New approaches to end weight

Jason Grafmiller
Department of Linguistics
Stanford University

Stephanie Shih
Department of Linguistics
Stanford University

Department of Linguistics
University of California, Berkeley

[Variation & Typology: New Trends in Syntactic Research | 26 August 2011]

the Principle of End Weight

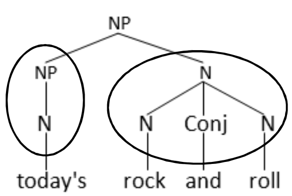
- “Phrases are presented in order of increasing weight.” (Wasow 2002: 3; following Behagel 1909; Quirk et al. 1985)
  [1] *peas and carrots > carrots and peas*
  [2] *the attitude of people who are really into classical music and feel that if it’s not seventy-five years old, it hasn’t stood the test of time >*
  people who are really into classical music and feel that if it’s not seventy-five years old, it hasn’t stood the test of time’s attitude
- Facilitates planning, production, and parsing
- Peripheral weight effects vary cross-linguistically (e.g. Yamashita 2001)

What is ‘weight’?

Syntax

Weight as syntactic complexity

- heavy constituents are structurally more complex
- Number of syntactic nodes (e.g., Ferreira 1991; Hawkins 1994)



What is ‘weight’?

Processing load

Weight as structural integration cost

- heavy constituents require more computational effort
- Cost of relating an input into a projected structure depends on intervening computations
- Dependency Locality Theory (Gibson 2000; Temperley 2007)
  - Each new referent (discourse new NP or finite verb) adds to integration cost.

What is 'weight'?

Phonology

Weight as phonological complexity

- Heavy constituents have complex prosodic properties
- Number of primary stressed syllables (Anttila et al. 2010; following Selkirk 1984; Zec and Inkelas 1990)

Weight as phonological 'weight'

- Number of syllables (Benor and Levy 2006; McDonald et al. 1993; a.o.)

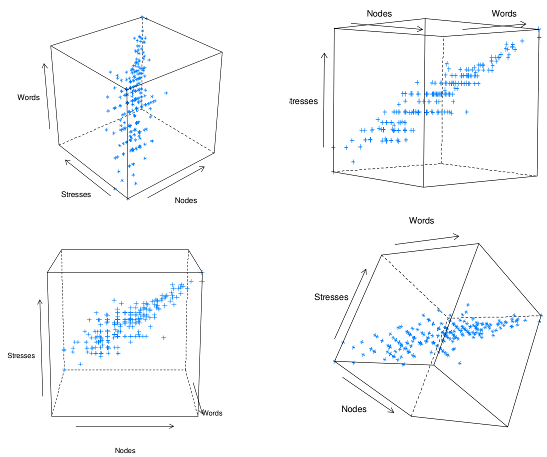
What is 'weight'?

Word count

Weight as length (word count)

- Many studies have used word count as a proxy for other weight factors (e.g, Wasow 2002; Szmrecsányi 2004; Bresnan and Ford 2010)
- Correlated with many other measures

High correlation of factors



Research Questions

[1] What is end weight?

- Most corpus-based studies of syntactic alternations focus on syntactic/processing weight
- Phonological weight hasn't been studied in the same way (cf., Anttila et al. 2010)
- Multiple theories of weight are rarely evaluated concurrently on the same data (cf., Szmrecsányi 2004)

[2] Methodological question:

- What is the best way to investigate and evaluate highly correlated variables?

The Data

- Two constructions in spoken American English
  (Switchboard Corpus; Godfrey and McDaniels 1992)

[1] Genitive Alternation (Shih et al., to appear)
- *'s*-genitive ~ *of*-genitive
- e.g., *the car's wheel ~ the wheel of the car*

[2] Dative Alternation (Bresnan et al. 2007)
- double object ~ prepositional dative (*to*)
- e.g., *give the dog the bone ~ give the bone to the dog*

Weight measures investigated

End Weight
Random Forests
Model Averaging
Discussion

- Syntactic nodes
- Referents (discourse new)
- Words
- Syllables
- Primary stressed syllables

Genitives model

End Weight
Random Forests
Model Averaging
Discussion

- 663 *of*-genitives + 460 *s*-genitives = 1123 total
- Control Predictors: Possessor animacy, final sibilancy, rhythm
  (Shih et al., to appear)
- Comparative weight (Bresnan and Ford 2010)

Comparative weight = log(possessor weight) – log(possessum weight)

s-genitive favored ←――――――― 0 ―――――――→ *of*-genitive favored

– 0 +

(*Referent counts were not log-transformed)

