “AntConc” and “AntMover”

35