15th International Morphology Meeting
Vienna, February 9-12

Explaining English Compound Stress Analogically

Sabine Arndt-Lappe, Universität Siegen,
Melanie J. Bell, Anglia Ruskin University

Objectives

- empirical
  - corpus study
  - test how compound stress assignement in the corpus can be modelled with the help of a computational analogical model (AM, Skousen 1989, 1992, Skousen et al. 2002 et seq.)

- theoretical
  - English compound stress functions analogically
  - AM is able to model and predict the interplay of different types of factor influencing compound stress
  - An analogical approach to compound stress assignment is superior to a categorical, rule-based approach

English compound stress

- variation between two possibilities: left stress, right stress
- left-stressed examples

ápple juice	*wíndow washer*	*Óxford street*
téabag	*cátfood*	*chéese man*
méal time	*scíence group*	

- right-stressed examples

apple píe	*glass dóor*
car rádio	*gold éarring*
easter hóliday	*kitchen dóor*

- predictive factors include constituent family and semantic properties

(cf. e.g. Plag 2006, Plag, Kunter, Lappe 2007; Plag, Kunter, Lappe, Braun 2008, Arndt-Lappe 2011a, Bell 2011)

Semantic effects

- Certain semantic relations are right-stressed (e.g. ‘locative’ compounds, *Boston hárbour*).
- Certain semantic classes of constituents trigger right stress (e.g. substance nouns as N1, *silk shírt*).
- Lexicalised semantics goes together with left stress (*sílk worm*).

Constituent family effects

Stress is assigned by analogy with compounds in the mental lexicon that share either N1 or N2.

Óxford	*Street*	*Oxford*	*Róad*	*state*	*administrátion*
Régent	*Street*	*Mill*	*Róad*	*state*	*búdget*
Hárley	*Street*	*Upland*	*Róad*	*state*	*bénefits*
...	*Street*	*...*	*Róad*	*státe*	*house*
				state	*fúnds*
				state	*...*
100 % left		0 % left		10 % left	

'constituent family stress bias'

Problems for accounts of the variation

- Variation is systematic and productive, but cannot be captured in terms of deterministic rules

- Not all compounds that have a particular N1or N2, and not all compounds with a certain semantics have the predicted stress pattern (e.g. 'made of'-relation: *apple píe* vs. *ápple juice*)

- Interaction between rather local effects (based on constituent family, affecting few compounds) and more general effects (based on semantics) is unclear.