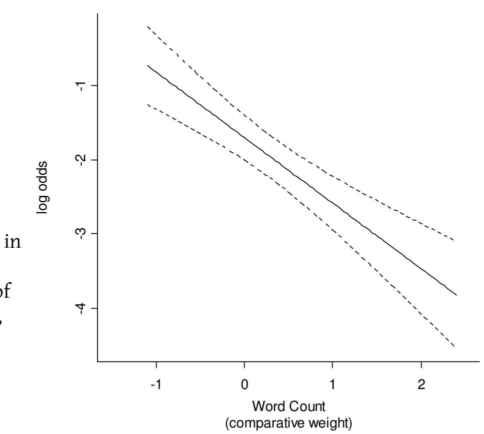
Genitives:

Heavy possessors favor *of*-gen

End Weight
Random Forests
Model Averaging
Discussion

- Higher log odds value = higher *s*-genitive likelihood
- Lower log odds value = higher *of*-genitive likelihood

➢ As the number of words in the possessor increases relative to the number of words in the possessum, an *of*-genitive becomes more likely.



Genitives:

Individual Regression Analysis

- Nodes
  - β = - 1. 234; z = -6.67; *p* < 0.000 (***)
- Words
  - β = - 0.884; z = -5.50; *p* < 0.000 (***)
- Referents
  - β = - 0.563; z = -3.71; *p* < 0.001 (**)
- Primary Stresses
  - β = - 0.525; z = -3.44; *p* < 0. 001 (**)
- Syllables
  - β = - 0.412; z = -3.42; *p* < 0.001 (**)

Datives Model

- 227 double objects + 183 prepositionals = 410 total
- Control Predictors: (Bresnan et al. 2007; Bresnan and Ford 2010)
  - Fixed effects: animacy of recipient, accessibility of recipient and theme, definiteness of recipient and theme
  - Random effect: verb
- Comparative weight (Bresnan and Ford 2010)

Comparative weight = log(recipient weight) – log(theme weight)

double object favored ← – ... 0 ... + → prepositional object favored

(*Referent counts were not log-transformed)

Datives:

Individual Regression Analysis

- Nodes
  - β = 1.312; z = 6.685; *p* < 0.000 (***)
- Words
  - β = 1.186; z = 6.877; *p* < 0.000 (***)
- Primary Stresses
  - β = 1.013; z = 6.304; *p* < 0. 000 (***)
- Syllables
  - β = 1.040; z = 6.086; *p* < 0.000 (***)
- Referents
  - β = 0.207; z = 1.305; *p* = .19

Methodology

- Controlled for other known variables influencing construction choice (Shih et al., to appear; Hinrichs and Szmrecsányi 2007; Bresnan et al. 2007; Bresnan and Ford 2010; a.o.)
- Conditional Random Forest Analysis (Hothorn et al. 2010; Strobl et al. 2009a; 2009b; a.o.)
  - Non-parametric, CART-based ensemble model
  - Conditional permutation accuracy variable importance measures
- Multimodel Inference (Model Averaging) (Burnham and Anderson 2002; 2004)
  - Full subset regression analysis of five weight predictors (32 models total)
  - Derived variable importance probabilities through comparative model weighting based on Akaike Information Criterion (AIC)

Random Forests

- Ensemble of classification or regression trees
  - random subsamples of data for each CART
  - random restricted set of predictor variables in each tree split

= diverse trees: variables have a greater chance of being included in the model when a stronger competitor is not.

- Detects contributions and behavior of predictor variables otherwise masked by competitors
- Suited to datasets with complex interactions and highly correlated predictor variables (Strobl et al. 2008; 2009a; 2009b)
- Greater accuracy than simple/mixed effect regression models for our data.

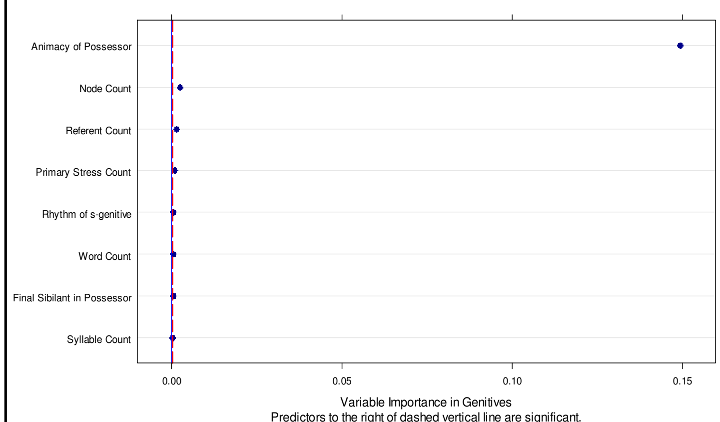
Conditional Variable Importance

- Conditional permutation accuracy
  - values of a predictor variable are randomly shuffled, breaking original association with response variable
  - the difference of model accuracy before and after shuffling tells us how important a variable is to the overall model
- Covers the individual impact of each predictor in the random forest model.

