Computational Models of Sentence Processing and Syntactic Acquisition

Byung-Doh Oh (오병도)

Dept. of Linguistics, The Ohio State University
Collaborators: Christian Clark, Lifeng Jin, William Schuler

Feb. 14, 2022, Dongguk University

The Ohio State University

Introduction

Why study language computationally?

Both humans and computers can “learn” and “process” language

Unlike human subjects, we have more control over computers

- Input data, architecture, and training objective
- Thorough inspection of model predictions

How can we use computational modeling to shed light on human language processing and acquisition?

Roadmap

Modeling sentence processing with left-corner parsing (Oh, Clark, & Schuler, 2021, forthcoming)

Modeling syntactic acquisition with unsupervised PCFG induction (Jin, Oh, & Schuler, 2021)

Conclusion and future directions

Modeling sentence processing with left-corner parsing

Oh, Clark, and Schuler (2021). Surprisal estimators for human reading times need character models. In *Proc. ACL*.
Oh, Clark, and Schuler (forthcoming). Comparison of structural parsers and neural language models as surprisal estimators. In *Frontiers in AI*.

Expectation-based theories of sentence processing

Processing difficulty is determined by *predictability* in context (Hale, 2001; Levy, 2008)

Predictability can be quantified through information-theoretic *surprisal* (놀라움; Shannon, 1948)

Strong correlation with (human) behavioral and neural measures of processing difficulty
(Demberg & Keller, 2008; Roark et al., 2009; Shain et al., 2020; Smith & Levy, 2013, inter alia)

Expectation-based theories of sentence processing

Surprisal (놀라움)

$$\mathrm{S}(w_t) \stackrel{\mathrm{def}}{=} -\log_2 \mathrm{P}(w_t \mid w_1, w_2, \ldots, w_{t-1})$$

- Can be calculated from any probability model over words
- Open question how to best estimate the language comprehender's probability model

Language models (언어 모델; Goodkind & Bicknell, 2018; Smith & Levy, 2013; Wilcox et al., 2020)

- Trained to predict the next word

Incremental parsers (점진적 파서; Hale et al., 2018; Jin & Schuler, 2020; van Schijndel et al., 2013)

- Trained to predict the next word and (usually syntactic) structure
- Maintains *multiple hypotheses* about structure in *parallel*

In this work

Incremental left-corner parser trained to predict common linguistic abstractions

- Syntactic tree structure with rich node labels (Oh & Schuler, 2021)
- Morphological rules for observed word

Evaluation of parser and LM surprisal on measures of processing difficulty

- Self-paced reading times (Futrell et al., 2021)
- Eye-gaze durations (Kennedy et al., 2003)
- fMRI blood oxygenation level-dependent signals (Shain et al., 2020)

Model description

Left-corner parsing

$$\mathsf{P}(w_t\ q_t \mid q_{t-1}) \;=\; \sum_{\ell_t, g_t} \mathsf{P}(\ell_t \mid q_{t-1}) \cdot \mathsf{P}(w_t \mid q_{t-1}\ \ell_t) \cdot \mathsf{P}(g_t \mid q_{t-1}\ \ell_t\ w_t) \cdot \mathsf{P}(q_t \mid q_{t-1}\ \ell_t\ w_t\ g_t)$$

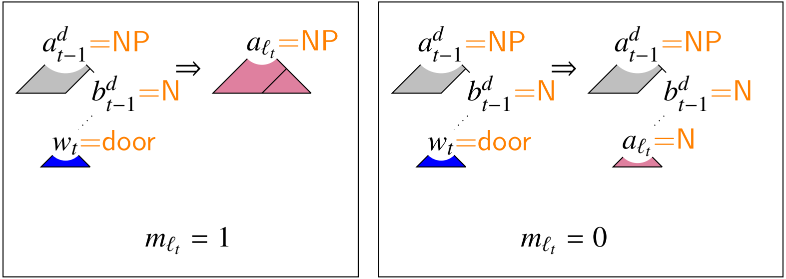
- w_t: Observed word
- q_t: Hidden states representing partial tree structures
- ℓ_t: Lexical decision
- g_t: Grammatical decision

