Learning both variability and exceptionality in probabilistic OT grammars

Society for Computation in Linguistics
Inaugural Meeting, January 2018
Salt Lake City, UT

Aleksei Nazarov
(U of Huddersfield)

Outline

- Existing models for learning phonologically exceptional words (Pater 2010, Becker 2009, Coetzee 2009):
  - Based on “classic” inconsistency to find exceptions (Tesar 1995)
  - Cannot find exceptions in the face of within-word variability

- Proposal:
  - Opposite ranking trends in probabilistic grammar (“soft inconsistency”) trigger assumption of exceptionality

- Tested on (simplified) Hebrew dataset with both exceptionality and variability

Introduction

Exceptionality

- Phonological processes may have exceptions
  - Dutch 2-syllable monomorphemes: initial stress (e.g., van der Hulst 1984)

    ká.nɔn “canon” ɑ́p.pəl “apple”
    kó.bra “cobra”

  - Exceptional words go against the rule and must be accounted for

    ka.nɔ́n “cannon” ɑp.pɛ́l “gathering” (?)

Variability

- Phonological processes may have variable outputs
  - English between-word place assimilation: optional

    /gɹin bɔks/ → [gɹi**n** bɔks] ~ [gɹi**m** bɔks]
    /ɪn bɛd/ → [ɪ**n** bɛd] ~ [ɪ**m** bɛd] (Coetzee & Pater 2011)

  - Phonology has to accommodate for this variable pronunciation

Exceptionality & Variability

- Variable outputs and exceptionality co-occur
  - Modern Hebrew: stops /p,b,k/ optionally spirantize to fricatives [f,v,χ] after a vowel (Temkin-Martínez 2010)

    /me**k**ase/ → [me**k**ase ~ me**χ**ase]

  - Exceptional words: spirantization is mandatory, or blocked

    /ma**k**ar/ → [ma**χ**ar] *****k** (mandatory)
    /da**k**ar/ → [da**k**ar] *****χ** (blocked)

OT: Variability

- Variability in OT proper: constraint ranking not fixed, but partially determined by chance
  - Partially Ordered Grammars (Anttila 2002): constraint ranking is underspecified

    {A, B} >> C >> D
  - Stochastic OT (Boersma 1998): random noise added to constraint ranks
  - Pairwise Ranking Grammars (Jarosz 2015): each pairwise ranking (A >> B, B >> C, A >> C) has a probability between 0 and 1

Pairwise Ranking

- E.g., p(Faith >> Agree[Place]) = 60%

/ɪ**n** bɛd/	Faith	Agree[Place]
☞ ɪ**n** bɛd		*
ɪ**m** bɛd	*!	

- Therefore: p(Agree[Place] >> Faith) = 40%

/ɪ**n** bɛd/	Agree[Place]	Faith
ɪ**n** bɛd	*!	
☞ ɪ**m** bɛd		*

Pairwise Ranking

- Problem: sampling independently from pairwise ranking probabilities can yield logically impossible rankings
  - p(A >> B) = p(B >> C) = p(A >> C) = 0.5
    could yield #{A >> B, B >> C, C >> A}

- Jarosz (2015): Sample one constraint pair at a time (in random order) to build a full ranking
  - The actual ranking probability used at each step is conditional on the entire ranking built so far
  - For instance, p(A>>C) may equal 0, but
    p(A >> C| A >> B, B >> C) = 1

Learning variability

- Learning probabilistic constraint ranking (e.g., Jarosz 2015; but also Boersma 1998):
  - If you encounter 6 tokens that require Faith >> Agree[Place]
  - And you encounter 4 tokens that require Agree[Place] >> Faith
  - Then p(Faith >> Agree[Place]) approaches 60%

- Opposite ranking requirements yield $0 < p < 100\%$

OT: Exceptions

- Exceptionality in OT: assume different rankings for each word
  - Cophonologies (Nouveau 1994, Inkelas 1998): groups of words may be fed through separate grammars (separate rankings)
    - Sublexical phonology (Becker and Gouskova 2016): similar to cophonologies but with weighted constraints and mechanism for predicting whether new words are exceptions
  - Indexed constraint theory (Kraska-Szlenk 1995, Pater 2000): there is one grammar, but some constraints (the indexed constraints) are only violated for a subset of the lexicon

Indexed constraints

- For ká.nɔn, Non-Finality >> Rightmost

/kanɔn/	Rightmost(*i*)	Non-Finality	Rightmost
☞ (ká.nɔn)			*
ka(nɔ́n)		*!	

