

# Evaluating error detection in parsing second-language learner data using mal-rules

**Dan Flickinger**

CSLI, Stanford University

DELPH-IN Summit 2021

21 July 2021

# Data collection setting: Up366 in Beijing

- Commercial product for grammar-checking in use since Feb 2020
- 2,000 middle/high school students in China learning English
- Each day, one Chinese sentence to translate into English
- ACE with ERG + mal-rules running on servers

See my summit talks from 2013, 2019 and 2020 for details

Note that we are still using the standard Redwoods parse ranking model, not yet informed by training data that includes mal-rules. To improve precision, some coverage is sacrificed by masking or blocking a number of lexical entries and constructions. Ideally, this brittle use of masking can be reduced or eliminated once a malrule-informed model is put into use (see Luis's talk).

- Evaluate spelling, punctuation, capitalization and grammar
- Each month, we get an anonymized log of student inputs



# Sentence-a-Day examples

Presented to the student:

在周日，我经常和我妈妈一起去我家附近的一家购物中心购物。

Sample correct translation:

I usually go shopping with my mother in a shopping mall near my home on Sunday.

Examples of incorrect student answers:

- On *sunday*, I often go shopping in a shopping mall near my home with my mother.
- On *Sunday*, I often go shopping in a shopping center around our house with my mother.
- On Sunday, I often go shopping *with* in a shopping center near my home with my mother.
- On Sunday, I often go shopping *in shopping center* around my house with my mother.
- I often go shopping with my mom *on the near*



# Evaluating system responses

Four possible coarse-grained pairings:

- **CC** Correct answer, rightly evaluated as correct
- **CX** Correct answer, wrongly evaluated as incorrect
- **XC** Incorrect answer, wrongly evaluated as correct
- **XX** Incorrect answer, rightly evaluated as incorrect



# Student data from Month 3

- 51,360 student answers over the 27 days
- Sampled every tenth sentence from time-ordered spreadsheet
- Began manual annotation of this 5136-sentence sample

Average sentence length: 12.6 tokens

Unique answers: 3635 of the 5136



# Manually evaluating system responses

- **CX** 50 of 5136 [1%]: correct, but reported with error(s)
  - 10 of 640 unparsed
    - Many people pursue happiness [ all their lives. ]*
  - 1 of 42 fragments
    - Establishing [ the natural reserve is a good way to provide more living space for these wild animals. ]*
  - 39 of 1334 robust sentence analyses
- **XX** 1966 of 5136 [38.3%]: error, correctly reported as error
  - 1295 of 1334 robust sentence analyses have errors
  - 630 of 640 unparsed are ungrammatical
  - 41 of 42 fragments are indeed fragments



# Full data sets for months 1, 7, 15

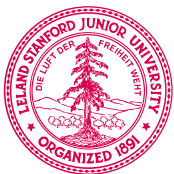
- ERG+mal defines 795 error codes
- Lexemes, lexical rules, constructions, punctuation, capitalization

Month	Total Items	Robust Items	Error Codes	Unique Codes
Month 1	317965	79922 (25%)	164455	506
Month 7	98064	19272 (20%)	44133	389
Month 15	67150	16080 (24%)	29005	331