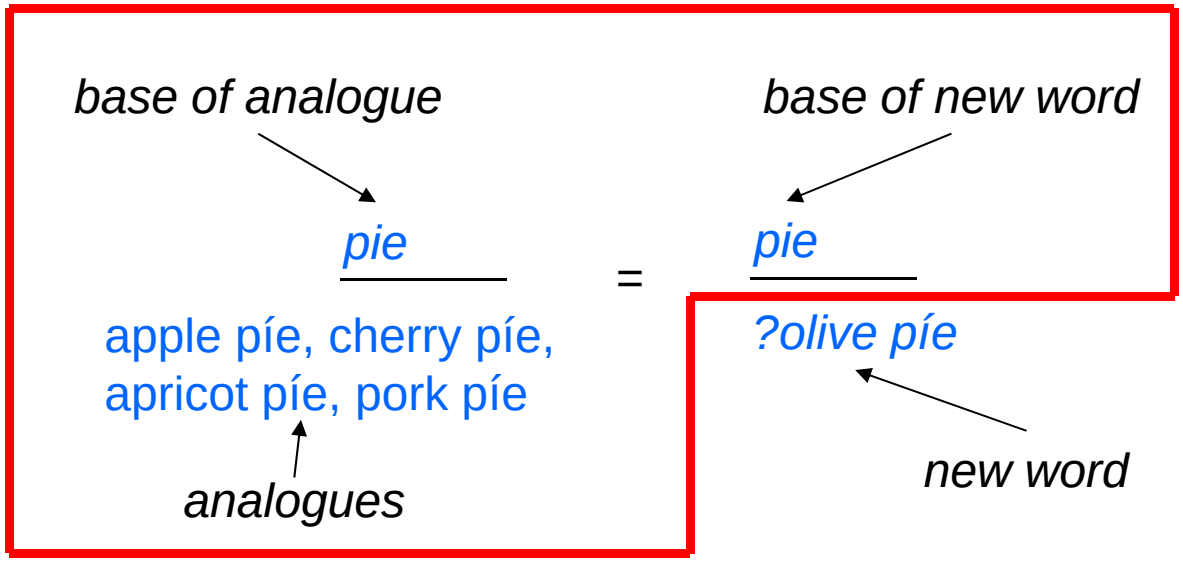
Proposal

- an analogical view of word-formation solves many of the problems of approaches involving deterministic rules
- especially:
  - variability / leakage of rules
  - interaction of local and general effects (an aspect of productivity)
- The criticism against analogical views of word-formation that they are
  - vague
  - non-predictive, and
  - not testable

  is not true for some computational analogical models (in particular: AM, Skousen 1989, 1992, Skousen et al. 2002).

The hypothesis:
Compound stress is assigned by analogy

lexemes in the Mental Lexicon



necessary assumption: analogues can be sets of words

computational analogical models

- can operationalise the notions of 'similarity' and 'sets of analogues' (alternative term: 'analogical sets')
- can model variation, categorial behavior and leakage