- For ka.nɔ́n, Rightmost(i) >> Non-Finality

/kanɔn(*i*)/	Rightmost(*i*)	Non-Finality	Rightmost
(ká.nɔn)	*!		*
☞ ka(nɔ́n)		*	

Learning exceptionality

- Indexation absent in phonetic form
  - How to find which words are exceptional?

- Existing approach (Pater 2010) uses a categorical method: Recursive Constraint Demotion (RCD: Tesar 1995)
  - RCD finds partial constraint ranking logically necessary to generate a dataset
  - At each stage of ranking, asks if the data require a ranking A >> B given the ranking already established

Learning exceptionality

- Diagnostic for exceptions (Pater 2010): assume there is exceptionality...
  - If some words require Cons1 >> Cons2, while other words require Cons2 >> Cons1
    (i.e., if there is logical *inconsistency*)

Learning exceptionality

- Diagnostic for exceptions (Pater 2010): assume there is exceptionality...
  - If ká.nɔn requires Non-Fin >> Rightmost, and ká.nɔn requires Rightmost >> Non-Fin
    (i.e., if there is logical *inconsistency*)

Learning exceptionality

- Diagnostic for exceptions (Pater 2010): assume there is exceptionality...
  - If ká.nɔn requires Non-Fin >> Rightmost, and ká.nɔn requires Rightmost >> Non-Fin
    (i.e., if there is logical *inconsistency*)

- Becker (2009), Coetzee (2009): methods to go from "there is exceptionality" to specific indexed constraints

Learning exceptionality

- Diagnostic for exceptions (Pater 2010): assume there is exceptionality...
  - If ká.nɔn requires Non-Fin >> Rightmost, and ká.nɔn requires Rightmost >> Non-Fin
    (i.e., if there is logical *inconsistency*)

- Opposite ranking requirements yield indexed constraints

Variability & Exceptionality

- This discovery procedure is not sufficient to find exceptions in the face of variability:
  - Opposite ranking requirements mean either variability (Jarosz 2015) or exceptions (Pater 2010)
  - RCD incompatible with variability

- How can we handle both kinds of variation in the same data set?

Variability & Exceptionality

- Moore-Cantwell (2017): discovery procedure for relative memory strength of features
  - e.g. /k/ with a strong [-continuant] feature remains [k], /k/ with a weak [-continuant] feature varies between [k] and [**χ**]

- Not sufficient for cases that require diacritics (full picture of Hebrew does require diacritics; see Temkin-Martínez 2010)

Proposal: “soft inconsistency”

- Existing diagnostic for exceptions (Pater 2010) – assume there is exceptionality...
  - If some words require Cons1 >> Cons2, while other words require Cons2 >> Cons1.

Proposal: “soft inconsistency”

- Existing diagnostic for exceptions (Pater 2010) – assume there is exceptionality...
  - If some words require Cons1 >> Cons2, while other words require Cons2 >> Cons1.
- **Proposal** – assume there is exceptionality:
  - If some words have p(Cons1 >> Cons2) above 50%, while other words have p(Cons1 >> Cons2) below 50%.

Proposal: “soft inconsistency”

- Existing diagnostic for exceptions (Pater 2010) – assume there is exceptionality...
  - If some words require Cons1 >> Cons2, while other words require Cons2 >> Cons1.
- **Proposal** – assume there is exceptionality:
  - If some words have p(Cons1 >> Cons2) above 50% + threshold, while other words have p(Cons1 >> Cons2) below 50% - threshold.

Proposal: “soft inconsistency”

- Existing diagnostic for exceptions (Pater 2010) – assume there is exceptionality...
  - If some words require Cons1 >> Cons2, while other words require Cons2 >> Cons1.
- **Proposal** – assume there is exceptionality:
  - If some words have p(Cons1 >> Cons2) above e.g. 60%, while other words have p(Cons1 >> Cons2) below e.g. 40%.