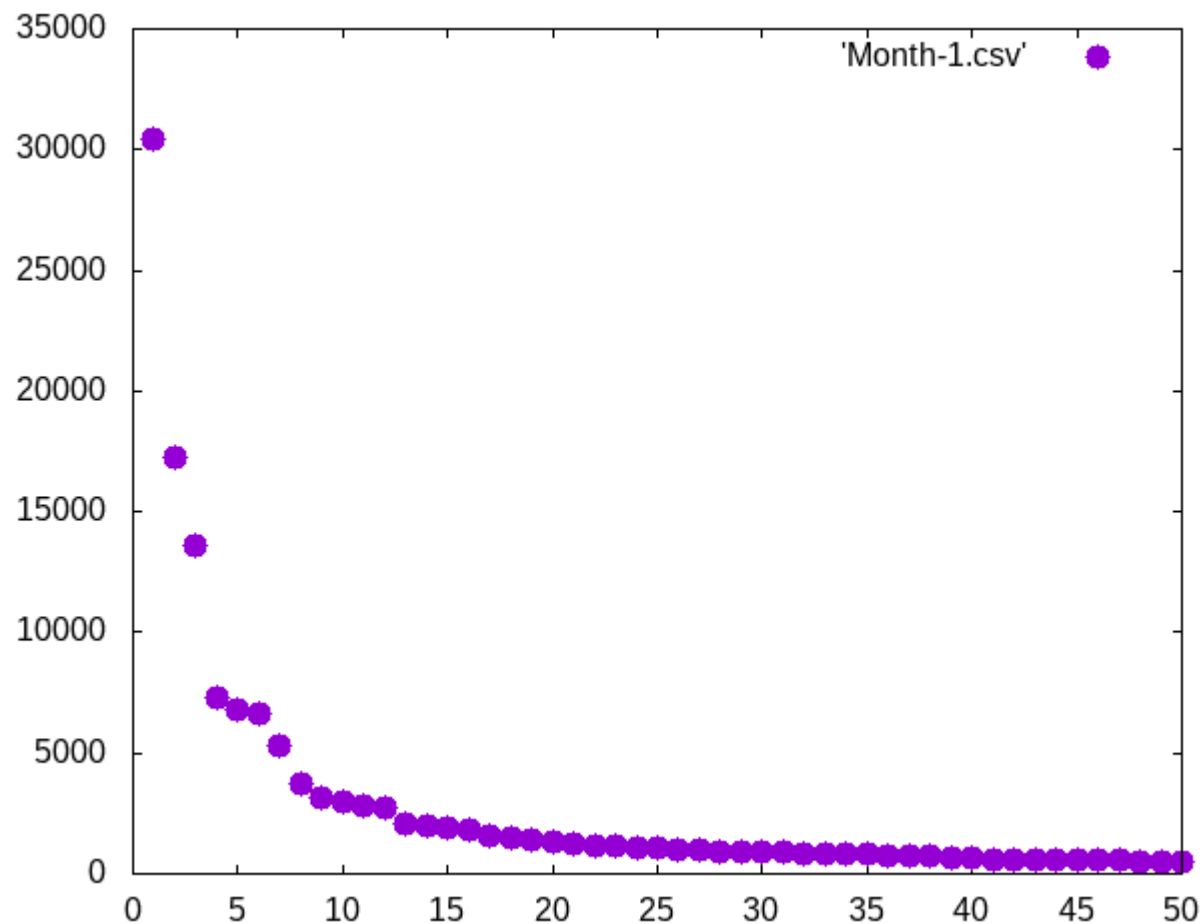
## Month 1 most frequent errors

Frequency	Error code	Example
30443	comma_sdwch_pct_rbst	A tall,green tree appeared.
17252	hdn_bnp_c_rbst	He went to house.
13616	third_sg_fin_v_ilr_rbst	Nancy write a letter.
7308	w_noinitcap_dlr_rbst	the cat slept.
6836	comma_prefix_pct_rbst	They arrived ,in the city.
6595	cl-cl_runon-cma_c_rbst	The dog barked, the cat ran.
5293	w_hasnoninitcap_dlr_rbst	We slept During the lecture.
3696	v_np_pp_le_rbst	He went the restaurant.
3121	i_spelled_l_rbst	He thought I won.
2983	v_vp_be-bse-p_le_rbst	The dog is chase the cats.
2854	a_det_mass_rbst	We found a pollution.
2766	a_det_rbst	She saw a owl.
2114	period_sdwch_pct_rbst	1.Cats
2010	hdn-aj_rc-asym_c_rbst	The cat, which we admire arose.
1878	ago_p1_rbst	They will arrive two days ago.