- analogy becomes predictive and predictable
- the theory becomes testable
- algorithm:
  - AM(L) (Skousen 1989, 1992 et seq.; an alternative: TiMBL, Daelemans et al. 1999 et seq.)

The basic architecture of an analogical model

exemplars in the lexicon

feature 1	feature 2	feature 3	feature 4	stress
afternoon	break	no	temporal	right
cat	food	no	no	left
chocolate	raisin	no	no	right
coffee	jar	no	no	left
sports	center	no	no	left

set of analogues / analogical set

feature 1	feature 2	feature 3	feature 4	stress
apple	pie	material	no	right
cherry	pie	material	no	right
pork	pie	material	no	right
chicken	burger	material	no	left
olive	oil	material	no	right

4x right, 1x left

new word

feature 1	feature 2	feature 3	feature 4	stress
olive	pie	material	no	???

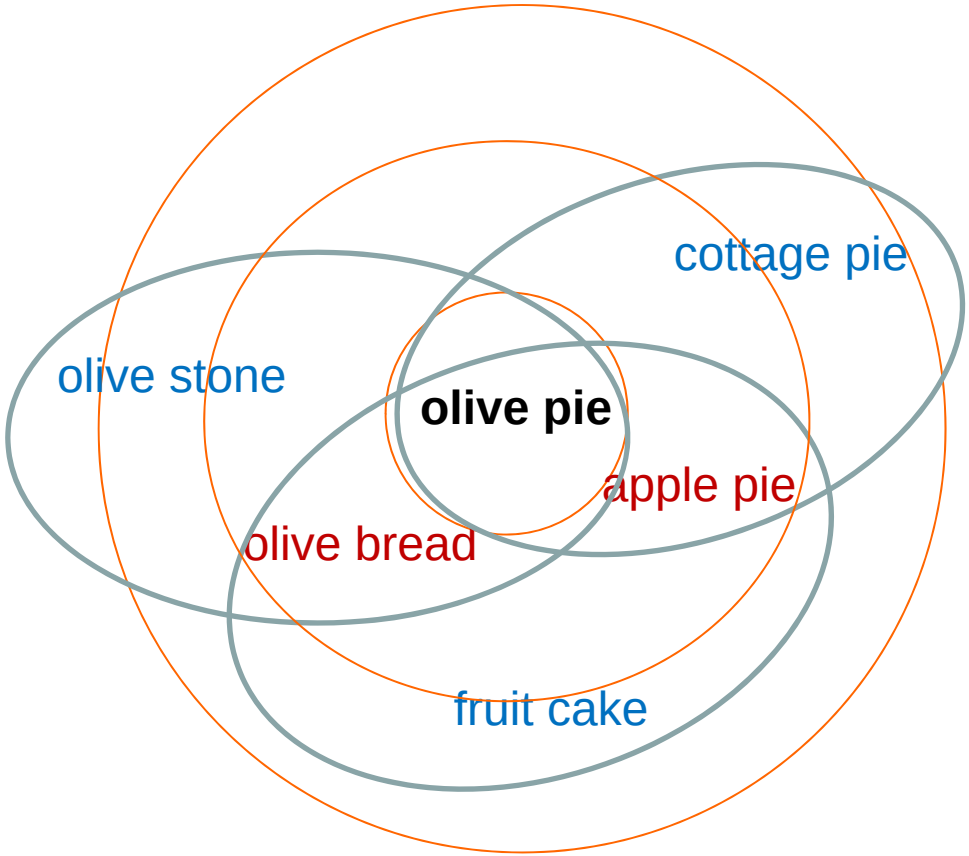
stress: right (majority choice)