Implementation

- Framework used:
  - Expectation Driven Learning (EDL, Jarosz 2015)

- Directly estimates pairwise ranking probs for Pairwise Ranking Grammars
  - Perfect tool for assessing soft inconsistency

“some words: p(Cons1 >> Cons2) > 50%,
other words: p(Cons1 >> Cons2) < 50%”

EDL + soft inconsistency

- Initial state:
  - data without exception indices
  - set of pre-defined constraints (but no indexed constraints)
  - 50/50 ranking probabilities for all pairs

- Probabilities updated through multiple rounds of estimation and re-estimation

  (Expectation Maximization, Dempster 1977)

Pseudocode

- Repeat procedure till 95% training data accuracy
  1. For each pair of phonological constraints:
     - Re-estimate p(Cons1 >> Cons2|all words) and p(Cons1 >> Cons2|one word)
     - **Soft Inconsistency diagnostic**: find which words are exceptional w.r.t. this constraint pair and the lexicon
  2. Find which constraint pair has greatest ranking probability divergence between exceptions and the overall pattern
  3. Mark the exceptional words for that constraint pair with an index (and add an indexed constraint to the constraint set)

Case study: Modern Hebrew

Hebrew spirantization

- Proposal tested on (simplified) Hebrew:
  - Variable spirantizatio n
    /me**k**ase/ → [me**k**ase ~ me**χ**ase]
  - Exceptions: ~~variable~~ or ~~spirantization~~
    /ma**k**ar/ → [ma**χ**ar] *__k__
    /da**k**ar/ → [da**k**ar] *__χ__
- Questions:
  - Will learner find the variable default rule?
  - Will learner account for exceptions?

Hebrew spirantization

- Abridged version of Temkin-Martínez' (2010) judgment task data
  - Computed the proportion of the following types in the corpus
    - ak → ak ~ aχ
    - ak → ak
    - ak → aχ
    - /b,p,k/ not after a vowel
  - Any non-postvocalic variability leveled

Hebrew spirantization

- Actual data set used:

underlying stops Postvocalic	**underlying stops non-postvocalic**
/me**k**ase/ → [me**k**ase ~ me**χ**ase]	/lin**p**oʃ/ → [lin**p**oʃ]
/ʃa**b**ar/ → [ʃa**b**ar ~ ʃa**v**ar]	/lis**b**ol/ → [lis**b**ol]
/da**k**ar/ → [da**k**ar]	/liʃ**k**oa/ → [liʃ**k**oa]
/me**b**arer/ → [me**v**arer]	/liʃ**p**oχ/ → [liʃ**p**oχ]
/me**b**atel/ → [me**v**atel]	/liz**k**ot/ → [liz**k**ot]
/ga**b**a/ → [ga**v**a]	

Phonological Constraints

- Very simple constraints used in simulation (following Temkin-Martínez 2010):
  - *Stop: no stops
  - *Fricative: no fricatives
  - *[Vowel-Stop]: no vowel-stop sequences

  - Ident: don’t change any segment
  - Max: don’t delete any segment

Phonological Constraints

- Desired rankings for default pattern:
  - *[Vowel-Stop] >> NoChange

    to ensure possibility of ak → aχ
  - NoChange >> *Stop

    to ensure that tk → tk
  - p(*Fricative >> *[Vowel-Stop]) ≈ 0.5

    to ensure variability ak → ak ~ aχ

Phonological Constraints

- Desired account for exceptions:
  - /da**k**ar/ → [da**k**ar] (never a fricative)
    - Highly ranked indexed constraint *Fricative(*i*) or NoChange(*i*)
  - /ga**b**a/ → [ga**v**a] (always a fricative)
    - Highly ranked indexed constraint *Stop(*i*) or *[Vowel-Stop](*i*)

Simulation Results

Results

- Learner was run 100 times
  - Each time, learner was allowed 80 iterations of the learning procedure
  - All 100 runs reached at least 95% accuracy on the training data within 14 to 36 iterations
- Results of each run evaluated on:
  - Accuracy on the default pattern
  - Marking all and only exceptions with indices

Accuracy on default

- Desired default pattern after a vowel:
  - Underlying stops optionally change into a fricative ak → aχ
  - The vowel preceding a stop is never deleted *ak → k
- This pattern was observed 99% of the time

Accuracy on default

- Desired default pattern not after a vowel:
  - Underlying stops never change into a fricative
    tk → tk
  - The consonant preceding a stop is never deleted
    *tk → k
- This pattern was observed 97% of the time

Exception marking

- Words that follow the default pattern **should not** be marked as exceptions
  - This is followed 100% of the time in ak → ak ~ aχ cases
  - Followed 98% of the time in tk → tk cases

Exception marking

- Exceptional words **should** be marked as exceptions
  - This is followed 100% of the time in ak → aχ cases
  - Followed 51% of the time in ak → ak cases
    - These are harder to find a basis for exceptionality for
  - Overall: 88% of exceptions receive exception marking