# Month 1 most frequent errors



## Month 7 most frequent errors

Frequency	Error code	Example
8037	comma_sdwch_pct_rbst	A tall,green tree appeared.
4320	w_noinitcap_dlr_rbst	the cat slept.
2725	hdn_bnp_c_rbst	He went to house.
2210	period_sdwch_pct_rbst	1.Cats
2088	third_sg_fin_v_ilr_rbst	Nancy write a letter.
1913	comma_prefix_pct_rbst	They arrived ,in the city.
1844	v_-_le_rbst	<b>She buys.</b>
1493	v-v_crd-fin-ncj_c_rbst	They ate cookies, drank milk.
1453	not_mv_adv1_rbst	<b>We not buy those books.</b>
1386	cl-cl_runon-cma_c_rbst	The dog barked, the cat ran.
1074	w_hasnoninitcap_dlr_rbst	We slept During the lecture.
1009	cl_rc-fin-nwh_c_rbst	The supermarket, we often go to, is closed.
760	vp_sbrd-prd-aj_rbst	Kim moved slow.
725	cl-frg_runon_c_rbst	We will win. In the morning.
708	v_vp_be-bse-p_le_rbst	The dog is chase the cats.

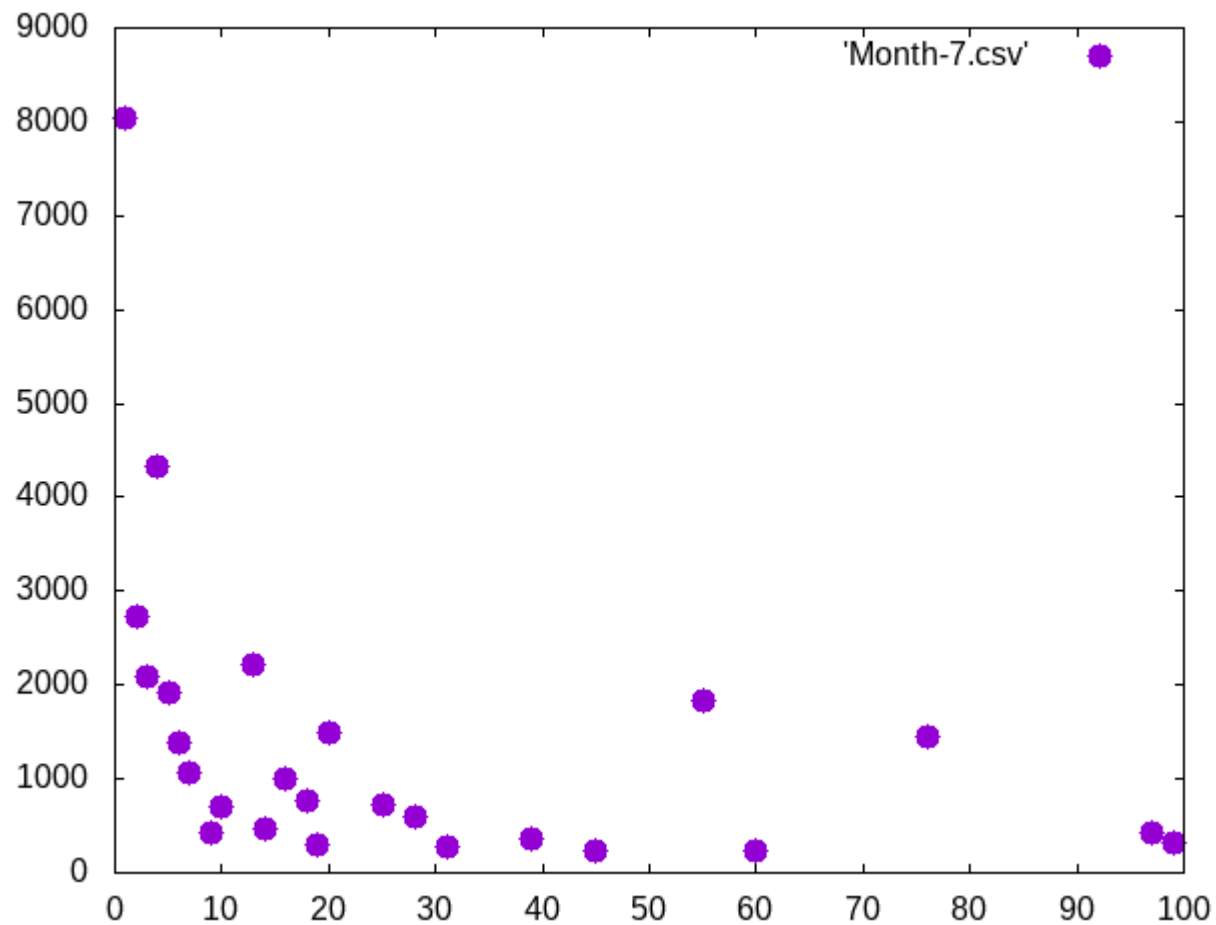