Defines a fixed number of decisions at every word

Model description

Left-corner parsing

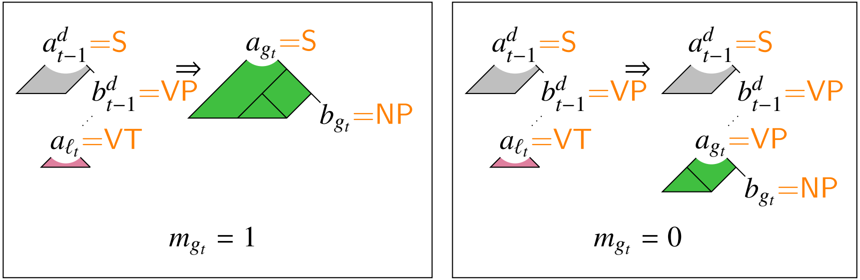
$$\mathsf{P}(w_t\ q_t \mid q_{t-1}) = \sum_{\ell_t, g_t} \mathsf{P}(\ell_t \mid q_{t-1}) \cdot \mathsf{P}(w_t \mid q_{t-1}\ \ell_t) \cdot \mathsf{P}(g_t \mid q_{t-1}\ \ell_t\ w_t) \cdot \mathsf{P}(q_t \mid q_{t-1}\ \ell_t\ w_t\ g_t)$$



Model description

Left-corner parsing

$$\mathsf{P}(w_t\ q_t \mid q_{t-1}) = \sum_{\ell_t, g_t} \mathsf{P}(\ell_t \mid q_{t-1}) \cdot \mathsf{P}(w_t \mid q_{t-1}\ \ell_t) \cdot \color{blue}{\mathsf{P}(g_t \mid q_{t-1}\ \ell_t\ w_t)} \cdot \mathsf{P}(q_t \mid q_{t-1}\ \ell_t\ w_t\ g_t)$$



Model description

Left-corner parsing

$$\mathsf{P}(w_t\ q_t \mid q_{t-1}) = \sum_{\ell_t, g_t} \mathsf{P}(\ell_t \mid q_{t-1}) \cdot \mathsf{P}(w_t \mid q_{t-1}\ \ell_t) \cdot \mathsf{P}(g_t \mid q_{t-1}\ \ell_t\ w_t) \cdot \mathsf{P}(q_t \mid q_{t-1}\ \ell_t\ w_t\ g_t)$$

Character-based word model

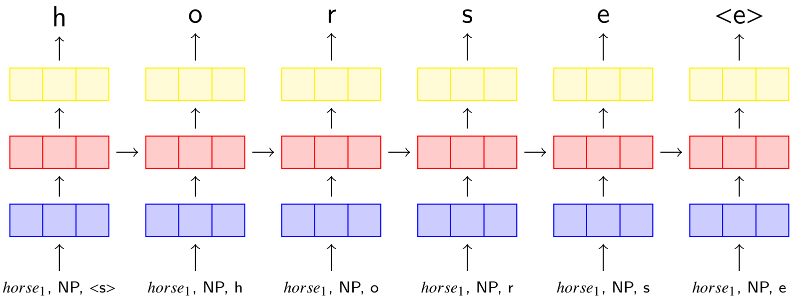
$$\mathsf{P}(w_t \mid q_{t-1}\ \ell_t) = \sum_{x_t, r_t} \mathsf{P}(x_t \mid q_{t-1}\ \ell_t) \cdot \mathsf{P}(r_t \mid q_{t-1}\ \ell_t\ x_t) \cdot \mathsf{P}(w_t \mid q_{t-1}\ \ell_t\ x_t\ r_t)$$

To a lemma x_t, apply a morphological rule r_t for word w_t (to *horse* apply $*\!\rightarrow\!*$s for *horses*)