Results Summary

1. 100 out of 100 runs reached at least 95% data accuracy within 14 to 36 iterations out of 80

2. Overall accuracy on default pattern > 95%

3. Overall accuracy of exception identification:
   - Non-exceptions remain unindexed: 99% of cases
   - Intended exceptions are indexed: 88% of cases

Results Summary

- Proposed discovery procedure effective in discovering exceptions despite variability
  - Apart from exceptions that go in the general direction of the data (most data points retain underlying stops), in this case [dakar] (51%)
  - This is a consequence of Soft Inconsistency
    - Only designed to distinguish exceptions that go in **opposite** direction of general pattern in data

Conclusion

Summary

- No existing discovery procedure that can find exceptional words in the face of variable pronunciation (like in Hebrew)

- My solution: modify Pater’s (2010) discovery procedure to be more flexible with variability
  - Embedded in Jarosz’ (2015) Expectation Driven Learning framework

Summary

- Tested exceptions discovery procedure on simplified Hebrew data:
  - Was able to match training data
  - Learned default variable pattern to great degree of accuracy
  - Learned exceptions to high degree of accuracy (>90%)
    - Exceptions of the form ak → ak not learned as well; hope to fix in new version of learner

Future Work

- Future directions:
  - Test on larger, non-simplified test cases
  - Experiment with different ways of presenting variation (logical, as in this case, or as frequencies)
  - Integrate with various theoretical frameworks of exceptions

Future Work

- Future directions:
  - Test on larger, non-simplified test cases
    - Full version of same Hebrew data set
    - Dutch reduction (Kager 1989)
    - ...

Future Work

- Future directions:
  - Experiment with different ways of presenting variation (logical, as in this case, or as frequencies)
    - In this case, learner did not match frequencies of each variant, but simply tried to generate some grammatical variant of each word
    - Frequency-matching learner will make different predictions

Future Work

- Future directions:
  - Integrate with various theoretical frameworks of exceptions
    - Here, exceptions implemented in Indexed Constraint Theory
    - However, Soft Inconsistency is in principle also compatible with Cophonology Theory and Sublexical Phonology
    - It will be important to explore the issues that arise in these frameworks

Thank you!

Acknowledgments

- Many thanks to
  - Adam Albright
  - Paul Boersma
  - Silke Hamann
  - Gaja Jarosz
  - Giorgio Magri
  - John McCarthy
  - Marc van Oostendorp
  - Joe Pater
  - Robert Staubs
  - Kristine Yu
  - audiences at UMass Amherst, the University of Manchester, the University of Amsterdam, MIT, the University of Connecticut, and the Annual Meeting on Phonology 2017

References

- Anttila, Arto. 2002. “Morphologically Conditioned Phonological Alternations.” NLLT 20(1): 1–42.
- Apoussidou, Diana. 2007. The learnability of metrical phonology. PhD dissertation, University of Amsterdam.
- Becker, Michael. 2009. *Phonological Trends in the Lexicon: The Role of Constraints*. PhD dissertation, University of Massachusetts Amherst.
- Beguš, Gašper and Aleksei Nazarov. Under review. *Lexicon against Naturalness: Unnatural Gradient Phonotactic Restrictions and Their Origins*.
- Boersma, Paul. 1998. *Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives*. Holland Academic Graphics/IFOTT.
- Coetzee, Andries. 2009. “Learning Lexical Indexation.” *Phonology* 26 (1): 109–45.
- Coetzee, Andries, and Joe Pater. 2011. “The Place of Variation in Phonological Theory.” In *The Blackwell Companion to Phonology*, edited by Marc van Oostendorp, Colin J Ewen, Elizabeth Hume, and Keren Rice, 401–34. Oxford: Wiley-Blackwell.
- Della Pietra, S., V.J. Della Pietra and J.D. Lafferty. 1997. Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19: 380-393.
- Dempster, A. P.; N. M. Laird; and D. B. Rubin. 1977. “Maximum Likelihood from Incomplete Data via the EM Algorithm.” *Journal of the Royal Statistical Society. Series B (Methodological)* 39(1).1–38.