# Drop in frequency of Month 7 from Month 1

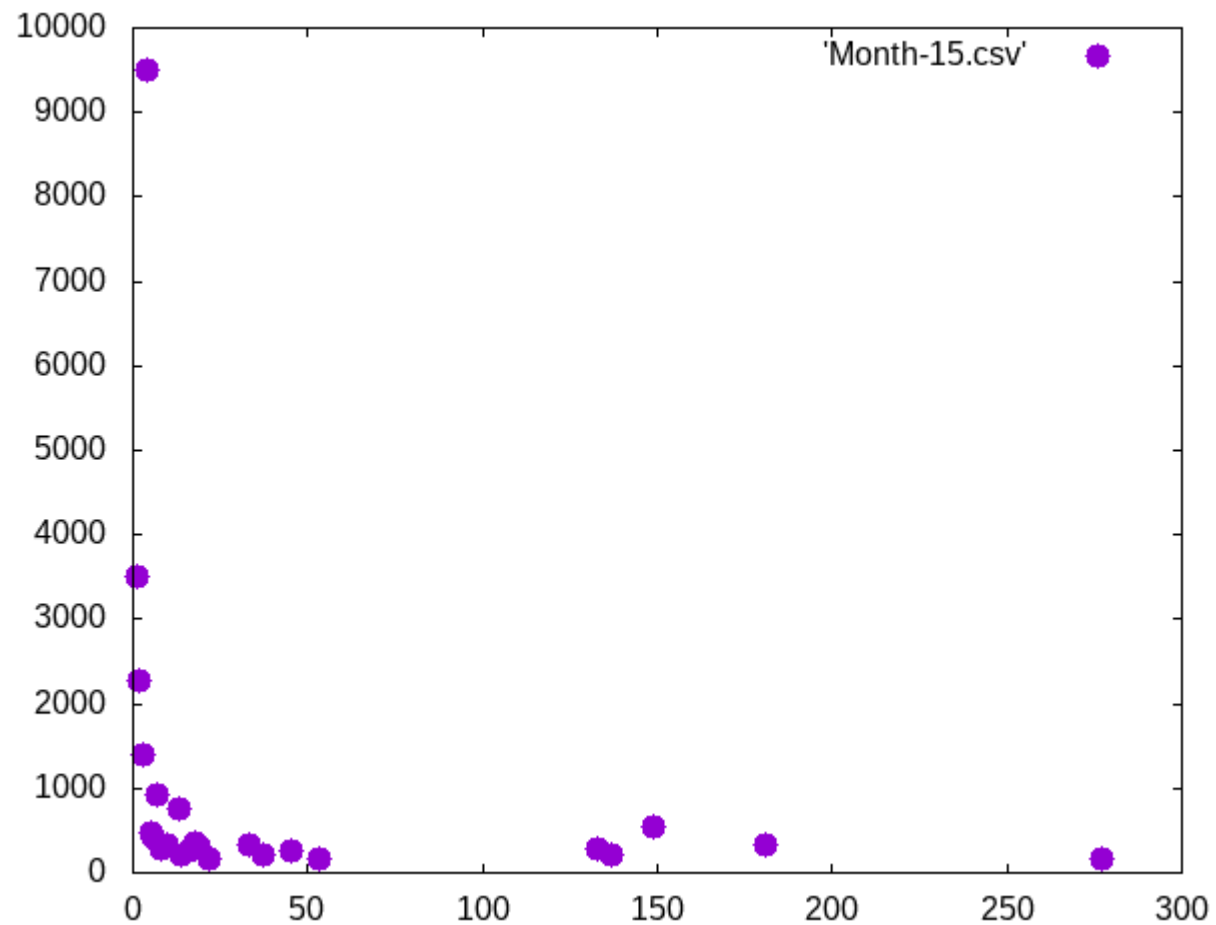
Frequency	Error code	Example
30443	comma_sdwch_pct_rbst	A tall,green tree appeared.
17252	hdn_bnp_c_rbst	He went to house.
13616	third_sg_fin_v_ilr_rbst	Nancy write a letter.
7308	w_noinitcap_dlr_rbst	the cat slept.
6836	comma_prefix_pct_rbst	They arrived ,in the city.
6595	cl-cl_runon-cma_c_rbst	The dog barked, the cat ran.
5293	w_hasnoninitcap_dlr_rbst	We slept During the lecture.
3696	v_np_pp_le_rbst	<b>He went the restaurant.</b>
3121	i_spelled_l_rbst	He thought I won.
2983	v_vp_be-bse-p_le_rbst	The dog is chase the cats.
2854	a_det_mass_rbst	<b>We found a pollution.</b>
2766	a_det_rbst	She saw a owl.
2114	period_sdwch_pct_rbst	1.Cats
2010	hdn-aj_rc-asym_c_rbst	The cat, which we admire arose.
1878	ago_p1_rbst	<b>They will arrive two days ago.</b>