Computing Analogical Sets

the similarity space

different degrees

different dimensions

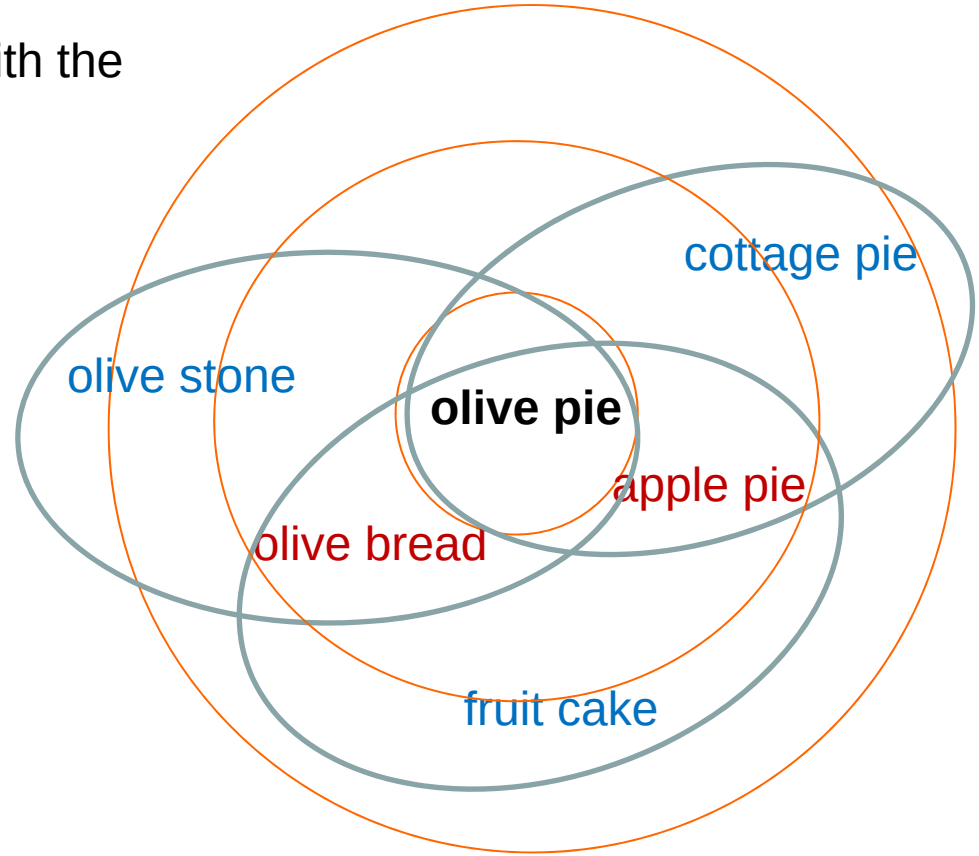


Computing Analogical Sets

the similarity space

different degrees

AM starts with the most similar exemplars.



different dimensions

Along all dimensions, AM tries to include more distant exemplars.