Model description

Character-based word model

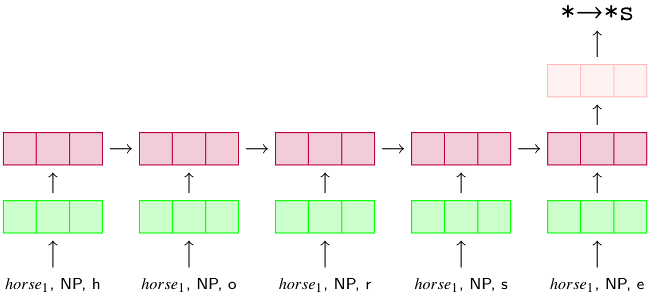
$$\mathrm{P}(w_t \mid q_{t-1}\ \ell_t) = \sum_{x_t, r_t} \mathrm{P}(x_t \mid q_{t-1}\ \ell_t) \cdot \mathrm{P}(r_t \mid q_{t-1}\ \ell_t\ x_t) \cdot \mathrm{P}(w_t \mid q_{t-1}\ \ell_t\ x_t\ r_t)$$



Model description

Character-based word model

$$\mathsf{P}(w_t \mid q_{t-1}\ \ell_t) \ = \ \sum_{x_t, r_t} \mathsf{P}(x_t \mid q_{t-1}\ \ell_t) \ \cdot \ \color{blue}{\mathsf{P}(r_t \mid q_{t-1}\ \ell_t\ x_t)} \ \cdot \ \mathsf{P}(w_t \mid q_{t-1}\ \ell_t\ x_t\ r_t)$$



Surprisal estimation

Parser trained on WSJ02-21 (Marcus et al., 1993)

Surprisal estimated using beam search (빔 탐색)

- Full model (*Structural*)
- Baseline 1: No syntactic category labels for ℓ_t, g_t (*-cat*)
- Baseline 2: Relative frequency estimation for w_t (*-morph*)

Evaluation

Evaluation on three datasets collected during naturalistic language processing

- Natural Stories self-paced reading (Futrell et al., 2021)
- Dundee eye-tracking (Kennedy et al., 2003)
- Natural Stories fMRI (Shain et al., 2020)

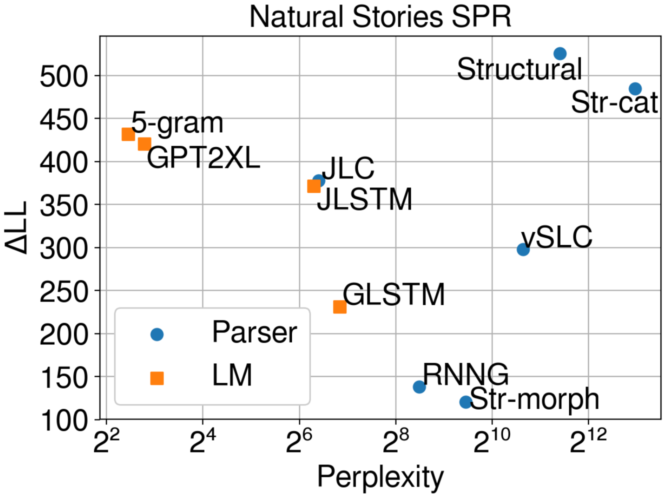
Evaluation metrics (Goodkind & Bicknell, 2018; Hao et al., 2020; Wilcox et al., 2020)

- Perplexity (혼란도): How well does model X predict the next word? ($\downarrow$)
- Δlog-likelihood (Δ 로그우도): How well does surprisal from model X predict the data? ($\uparrow$)

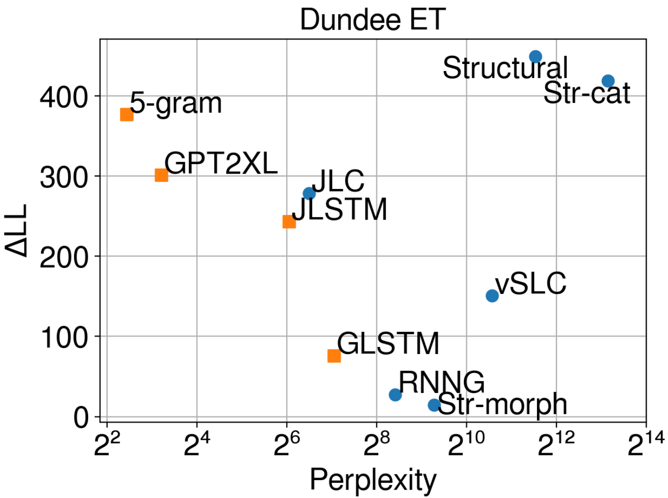
Comparison against surprisal estimates from various models