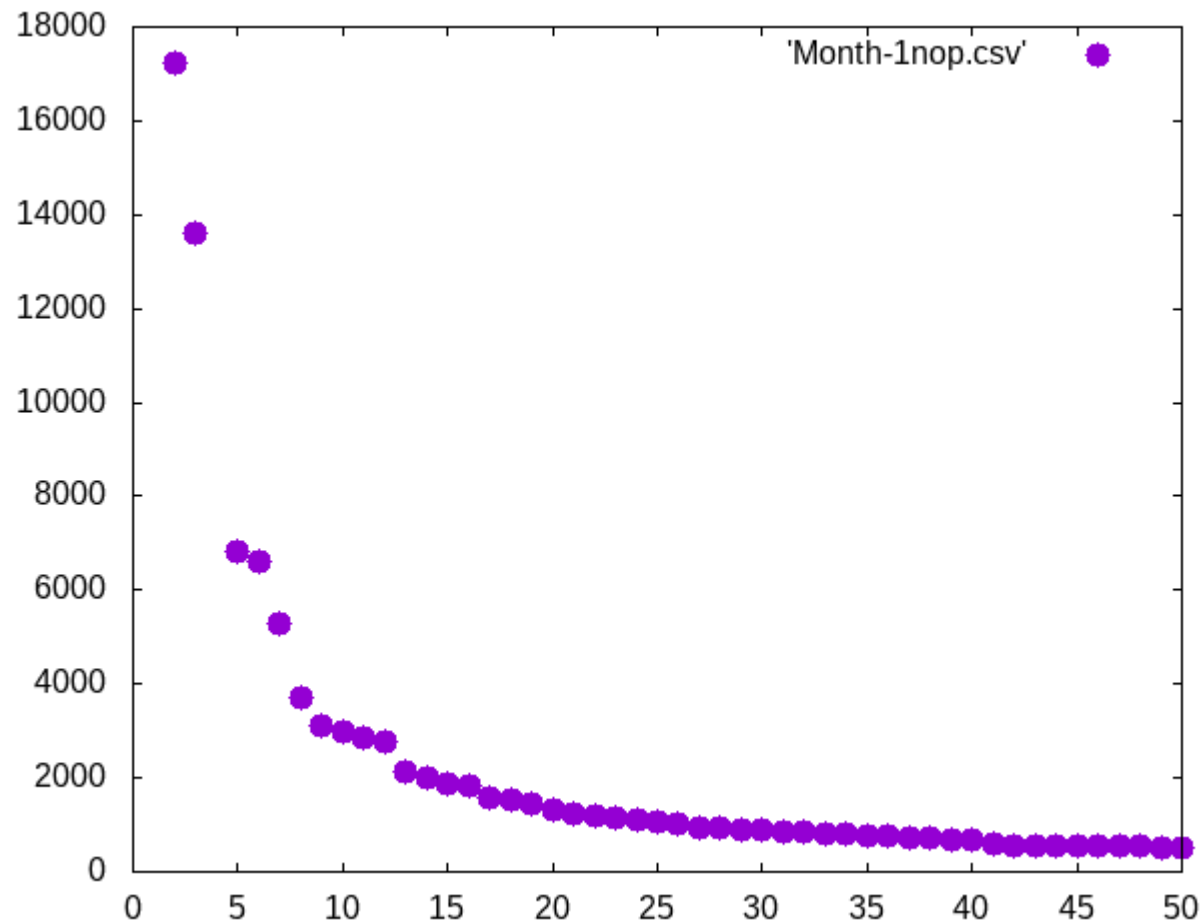
# Month 7 most frequent errors



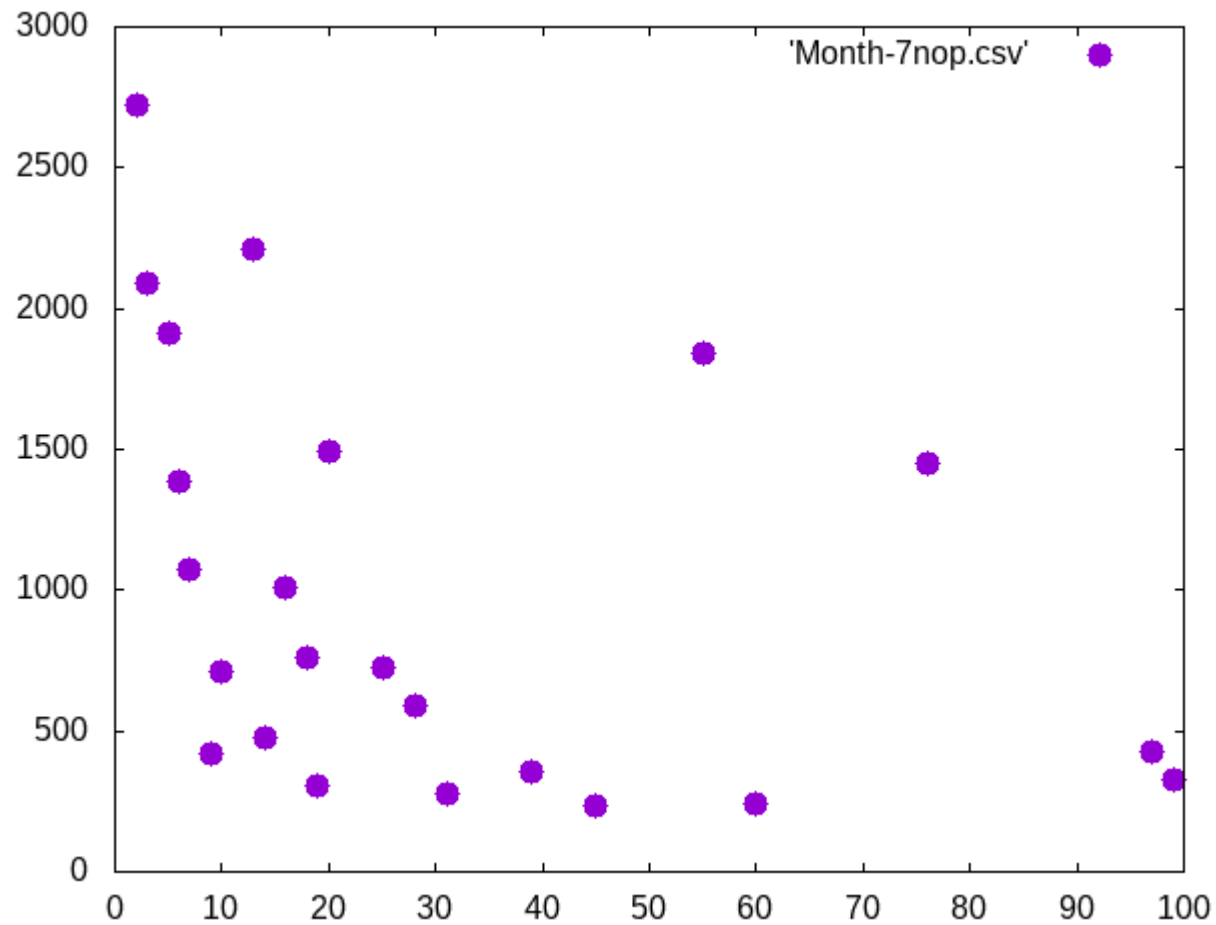
# Month 15 most frequent errors



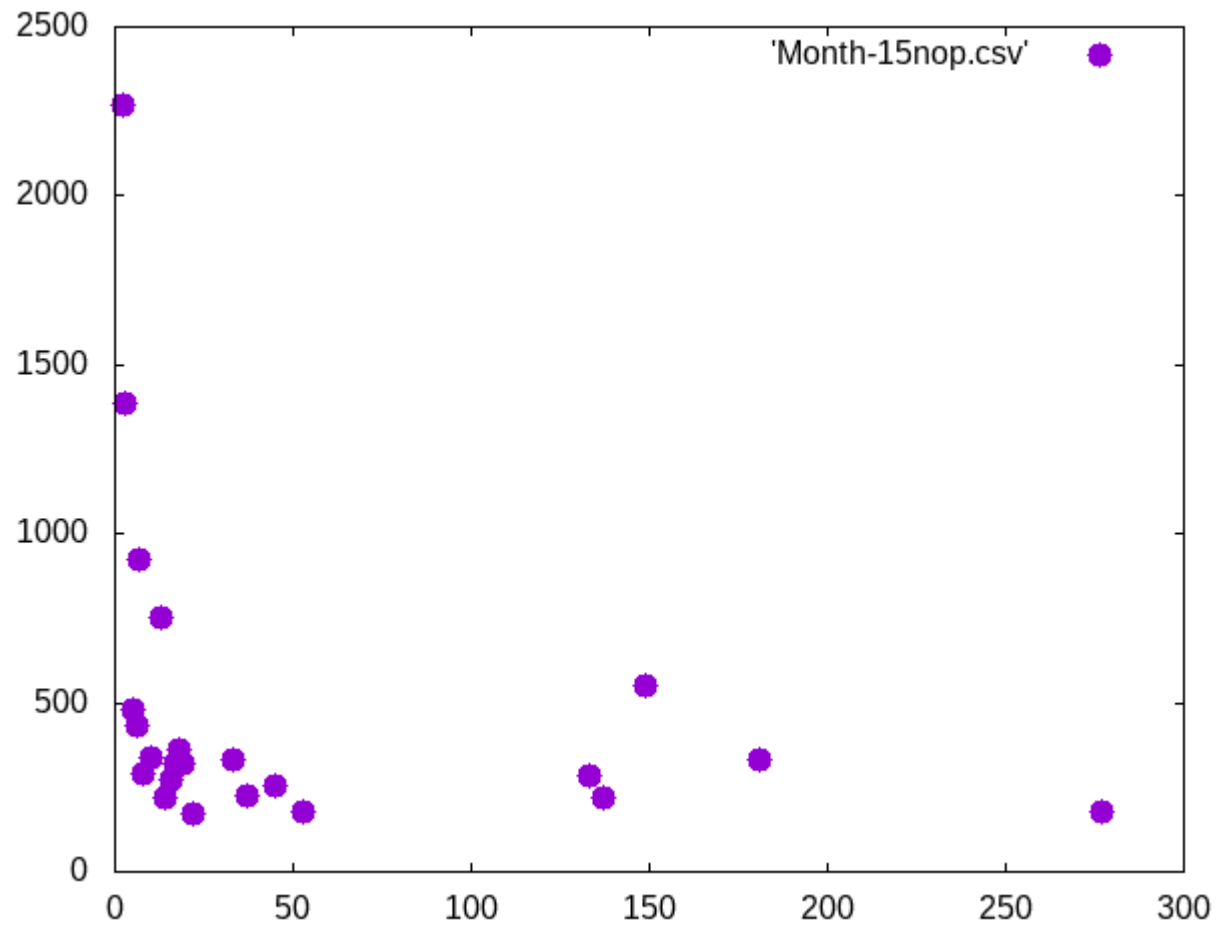
# Month 1 excluding punctuation errors



# Month 7 excluding punctuation errors



# Month 15 excluding punctuation errors





# Summary

- Analysis of a half million student sentences over 15 months
- Parse coverage above 90%; false error reports less than 1%
- Errors found in 20-25% of parsed sentences
- Broadly similar frequency counts among top 25 error types
- Some frequent Month 1 errors drop in Months 7 and 15
- Absent knowledge of student identity, difficult to study individual improvements in writing proficiency