Model Parameters

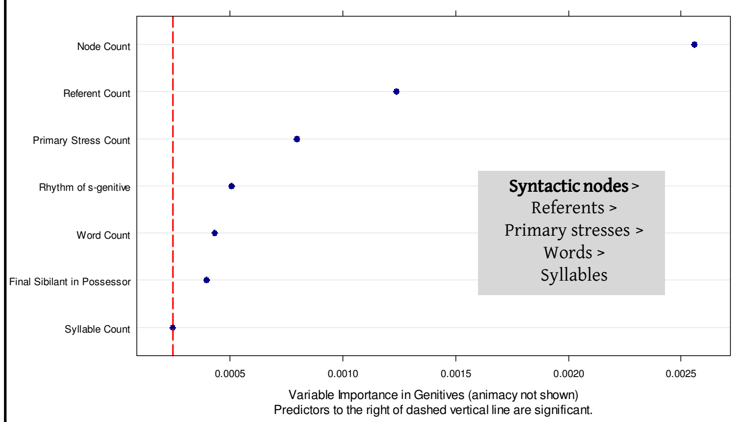
- Model parameters
  - Genitives: ntree = 2000, mtry = 3
  - Datives: ntree = 8000, mtry = 3
- Model stability verified on at least two random seeds.

Genitives Variable Importance



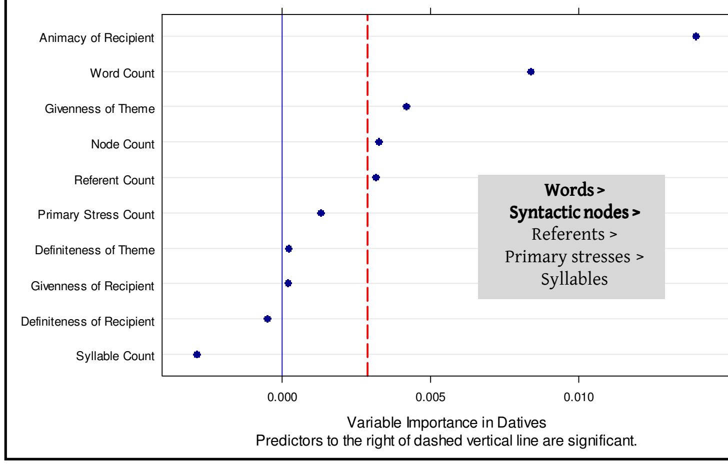
Random Forests | Results

Genitives Variable Importance



Random Forests | Results

Datives Variable Importance



Model Averaging using AIC

- Model averaging does not assume a single "best" model.
  - Inferences better reflect uncertainty in parameter estimates
- Provides variable importance ranking based on evidence from all possible combinations of predictors
- The Akaike Information Criterion (AIC) is an *estimate* of the distance from a fitted model g to unknown reality f.

AIC = -2 log(likelihood) + 2k

Model Averaging using AIC

- In a set of models, we can compare AIC values by scaling them:

$$\Delta_i = AIC_i - AIC_{min}$$

  - Models with $\Delta_i \leq 2$ have strong support
  - Models with $\Delta_i > 10$ have little support
- The Akaike weight w_i denotes the probability that a model i is the best approximation of the data in the set of models r.

$$w_i = e^{(-0.5\,\Delta_i)} / \Sigma e^{(-0.5\,\Delta_r)}$$

Genitives:

Model Averaging Results

End Weight
Random Forests
Model Averaging
Discussion

Weight measure in model	LogLik	AIC	Δ_i	w_i
Nodes, Stress, Refs, Words	-391.84	799.67	0.00	0.38
Nodes, Stress, Refs	-393.36	800.71	1.03	0.22
Nodes, Stress, Refs, Words, Syll	-391.82	801.63	1.96	0.14
Nodes, Stress, Refs, Syll	-393.35	802.71	3.031	0.08
…				
Stresses	-415.53	841.08	41.40	3.8 e-10
Syllables	-415.75	841.50	41.83	3.1 e-10
None	-421.64	851.28	51.61	2.34 e-12

Genitives | Model Averaging

Variable Importance

End Weight
Random Forests
Model Averaging
Discussion

- Importance of individual variables is calculated by adding the weights of all the models containing the variable.