- LMs: 5-gram (Heafield et al., 2013), GLSTM (Gulordava et al., 2018), JLSTM (Jozefowicz et al., 2016), GPT2XL (Radford et al., 2019)
- Parsers: RNNG (Hale et al., 2018), vSLC (van Schijndel et al., 2013), JLC (Jin & Schuler, 2020)

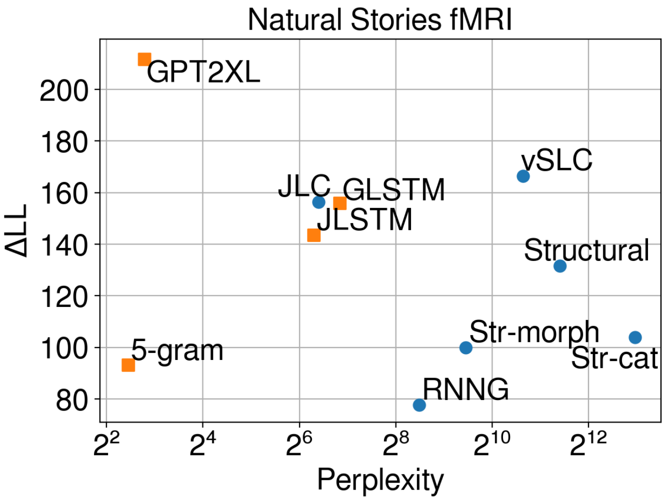
Results (self-paced reading)



Results (eye-tracking)



Results (fMRI)



Summary

Incremental left-corner parser with common linguistic abstractions

Surprisal estimates show better fits to human response data

- Better than large-scale neural LMs on SPR and ET data

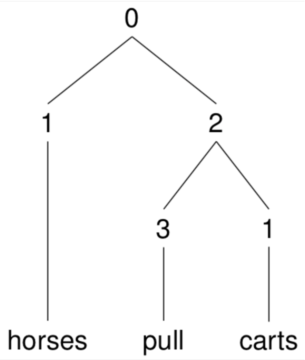
New nuance to the relationship between perplexity and predictive power (Goodkind & Bicknell, 2018; Wilcox et al., 2020)

Modeling syntactic acquisition with unsupervised PCFG induction

Jin, Oh, and Schuler (2021). Character-based PCFG induction for modeling the syntactic acquisition of morphologically rich languages. In *Findings of EMNLP*.

Unsupervised PCFG induction

horses pull carts $\longrightarrow$



Unsupervised PCFG induction

Nonterminal expansion probabilities (비단말 확장):

$P(0 \rightarrow 1\ 2)$

$P(2 \rightarrow 3\ 1)$

...

Terminal expansion probabilities (단말 확장):

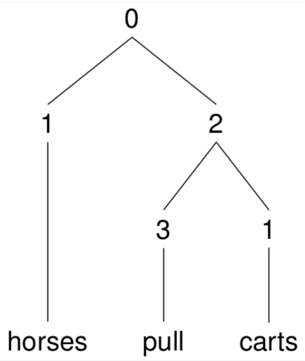
$P(1 \rightarrow \text{horses})$

$P(3 \rightarrow \text{pull})$

$P(1 \rightarrow \text{carts})$

...

$\rightarrow$



Unsupervised PCFG induction

Shows the extent to which grammars can be learned from distributional data alone

Recent neural approaches fairly successful (Kim et al., 2019; Yang et al., 2021; Zhu et al., 2020)

However, word-based PCFGs cannot inspect word affixes

Terminal expansion probabilities:

P(1 $\rightarrow$ horses)

P(3 $\rightarrow$ pull)

P(1 $\rightarrow$ carts)

Unsupervised PCFG induction

Child language learners are sensitive to functional affixes (Dye et al., 2019; Haryu & Kajikawa, 2016; Mintz, 2013)

Word-based models are less appropriate for morphologically rich languages

This work presents

- A character-based model for neural PCFG induction
- Experiments on child-directed speech corpora