References

- Guinn, Devon and Aleksei Nazarov. In progress. *Evidence for phonological features and phonotactics in beatboxing vocal percussion*.
- Harris, Zellig. 1946. ”From Morpheme to Utterance.” *Language* 22:161-83.
- Hayes, B. and C. Wilson. 2008. A maximum entropy model of phonotactics and phontactic learning. *Linguistic Inquiry* 39: 379-440.
- van der Hulst, Harry. 1984. *Syllable Structure and Stress in Dutch*. Linguistic Models 8. Dordrecht: Foris Publications.
- Inkelas, Sharon. 1998. “The Theoretical Status of Morphologically Conditioned Phonology: A Case Study from Dominance.” *Yearbook of Morphology* 1997: 121–55.
- Jäger, Gerhard. 2007. “Maximum entropy models and Stochastic Optimality Theory.” In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling & Chris Manning (eds.), *Architectures, rules, and preferences: variations on themes by Joan W. Bresnan*, Stanford: CSLI, 467–479.
- Jarosz,Gaja. 2006. “Rich Lexicons and Restrictive Grammars: Maximum Likelihood Learning in Optimality Theory.” PhD dissertation, Johns Hopkins University.
- Jarosz, Gaja. 2013. Learning with hidden structure in Optimality Theory and Harmonic Grammar: beyond Robust Interpretive Parsing. *Phonology* 30 1:27:71.
- Jarosz, Gaja. 2015. “Expectation Driven Learning of Phonology.” Ms., University of Massachusetts Amherst.
- Jarosz, Gaja. 2016. “Learning Opaque and Transparent Interactions in Harmonic Serialism.” In Gunnar Ólafur Hansson, Ashley Farris-Trimble, Kevin McMullin, Douglas Pulleyblank (eds.), *Proceedings of the 2015 Annual Meeting on Phonology*, Linguistic Society of America.

References

- Kager, René. 1989. *A Metrical Theory of Stress and Destressing in English and Dutch*. Dordrecht - Holland/Providence RI - USA: Foris.
- Kraska-Szlenk, Iwona. 1995. “The Phonology of Stress in Polish.” PhD dissertation, University of Illinois, Urbana-Champaign.
- Lin, Y. 2005. *Learning features and segments from waveforms: A statistical model of early phonological acquisition*. Doctoral dissertation, UCLA.
- Mielke, Jeff. 2004. *The emergence of distinctive features*. PhD dissertation, Ohio SU.
- Moore-Cantwell, Claire. 2017. “Concurrent learning of the lexicon and phonology.” LSA 2017.
- Mullin, Kevin. 2011. *The Necessity of Diacritics for Descriptive Adequacy.* RUMMIT 2011.
- Nazarov, Aleksei. 2014. “A radically emergentist approach to phonological features: implications for grammars.” *Nordlyd* 41:1, 21-58.
- Nazarov, Aleksei. 2016. *Extending Hidden Structure Learning: Features, Opacity, and Exceptions*. Doctoral dissertation, University of Massachusetts Amherst.
- Nazarov, Aleksei. Under revision. ‘Reduction Domains between feet and syllables: Evidence from Dutch.‘
- Nazarov, Aleksei and Gaja Jarosz. 2017. “Learning Parametric Stress without Domain-Specific Mechanisms.” In Karen Jesney, Charlie O’Hara, Caitlin Smith, and Rachel Walker (eds.), *Proceedings of the 2016 Meeting on Phonology*, Linguistic Society of America.

References

- Nazarov, Aleksei and Joe Pater. 2017. “Learning opacity in Stratal Maximum Entropy Grammar.” *Phonology* 34:2, 299-324.
- Nosofsky, R.M. 1989. Further tests of an exemplar-similarity approach to relating identification and categorization. *Perception and Psychophysics* 45: 279-290.
- Nouveau, Dominique. 1994. *Language Acquisition, Metrical Theory, and Optimality. A Study of Dutch Word Stress*. Utrecht: OTS Publications.
- van Oostendorp, Marc. 1997. “Lexical Variatie in Optimaliteitstheorie.” *Nederlandse Taalkunde* 2: 133–54.
- Van Oostendorp, Marc. 2012. “Quantity and the Three-Syllable Window in Dutch Stress.” *Language and Linguistics Compass* 6 (6): 343–58. doi:10.1002/lnc3.339.
- Osadcha, Iryna. 2014. *On Nominal Stress in Ukrainian*. Ms. (Generals Paper), University of Toronto.
- Pater, Joe. 2000. “Non-Uniformity in English Secondary Stress: The Role of Ranked and Lexically Specific Constraints.” *Phonology* 17 (2): 237–74.
- Pater, Joe. 2010. “Morpheme-Specific Phonology: Constraint Indexation and Inconsistency Resolution.” In *Phonological Argumentation: Essays on Evidence and Motivation*, edited by Steve Parker, 123–54. London: Equinox Press.
- Pater, Joe, Karen Jesney, Robert Staubs, and Brian Smith. 2012. “Learning Probabilities over Underlying Representations.” In *Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology*, 62–71. Association for Computational Linguistics.