Weight measure	Variable Importance (Cumulative Prob)
Nodes	0.996
Stresses	0.984
Referents	0.839
Words	0.610
Syllables	0.273

Datives:

Model Averaging Results

End Weight
Random Forests
Model Averaging
Discussion

Weight measure in model	LogLik	AIC	Δ_i	w_i
Words	-190.89	397.77	0.00	0.16
Nodes	-191.29	398.58	0.81	0.10
Words, Nodes	-190.47	398.94	1.17	0.09
Words, Stress	-190.49	398.99	1.22	0.08
…				
Stresses	-190.01	402.03	4.26	0.019
Syllables	-191.09	402.19	4.42	0.017
None	-191.11	402.21	4.44	0.017

Datives | Model Averaging

Variable Importance

End Weight
Random Forests
Model Averaging
Discussion

Weight measure	Variable Importance (Cumulative Prob)
Words	0.716
Nodes	0.541
Stresses	0.337
Syllables	0.281
Referents	0.275

Results

Model Averaging vs. Random Forests

	Model Averaging	Random Forest
Genitives	Syntactic nodes > Primary stresses > Referents > Words > Syllables	Syntactic nodes > Referents > Primary stresses > Words > Syllables

Results

Model Averaging vs. Random Forests

	Model Averaging	Random Forest
Datives	Words > Syntactic nodes > Primary stresses > Syllables > Referents	Words > Syntactic nodes > Referents > Primary stresses > Syllables

Research Questions

[1] What is end weight?

→ What is the best measure of weight?

[2] Methodological questions:

→ What is the best way to investigate and evaluate highly correlated variables?

Discussion

Processing-based Weight Measures

Referents, in comparison to other measures, are not a reliable measure of weight.

= Non-given and definite nouns and finite verbs (Gibson 1998; 2000)

- What else can contribute to integration costs?

	the	*green*	*ball*
Gibson:			x = 1 new referent
alternatively:		x	x = 2 new referents

- Redefinition of 'referents' → content words?
  – Dependency Length Minimization (Temperley 2007, 2008; Gildea & Temperley 2010)

Discussion

Phonological Weight Measures

Below the prosodic hierarchy…
- Syllables: rank low as a good independent measure of weight in genitive and dative construction choice.

- Do possible phonetic correlates of weight or complexity play into end weight effects?
  - e.g., duration, complexity of segments, syllable weight or complexity of syllable structure? (e.g., Benor and Levy 2006)

Prosodic Weight Measures
- Primary stresses: high-ranking predictor in genitives.
- Prosodic theory of end weight (=number of primary stresses) is not entirely syntax-independent.
  - i.e., phonological words ≈ content word

Discussion

Syntactic Weight

End Weight
Random Forests
Model Averaging
Discussion

Syntactic Complexity (number of syntactic nodes)
- Consistently one of the highest ranking predictors.
- Highest ranking individual predictor for genitives.
- Second highest ranking for datives.

- Is ‘weight’ purely syntactic?
  - English binomial ordering studies: number of syllables affect ordering of nouns in binomial pairs (Wright et al. 2005; cf., McDonald et al. 1993; Benor & Levy 2006)

- At a higher-level domain (i.e., genitives, datives), syntactic complexity is one of the most salient manifestations of ‘weight’

- Also: possible confound between syntactic and prosodic complexity?

Discussion

Datives vs. Genitives

End Weight
Random Forests
Model Averaging
Discussion

	Model Averaging	Random Forest
Genitives	Syntactic nodes > Primary stresses > Referents > Words > Syllables	Syntactic nodes > Referents > Primary stresses > Words > Syllables
Datives	Words > Syntactic nodes > Primary stresses > Syllables > Referents	Words > Syntactic nodes > Referents > Primary stresses > Syllables

[Q]: What causes the apparent variation in variable importance between the genitive and dative constructions? Are different syntactic domains more sensitive to different components of weight?

Research Questions

End Weight
Random Forests
Model Averaging
Discussion

[1] What is end weight?

→ What is the best measure of weight?

[2] Methodological questions:

→ What is the best way to investigate and evaluate highly correlated variables?