Model description

Objective function: marginal probability of sentence σ

$$\mathsf{P}(\sigma) = \sum_{\tau \text{ for } \sigma} \prod_{\eta \in \tau \text{ s.t. } c_\eta \to c_{\eta 1}\ c_{\eta 2}} \mathsf{P}(c_\eta \to c_{\eta 1}\ c_{\eta 2}) \cdot \prod_{\eta \in \tau \text{ s.t. } c_\eta \to w_\eta} \mathsf{P}(c_\eta \to w_\eta)$$

"Split" model: nonterminal or terminal expansion?

$$\mathsf{P}(\mathsf{Term} \mid c_\eta) = \underset{\{0,1\}}{\mathsf{SoftMax}}(\mathsf{ResNet}_{\mathsf{split}}(\mathbf{v}_{\mathbf{c}_\eta}))$$

Model description

Objective function: marginal probability of sentence σ

$$\mathsf{P}(\sigma) = \sum_{\tau \text{ for } \sigma} \prod_{\eta \in \tau \text{ s.t. } c_\eta \to c_{\eta 1}\ c_{\eta 2}} \mathsf{P}(c_\eta \to c_{\eta 1}\ c_{\eta 2}) \cdot \prod_{\eta \in \tau \text{ s.t. } c_\eta \to w_\eta} \mathsf{P}(c_\eta \to w_\eta)$$

Nonterminal expansion probabilities

$$\mathsf{P}(c_\eta \to c_{\eta 1}\ c_{\eta 2}) = \mathsf{P}(\mathsf{Term}{=}0 \mid c_\eta) \cdot \underset{c_{\eta 1}, c_{\eta 2}}{\mathsf{SoftMax}}(\mathbf{W}_{\mathsf{nont}}\, \mathbf{v}_{c_\eta})$$

Model description

Objective function: marginal probability of sentence σ

$$\mathsf{P}(\sigma) = \sum_{\tau \text{ for } \sigma} \prod_{\eta \in \tau \text{ s.t. } c_\eta \to c_{\eta 1}\ c_{\eta 2}} \mathsf{P}(c_\eta \to c_{\eta 1}\ c_{\eta 2}) \cdot \prod_{\eta \in \tau \text{ s.t. } c_\eta \to w_\eta} \mathsf{P}(c_\eta \to w_\eta)$$

Character-based terminal expansion probabilities (*NeuralChar*)

$$\mathsf{P}(c_\eta \to w_\eta) = \mathsf{P}(\text{Term}{=}1 \mid c_\eta) \cdot \prod_{l_i \in \{l_1, \ldots, l_n\}} \mathsf{P}(l_i \mid c_\eta, l_1, \ldots, l_{i-1})$$

Model description

Objective function: marginal probability of sentence σ

$$\mathsf{P}(\sigma) = \sum_{\tau \text{ for } \sigma} \prod_{\eta \in \tau \text{ s.t. } c_\eta \to c_{\eta 1}\ c_{\eta 2}} \mathsf{P}(c_\eta \to c_{\eta 1}\ c_{\eta 2}) \cdot \prod_{\eta \in \tau \text{ s.t. } c_\eta \to w_\eta} \mathsf{P}(c_\eta \to w_\eta)$$

Word-based terminal expansion probabilities (*NeuralWord*)

$$\mathsf{P}(c_\eta \to w_\eta) = \mathsf{P}(\mathrm{Term}{=}1 \mid c_\eta) \cdot \underset{w_\eta}{\mathsf{SoftMax}}(\mathsf{ResNet}_{\mathsf{term}}(\mathbf{v}_{c_\eta}))$$

Evaluation

NeuralChar and *NeuralWord* trained and evaluated on transcriptions of child-directed speech from CHILDES (MacWhinney, 2000)

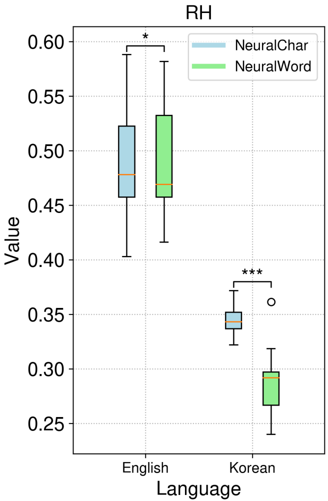
- English (Brown, 1973): Eve (1;6-2;3)
- Korean (Ryu et al., 2015): Jong (1;3-3;5)
- Chinese (Deng et al., 2018): Tong (1;0-4;5)
- German (Behrens, 2006): Leo (1;11-4;11)