It does so if the more distant exemplars behave like the more similar group w.r.t. stress assignment. => minimised uncertainty

Methodology

Data: NN constructs

- 406 compounds extracted from the British National Corpus, for which stress was produced and rated consistently across all 4 recordings done for Bell (2011), which have a constituent family for N1 or N2 or both

of these,

- 241 are left-stressed, 165 are right-stressed

Coded features

- N1, N2 (in spelling)
- semantic properties and relations found relevant in Bell (2011), Plag et al. (2007, 2008)

Setup

'leave-one-out', corpus is tested on itself

AM experiment – overall performance

features given as information source	**F-score$_{left}$* (% correct predictions)**	**F-score$_{right}$ (% correct predictions)**	**F-score$_{average}$ (% correct predictions)**
constituent family & semantic features	0.94 (95%)	0.92 (90%)	0.93 (93%)

*cf. Daelemans & Bosch 2005 for discussion of performance measures

⇒ AM predicts stress assignment correctly for 93% of the data!

⇒ Predictions are almost equally good for left and right stress!

How does AM do this?

three types of analogical set

small sets
(1-15 exemplars)

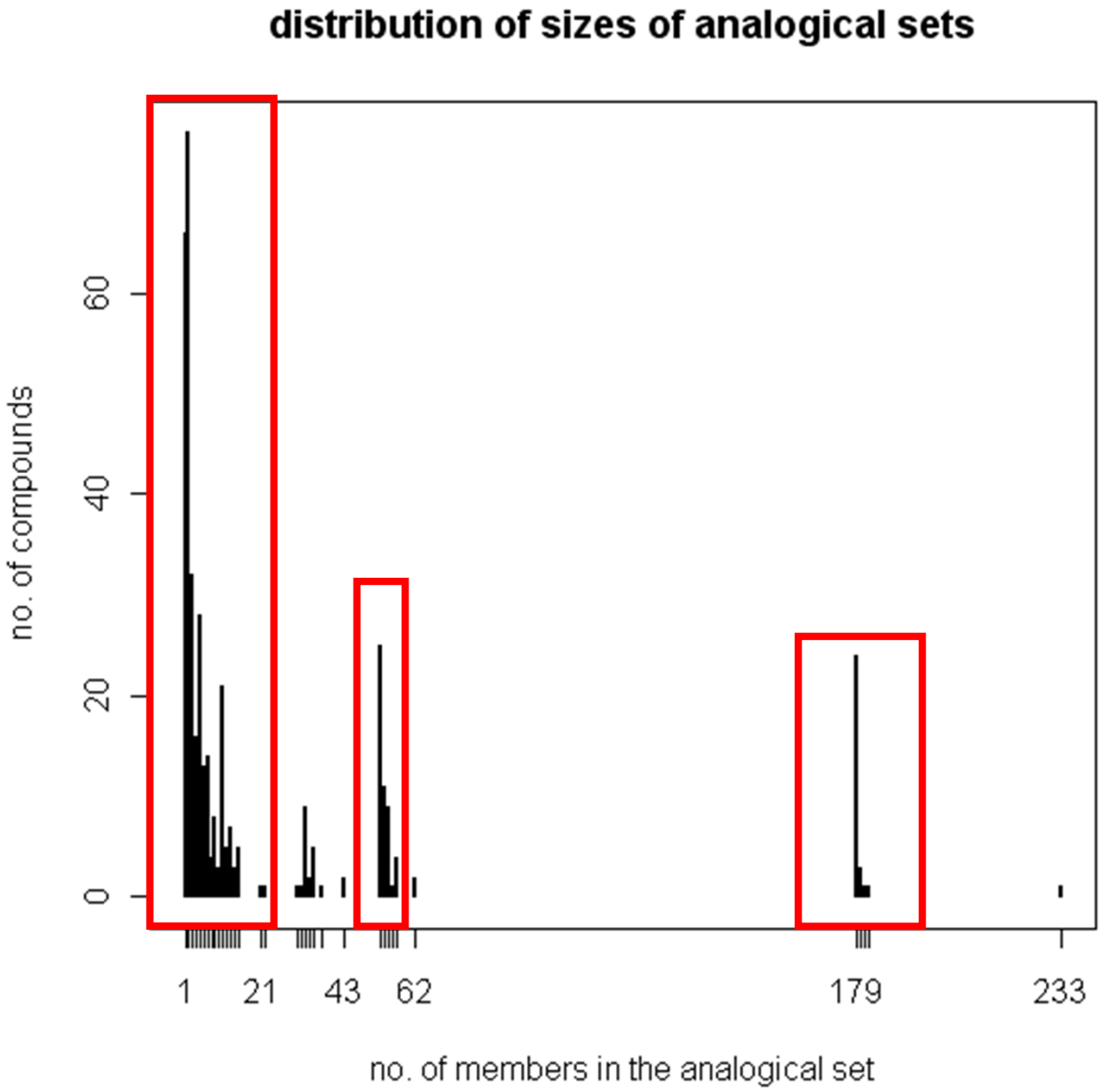
75% of the data

mid-size sets
(53-55 exemplars)

11% of the data

large sets
(179-183 exemplars)

7% of the data



Small analogical sets

- Clear effect of constituent family in all small sets (**all** sets are comprised of members of the constituent family)
- stress is sometimes left, sometimes right

examples, sets of two:

cat food	**cat** muck	convenience **food**
banana sandwich	lamb **sandwiches**	salmon **sandwiches**
football quiz	**football** nights	**football** party

⇒ **'local analogies'**

Mid-size analogical sets

- 53 exemplars that reoccur together in 45 analogical sets
- stress is right

examples:

baby boy	plastic wallet	bastard teacher	bitch teacher
cotton sheets	glass bowl	glass dish	gold jewellery
gold band	leather bags	toy cups	

⇒ **wider similarity space:** For these 45 new words, exemplars that share certain semantic features appear together in the analogical set. The fact that they reappear more often gives the impression of a 'rule'. But there is no rule.

⇒ **'less local analogies':** give the impression of a productive rule

What do compounds in the ‚mid-size sets' have in common?

- **semantic relations:** esp.
  - material — yes (34), no (19)
  - temporal — no
  - locative — no
  - copulative — no
- **semantic categories of N1:**
  - N1 is a time — no
  - N1 is a location — no
  - N1 is adjective-like — yes
  - N1 is a material — yes (48), no (5)
  - N1 is a social group — no