Discussion | Methodology

Model Averaging vs. Random Forest

End Weight
Random Forests
Model Averaging
Discussion

	Model Averaging	Random Forest
Pros	• handles small *n*, large *p* • less likely to lead to spurious significance • better at handling collinearity than single regression models	• handles small *n*, large *p* • deals well with correlations and high-order interactions • shows independent effects of predictors • eliminates order effects in single CARTs • more accurate than parametric regression models
Cons	• not immune to harmful effects of collinearity (at the model level) • long computing time when more predictors are present	• difficult to see main effects • long computing load and time • permutation importance cannot yet handle NA data (a minor annoyance)

Discussion | Methodology

Model Averaging vs. Random Forest

End Weight
Random Forests
Model Averaging
Discussion

	Model Averaging	Random Forest
Genitives	Syntactic nodes > Primary stresses > Referents > Words > Syllables	Syntactic nodes > Referents > Primary stresses > Words > Syllables
Datives	Words > Syntactic nodes > Primary stresses > Syllables > Referents	Words > Syntactic nodes > Referents > Primary stresses > Syllables

Model averaging and random forests provide similar results in variable importance ranking.

Future directions

Weight beyond English

End Weight
Random Forests
Model Averaging
Discussion

- How do measures of weight generalize beyond English?
- Is there a better proxy for cross-linguistic syntactic complexity?
  - i.e., morphological complexity as weight?

Conclusion

End Weight
Random Forests
Model Averaging
Discussion

- Two statistical methods more resistant to collinearity:
  - Conditional random forest analysis
  - Information-theoretic (AIC) model averaging
- Two alternations in spoken American English:
  - Genitives | Datives
- Tested syntactic, processing, and phonological measures of 'weight.'
  - Syntactic nodes (syntactic complexity)
  - Referents (processing dependencies, DLT)
  - Primary stress (phonological complexity)
  - Syllables (phonological weight)
  - Words (commonly used weight proxy)

End Weight
Random Forests
Model Averaging
Discussion

Conclusion

- Importance of weight measures varies by construction.
  - Genitives: syntactic nodes and primary stresses
  - Datives: words and syntactic nodes
- Syntactic complexity is a highly reliable predictor in both constructions.
- Syllable and referent counts as measures of weight are not reliable.
  - Phonological weight may capture weight effects only in lower-level ordering phenomena, e.g. binomial pairs

Thank you!

Thank you to Arto Anttila for the initial push and subsequent support on this project. Many thanks must go to Joan Bresnan for invaluable advice, discussion, and patience. Acknowledgements also to Matthew Adams, Susanne Gahl, Sharon Inkelas, Victor Kuperman, Beth Levin, Robin Melnick, Daphne Theijssen, Benedikt Szmrecsányi, Michael Wagner, Tom Wasow, Christoph Wolk, the audience at the *Development of Syntactic Alternations* workshop and the LSA Annual Meeting 2011, and the members of Quorum (UC Berkeley), CrISP (Stanford/UC Santa Cruz), and the Edinburgh Phonology Workshop (U of Edinburgh) for further discussion, support, skepticism, questions, and statistical aid.

This material is based in part upon work supported by the National Science Foundation under Grant Number BCS-1025602 to Stanford University for the research project "The Development of Syntactic Alternations" (PI Joan Bresnan). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Contact:

Jason Grafmiller
jasong1@stanford.edu

Stephanie Shih
stephsus@stanford.edu

Selected References

Anttila, Arto; Matthew Adams; and Michael Speriosu. 2010. The role of prosody in the English dative alternation. *Language and Cognitive Processes*.

Behagel, O. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen*. 25: 110-142.

Benor, Sarah Bunin and Roger Levy. 2006. The Chicken of the Egg? A Probabilistic Analysis of English Binomials. *Language*. 82(2): 233-278.

Bresnan, Joan; Anna Cueni; Tatiana Nikitina; and R. Harald Baayen. 2007. Predicting the Dative Alternation. in G. Bouma,; I. Kraemer; and J. Zwarts (ed). *Cognitive Foundations of Interpretation*. Royal Netherlands Academy of Science. 69-94.

Bresnan, Joan and Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language*. 86(1): 168-213.

Burnham, Kenneth P. and David R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, second edition. New York, NY: Springer Science+Business Media, Inc.

Burnham, Kenneth P. and David R. Anderson. 2004. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociolinguistic Methods Research*. 33: 261-304.

Carnegie Mellon University Pronouncing Dictionary. <http://www.speech.cs.cmu.edu/cgi-bin/cmudict> accessed 2008.