Evaluation metric: Recall-Homogeneity (Jin, Schwartz, et al., 2021)

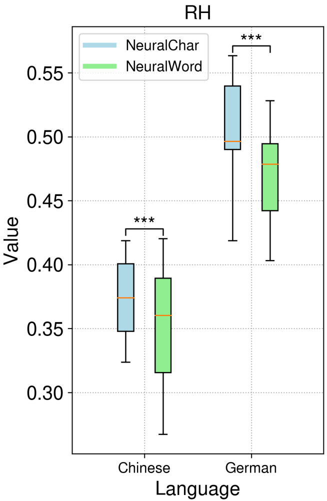
- Recall (재현율): How well does grammar X recall attested constituents? (↑)
- Homogeneity (순도): How homogeneous are syntactic categories of grammar X? (↑)

Results from 10 runs using 90 categories

Results



RH
NeuralChar
NeuralWord
*
***
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
Value
English
Korean
Language



RH
NeuralChar
NeuralWord
***
***
0.55
0.50
0.45
0.40
0.35
0.30
Value
Chinese
German
Language

Induced preterminal categories

Induced category	Count	Attested category (relative frequency)	Examples
NC-63	100	sf (1.0)	.
NC-29	73	npd+jxt (0.23), nq (0.12), ncn (0.12), npd+jcs (0.1), npd (0.1), nq+jcs (0.07), ncn+jcs (0.05)	이거는, 종현이, 이거, 이게, 아빠, 종현아, 종현이가, 이건
NC-62	48	sf (1.0)	?
NC-38	25	px+ef (0.32), pvg+ef (0.2), paa+ef (0.2), pvg+ep+ef (0.16)	와, 있어, 먹어, 갔었어, 졌다, 졌네, 했었어, 놀구요
NC-16	21	pvg+ecx (0.67), pvg+ecs (0.14), paa+ecc (0.1), paa+ef (0.1)	가져, 타러, 보고, 많아요, 알고, 길고, 작아요
NC-2	20	ncn (0.55), ncn+jcj (0.15), ncn+jcs (0.1), pad+ef (0.05), mag (0.05), ncn+jxt (0.05), pvd+ecs (0.05)	엄마, 엄마랑, 엄마가, 그래, 그냥, 엄마는, 그러고
NC-6	20	ii (1.0)	아이구, 아우, 아이고, 아휴, 오, 오오
NC-7	20	pad+ef (1.0)	그렇지, 그래, 그지, 그지요
NW-55	61	sf (1.0)	.
NW-32	51	ii (0.45), pad+ef (0.2), ncn (0.12), mag (0.08), maj (0.06)	그렇지, 어, 짠, 또, 아빠, 자, 엄마, 여기
NW-54	50	sf (1.0)	?
NW-0	46	ncn (0.35), npd+jxt (0.07)	이거는, 이게, 여기, 이, 물, 책, 엄마, 꽃
NW-14	39	sf (1.0)	.
NW-10	34	ncn+jcs (0.24), mag (0.15), ncn (0.06), pvg+ecs (0.06), ncn+jxc (0.06), nq (0.06), paa+ecs (0.06)	많이, 책도, 목이, 꽃이, 가렸네, 백일, 전신, 살이
NW-44	34	paa+ef (0.18), pvg+ef (0.15), ncn+jp+ef (0.09), pvg+ep+ef (0.06), mag (0.06), pvg+etm (0.06), pvg+ef+jxf (0.06), paa+ef+jxf (0.06)	적어요, 아빠가, 때야, 졌다, 졌네, 나와, 그냥, 목욕했어, 보네
NW-29	30	mag (0.2), ncn+jcs (0.1), ncn (0.1), paa+etm (0.1), npp (0.07), pvg+ecx (0.07), ncn+jxt (0.07)	종현이, 너, 다, 작은, 구두는, 진짜, 살, 디게