References

- Pearl, Lisa. 2007. *Necessary Bias in natural Language Learning*. PhD dissertation, U of Maryland.
- Pearl, Lisa. 2011. “When Unbiased Probabilistic Learning is Not Enough: Acquiring a Parametric System of Metrical Phonology.” *Language Acquisition* 18(2).87-120.
- Potts, C., J. Pater, K. Jesney, R. Bhatt and M. Becker. 2010. “Harmonic Grammar with Linear Programming: From linear systems to linguistic typology.” *Phonology* 27: 77-117.
- Prince, Alan S, and Bruce Tesar. 2004. “Learning Phonotactic Distributions.” In *Fixing Priorities: Constraints in Phonological Acquisition*, edited by René Kager, Joe Pater, and Wim Zonneveld, 245–91. Cambridge, Massachusetts: Cambridge University Press.
- Rasin, Ezer, Iddo Berger, Nur Lan, and Roni Katzir. to appear. “Rule-based Learning of Phonological Optionality and Opacity.” To appear in *Proceedings of NELS 48*.
- Rubach, Jerzy & Geert E. Booij. 1985. ‘A grid theory of stress in Polish.’ *Lingua* 66. 281–319.
- Stankiewicz, Edward. 1993. The Accentual Patterns of the Slavic Languages. Stanford, CA: Stanford UP.
- Staubs, R.D. 2014. *Computational modeling of learning biases in stress typology*. PhD dissertation, University of Massachusetts Amherst.
- Staubs, Robert, and Joe Pater. 2016. ‘Learning Serial Constraint-Based Grammars.’ In *Harmonic Grammar and Harmonic Serialism*, edited by John McCarthy and Joe Pater. London: Equinox Press.
- Tesar, Bruce. 1995. *Computational Optimality Theory*. Doctoral dissertation, University of Colorado.

References

- Tesar, Bruce. 1998. ‘An iterative strategy for language learning.’ *Lingua* 104. 131–145.
- Tesar, Bruce. 2014. *Output-Driven Phonology: Theory and Learning*. Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.
- Tesar, Bruce, and Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.
- Tesar, Bruce, John Alderete, Graham Horwood, Nazarré Merchant, Koichi Nishitani & Alan Prince. 2003. “Surgery in language learning.” *WCCFL* 22. 477–490.
- Temkin–Martínez, M. 2010. “Sources of Non-Conformity in Phonology: Variation and exceptionality in Modern Hebrew Spirantization.” PhD dissertation, USC.

Appendix

Ranking probability computation

- p(A >> B|data) is proportional to

    p(data|A >> B) *(estimate by sampling)*

    x

    p(A >> B) *(look up in grammar)*

(by Bayes' Law)

Choosing index

- When inconsistency is detected, select **target constraint pair**:
  - Pair with **maximum deviance** from the trend
  - Deviance: sum, for all inconsistent words *V* for constraint pair {A,B}, their absolute difference from the trend in terms of preferring A >> B

$$\text{Deviance} = \Sigma_{v \in V\{A,B\}} |p(A >> B|v) - p(A >> B|\text{data})|$$

Choosing index

- Reason for choosing pair with maximal deviance:
  - It takes a while for the learner to process/absorb the effects of a new indexed constraint
  - Therefore, the same indexed constraint will be pushed until the pair that gave rise to it is no longer the one with maximal deviance (which is a sign that the lexical constraint has been absorbed into the grammar: the exceptions have been accounted for)

Inducing constraints

- Within target pair {C, D}, **target constraint** is the one sent downward by the trend
  - If p(**C >> D**|data) > 0.5, target constraint is **D**
  - If p(**D >> C**|data) > 0.5, target constraint is **C**

- If grammar has C >> D, exceptional pattern should be D >> C
- Represented as D_i >> C >> D

Inducing constraints

- If target constraint **D** does not yet have an indexed variant:
  - Add D_i to constraint set, with p(D_i >> α) for every constraint α
  - Add index *i* to all words in *V*{C,D}, i.e., the ones inconsistent with C >> D

- If target constraint **D** does have an indexed variant D_i:
  - Add index *i* to all words in V{C,D} that did not have it before