Comrie, Bernard. 2003. On explaining language universals. in M. Tomasello (ed). *The New Pyschology of Language: Cognitive and Functional Approaches to Language Structure*. Mahwah, NJ: Erlbaum. 195-210.

Ferreira, Fernanda. 1991. Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*. 30: 210-233.

Gildea, Daniel and David Temperley. 2010. Do grammars minimize dependency length? *Cognitive Science*. 34: 286-310.

Gibson, Edward. 1998. Linguistic Complexity: locality of syntactic dependencies. *Cognition*. 68: 1-76.

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. in Y. Miyashita; A. Marantz; and W. O'Neil (ed). *Image, Language, Brain*. Cambridge, MA: MIT Press. 95-126.

Godfrey, J. Holliman and J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. *Proceedings of ICASSP-92*. 517-520.

Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Harrell, Frank E, Jr. 2009. Design: Design Package. R package version 2.3-0. http://CRAN.R-project.org/package=Design

Hinrichs, Lars and Benedikt Szmrecsányi. 2007. Recent changes in the function and frequency of standard English genitive constructions: a multivariate analysis of tagged corpora. *English Language and Linguistics*. 11(3): 437-474.

Hothorn, Torsten; Kurt Hornik; Carolin Strobl; and Achim Zeileis. 2010. A Laboratory for Recursive Partytioning. <http://cran.r-project.org/web/packages/party/party.pdf>

Malkiel, Yakov. 1959. Studies in irreversible conjunctions. *Lingua*. 8: 113-160.

Selected References, cont.

McDonald, Janet L.; Kathryn Bock; and Michael H. Kelly. 1993. Word and World Order: Semantic, Phonological, and Metrical Determinants of Serial Position. *Cognitive Psychology*. 25: 188-230.

Quirk, Randolph; Sidney Greenbaum; Geoffrey Leech; and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.

R Development Core Team. 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>

Selkirk, Elisabeth O. 1984. *Phonology and Syntax: the Relation between Sound and Structure*. Cambridge, MA: MIT Press.

Shih, Stephanie. 2010. Random forests for classification trees and categorical dependent variables: an informal quick start R guide. MS. <www.stanford.edu/~stephsus/R-randomforest-guide.pdf>

Shih, Stephanie; Jason Grafmiller; Richard Futrell; and Joan Bresnan. to appear. Rhythm's role in predicting genitive construction choice in spoken English. In Vogel, R. and R. Van de Vijver (ed). *Rhythm in phonetics, grammar, and cognition*.

Strobl, Carolin; Anne-Laure Boulesteix; Thomas Kneib; Thomas Augustin; and Achim Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics*. 9:307.

Strobl, Carolin; Torsten Hothorn; and Achim Zeileis. 2009. Party on! A new, conditional variable-importance measure for random forests available in party package. *The R Journal*. 1/2: 14-17.

Strobl, Carolin; James Malley; and Gerhard Tutz. 2009. An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random Forests. *Psychological Methods*. 14(4): 323-348.

Szmrecsányi, Benedikt. 2004. On operationalizing syntactic complexity. *Journées internationals d'Analyse statistique des Données Textualles*. 7: 1031-1038.

Szmrecsányi, Benedikt. 2008. Probabilistic determinants of genitive variation in spoken and written English: a multivariate comparison across time, space, and genres. in T. Nevalamen; I. Taavitsamen; P. Pahta; and M. Korhonen (ed). *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*. Amsterdam: Benjamins.

Temperley, David. 2007. Minimization of dependency length in written English. *Cognition*. 105: 300-333.

Temperley, David. 2008. Dependency length minimization in natural and artificial grammars. *Journal of Quantitative Linguistics*. 15: 256-282.

Wasow, Tom. 2002. *Postverbal Behavior*. Stanford, CA: CSLI Publications.

Wright, Saundra K.; Jennifer Hay; and Tessa Bent. 2005. Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics* 43(3): 531-561.

Yamashita, Hiroko and Franklin Chang. 2001. "Long before short" preference in the production of a head-final language. *Cognition*. 81: B45-B55.

Zec, Draga and Sharon Inkelas. 1990. Prosodically Constrained Syntax. in S. Inkelas and D. Zec (ed). *The Phonology-Syntax Connection*. Stanford, CA: Center for the Study of Language and Information.

Zubizarreta, Maria Luisa. 1998. *Prosody, Focus, and Word Order*. Cambridge, MA: The MIT Press.