Summary

A neural model for unsupervised PCFG induction

- Allows clean manipulation of terminal expansion model

Subword information leads to more accurate grammars on child-directed speech

- Bigger impact on morphologically richer languages

Further support for a distributional model of syntactic acquisition

Conclusion and future directions

Conclusion

Parser results show linguistic abstractions are important for capturing humanlike processing difficulty

Inducer results show subword information is important for grammar induction, especially for morphologically rich languages

Computational approaches can be used to model human sentence processing and syntactic acquisition

Future directions

Investigating the contribution of discourse-level information in sentence processing
(e.g. coreference; Jaffe, Oh, & Schuler, 2021)

Modeling language acquisition with more realistic input (e.g. acoustic signals; Shain & Elsner, 2020)

Connecting NLP/ML techniques to psycholinguistic research questions

Thank you for listening!

Parser code: https://github.com/byungdoh/acl21_semproc
Inducer code: https://github.com/lifengjin/charInduction
Regression code: https://github.com/modelblocks/modelblocks-release

oh.531@osu.edu
https://byungdoh.github.io

References I

Behrens, H. (2006). The input-output relationship in first language acquisition. *Language and Cognitive Processes, 21*(1-3), 2–24. https://doi.org/10.1080/01690960400001721

Brown, R. (1973). *A first language: The early stages*. Harvard University Press.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210. https://doi.org/10.1016/j.cognition.2008.07.008

Deng, X., Yip, V., Macwhinney, B., Matthews, S., Ziyin, M., Jing, Z., & Lam, H. (2018). A multimedia corpus of child Mandarin: The Tong Corpus. *The Journal of Chinese Linguistics*, *46*(1), 69–92. https://doi.org/10.1353/jcl.2018.0002

Dye, C., Kedar, Y., & Lust, B. (2019). From lexical to functional categories: New foundations for the study of language development. *First Language*, *39*(1), 9–32. https://doi.org/10.1177/0142723718809175

Futrell, R., Gibson, E., Tily, H. J., Blank, I., Vishnevetsky, A., Piantadosi, S., & Fedorenko, E. (2021). The Natural Stories corpus: A reading-time corpus of English texts containing rare syntactic constructions. *Language Resources and Evaluation*, *55*, 63–77. https://doi.org/10.1007/s10579-020-09503-7

Goodkind, A., & Bicknell, K. (2018). Predictive power of word surprisal for reading times is a linear function of language model quality. *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics*, 10–18. https://www.aclweb.org/anthology/W18-0102

References II

Gulordava, K., Bojanowski, P., Grave, E., Linzen, T., & Baroni, M. (2018). Colorless green recurrent networks dream hierarchically. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1195–1205. https://www.aclweb.org/anthology/N18-1108

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, 1–8. https://www.aclweb.org/anthology/N01-1021

Hale, J., Dyer, C., Kuncoro, A., & Brennan, J. (2018). Finding syntax in human encephalography with beam search. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 2727–2736. https://www.aclweb.org/anthology/P18-1254

Hao, Y., Mendelsohn, S., Sterneck, R., Martinez, R., & Frank, R. (2020). Probabilistic predictions of people perusing: Evaluating metrics of language model performance for psycholinguistic modeling. *Proceedings of the 10th Workshop on Cognitive Modeling and Computational Linguistics*, 75–86. https://www.aclweb.org/anthology/2020.cmcl-1.10

Haryu, E., & Kajikawa, S. (2016). Use of bound morphemes (noun particles) in word segmentation by Japanese-learning infants. *Journal of Memory and Language*, *88*, 18–27. https://doi.org/https://doi.org/10.1016/j.jml.2015.11.007

References III

Heafield, K., Pouzyrevsky, I., Clark, J. H., & Koehn, P. (2013). Scalable modified Kneser-Ney language model estimation. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 690–696. https://www.aclweb.org/anthology/P13-2121

Jaffe, E., Oh, B.-D., & Schuler, W. (2021). Coreference-aware surprisal predicts brain response. *Findings of the Association for Computational Linguistics: EMNLP 2021*, 3351–3356. https://aclanthology.org/2021.findings-emnlp.285