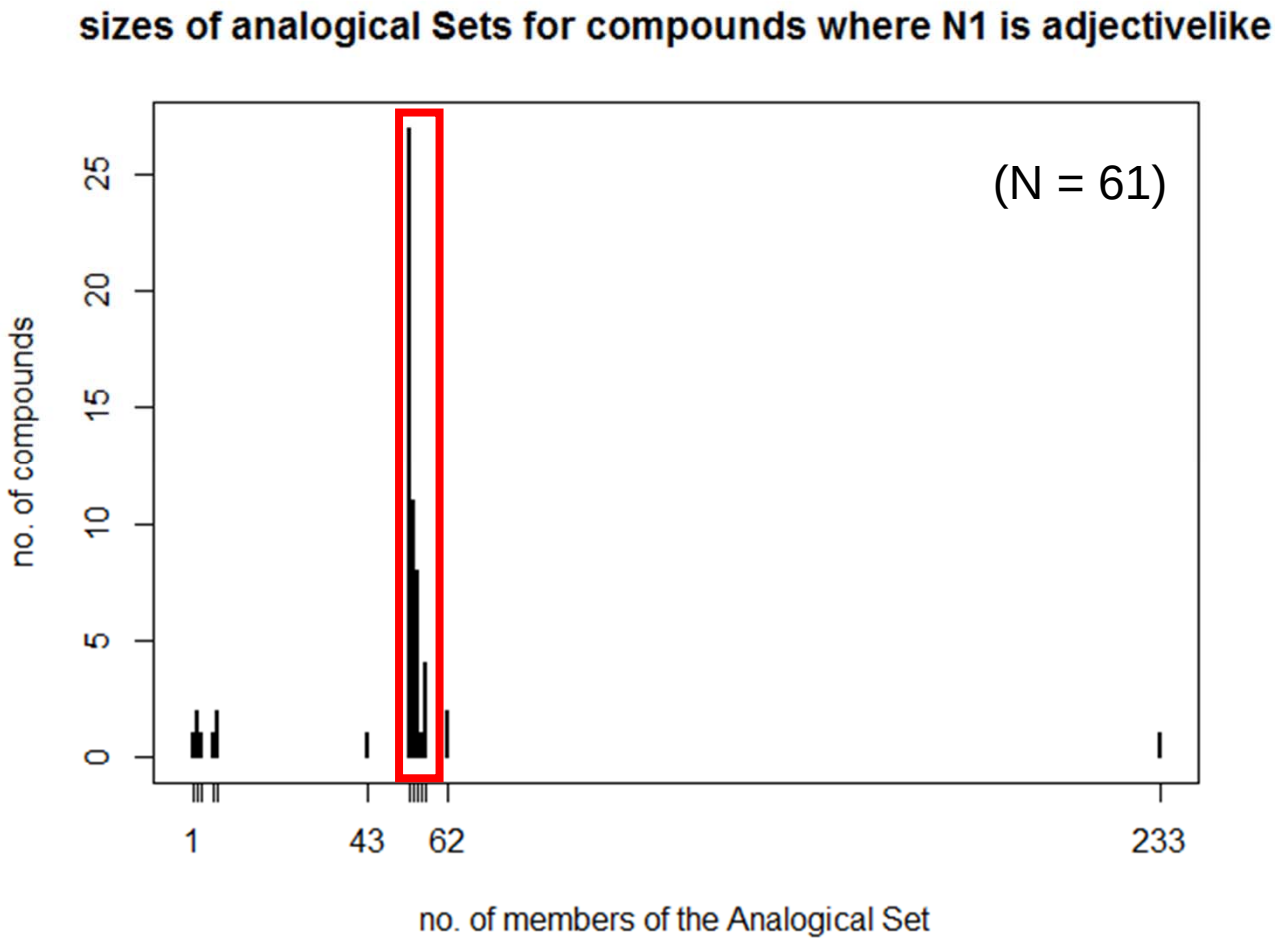
(categories & coding were taken from Bell 2011)

leakage in the mid-size set

- All 53 members of the mid-size Analogical Set share a semantic feature
  - ‚N1 has adjective-like qualities‘

BUT

- There are 61 ‚adjective-like‘ compounds in the dataset!‘



Large analogical sets

- 178 exemplars that reoccur together in 29 analogical sets
- stress is mostly left

examples

alarm business	antiques day	arm bands	art centre
assessment piece	attache case	baby stuff	banking job
begging bowl	bike things	bin day	bingo money

wide similarity space: analogical sets consist of exemplars NOT having a semantics that favours right stress.