Jin, L., Oh, B.-D., & Schuler, W. (2021). Character-based PCFG induction for modeling the syntactic acquisition of morphologically rich languages. *Findings of the Association for Computational Linguistics: EMNLP 2021*, 4367–4378. https://aclanthology.org/2021.findings-emnlp.371

Jin, L., & Schuler, W. (2020). Memory-bounded neural incremental parsing for psycholinguistic prediction. *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*, 48–61. https://www.aclweb.org/anthology/2020.iwpt-1.6

Jin, L., Schwartz, L., Doshi-Velez, F., Miller, T., & Schuler, W. (2021). Depth-bounded statistical PCFG induction as a model of human grammar acquisition. *Computational Linguistics*, *47*(1), 181–216. https://doi.org/10.1162/coli_a_00399

Jozefowicz, R., Vinyals, O., Schuster, M., Shazeer, N., & Wu, Y. (2016). Exploring the limits of language modeling. *arXiv*. https://arxiv.org/abs/1602.02410

Kennedy, A., Hill, R., & Pynte, J. (2003). The Dundee Corpus. *Proceedings of the 12th European conference on eye movement*.

References IV

Kim, Y., Dyer, C., & Rush, A. (2019). Compound probabilistic context-free grammars for grammar induction. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2369–2385. https://aclanthology.org/P19-1228

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk*. Lawrence Erlbaum Associates.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, *19*(2), 313–330. https://www.aclweb.org/anthology/J93-2004

Mintz, T. H. (2013). The segmentation of sub-lexical morphemes in English-learning 15-month olds. *Frontiers in Psychology*, *4*(24), 1–12. https://doi.org/https://doi.org/10.3389/fpsyg.2013.00024

Oh, B.-D., Clark, C., & Schuler, W. (2021). Surprisal estimators for human reading times need character models. *Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 3746–3757. https://aclanthology.org/2021.acl-long.290

Oh, B.-D., Clark, C., & Schuler, W. (forthcoming). Comparison of structural parsers and neural language models as surprisal estimators. *Frontiers in Artificial Intelligence*. https://www.frontiersin.org/articles/10.3389/frai.2022.777963

References V

Oh, B.-D., & Schuler, W. (2021). Contributions of propositional content and syntactic category information in sentence processing. *Proceedings of the 11th Workshop on Cognitive Modeling and Computational Linguistics*, 241–250. https://www.aclweb.org/anthology/2021.cmcl-1.28

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Technical Report*. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 324–333. https://www.aclweb.org/anthology/D09-1034

Ryu, J.-Y., Horie, K., & Shirai, Y. (2015). Acquisition of the Korean imperfective aspect markers –ko iss–and –a iss–by Japanese learners: A multiple-factor account. *Language Learning*, *65*(4), 791–823. https://doi.org/10.1111/lang.12132

Shain, C., Blank, I. A., van Schijndel, M., Schuler, W., & Fedorenko, E. (2020). fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*, *138*. https://doi.org/10.1016/j.neuropsychologia.2019.107307

Shain, C., & Elsner, M. (2020). Acquiring language from speech by learning to remember and predict. *Proceedings of the 24th Conference on Computational Natural Language Learning*, 195–214. https://www.aclweb.org/anthology/2020.conll-1.15

References VI

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*, 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

van Schijndel, M., Exley, A., & Schuler, W. (2013). A model of language processing as hierarchic sequential prediction. *Topics in Cognitive Science*, *5*(3), 522–540. https://doi.org/10.1111/tops.12034

Wilcox, E. G., Gauthier, J., Hu, J., Qian, P., & Levy, R. P. (2020). On the predictive power of neural language models for human real-time comprehension behavior. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*, 1707–1713. https://cognitivesciencesociety.org/cogsci20/papers/0375

Yang, S., Zhao, Y., & Tu, K. (2021). PCFGs can do better: Inducing probabilistic context-free grammars with many symbols. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1487–1498. https://aclanthology.org/2021.naacl-main.117

Zhu, H., Bisk, Y., & Neubig, G. (2020). The return of lexical dependencies: Neural lexicalized PCFGs. *Transactions of the Association for Computational Linguistics*, *8*, 647–661. https://aclanthology.org/2020.tacl-1.42