The fact that they share 'negative' values for the relevant semantics gives the impression of a 'default situation'.
(cf., e.g., Derwing & Skousen 1989, Eddington 2000 for inflection)

the most **non-local type of analogy** conceivable

Compound stress assignment by analogy

- AM is highly successful in modelling compound stress on the basis of constituent family and semantics

- one single mechanism produces effects for which three different mechanisms are invoked in other frameworks:
  - different degrees of productivity
  - local, exceptional analogies
  - rules and default rules

- key: interplay of local and less local analogies

Summary & conclusion – empirical level

- AM was used to test how the challenges could be solved within an analogical theory of word-formation
  - very good overall predictive power
  - non-deterministic behaviour is expected in an analogical model
  - interaction of local and less local generalisations as well as default situations are epiphenomena of 'gang behaviour' among analogical sets

summary and conclusion – theoretical level

- AM constitutes a testable version of a theory of word-formation that is based on analogy.
- rules vs. analogy? Compound stress assignment provides evidence
  - against rules that are deterministic and independent of the lexicon
  - in favour of an approach that assumes no strict distinction between the lexicon and rules, and that allows for systematic variability. An analogical approach of the type implemented in AM is one plausible possibility.

Thank you very much for your attention!

References

Arndt-Lappe, Sabine. in press. 47. Word-formation and analogy. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-Formation - An International Handbook of the Languages of Europe*. Berlin: de Gruyter Mouton.

Arndt-Lappe, Sabine. 2011. Towards an exemplar-based model of stress in English noun–noun compounds. *Jounal of Linguistics*.

Bauer, Laurie. 1983. *English Word-Formation*. Cambridge: CUP.

Bauer, Laurie. 2001. *Morphological Productivity*. Cambridge: CUP.

Becker, Thomas. 1990. *Analogie und morphologische Theorie* (Studien zur theoretischen Linguistik). München: Fink.

Bell, Melanie. 2011. *The English Noun Noun Construct: Its Prosody and Structure*. Cambridge: Ph.D. dissertation, University of Cambridge.

Bell, Melanie & Ingo Plag. 2010. Informativeness is a determinant of compound stress in English: Ms, submitted for publication.

Blevins, James P. & Juliette Blevins (eds.) (2009). *Analogy in Grammar*. Oxford: OUP.

Daelemans, Walter & Antal d. van Bosch. 2005. *Memory-Based Language Processing*. Cambridge: CUP.

Daelemans, Walter, Jakub Zavrel, Ko van der Sloot & Antal van den Bosch. 1999 et seq. *TiMBL: Tilburg Memory Based Learner* : Available from http://ilk.uvt.nl/timbl/.

Derwing, Bruce I. & Royal Skousen. 1989. Morphology in the mental lexicon: a new look at analogy. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1989*, 55–71. Dordrecht: Foris.

Krott, Andrea, Harald R. Baayen & Rob Schreuder. 2001. Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics* 39. 51–93.

Plag, Ingo. 2006. The variability of compound stress in English: structural, semantic, and analogical Factors. *English Language and Linguistics* 10(1). 143–172.

Plag, Ingo. 2010. Compound stress assignment by analogy: the constituent family bias. *Zeitschrift für Sprachwissenschaft* 29(2). 243–282.

Plag, Ingo, Gero Kunter & Sabine Lappe. 2007. Testing hypotheses about compound stress assignment in English: a corpus-based investigation. *Corpus Linguistics and Linguistic Theory* 3(2). 199–233.

Plag, Ingo, Gero Kunter, Sabine Lappe & Maria Braun. 2008. The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language* 84(4). 760–794.

Skousen, Royal. 1989. *Analogical Modeling of Language*. Dordrecht: Kluwer.

Skousen, Royal. 1992. *Analogy and Structure*. Dordrecht: Kluwer.

Skousen, Royal, Deryle Lonsdale & Dilworth B. Parkinson (eds.) (2002). *Analogical Modeling*. Amsterdam / Philadelphia: John Benjamins.

Skousen, Royal & Theron Stanford. 2007. *AM::Parallel*. available from http://humanities.byu.edu/am