The Darkening of English /l/: a Stochastic Stratal OT Analysis

Danielle Turton

University of Manchester

19th Manchester Phonology Meeting, 19th-21st May 2011

Research Questions

1. How can we account for variable phonological processes that are morphosyntactically conditioned?
2. More specifically, how do differences in application rate across different morphosyntactic environments arise diachronically, and how are they generated synchronically?

Outline

English /l/

/l/-darkening in English is a variable phonological process.
In syllable based accounts it is said that:

- Light [l] occurs in canonical onsets: e.g. *love, light*
- Dark [ɫ] occurs in canonical codas: e.g. *dull, fall*

Dark [ɫ]:

- delayed tongue-tip gesture
- dorsal gesture precedes the coronal
- shows a greater retraction and lowering of the tongue dorsum

Sproat & Fujimura (1993)

Outline

Hayes (2000)
Boersma & Hayes (2001)

- Study of /l/ in American English (Hayes, 2000)
- Looking at the light and dark allophones of /l/
- 10 speakers of American English gave acceptability ratings on pronunciations produced with both light and dark /l/ in different environments
- Boermsa & Hayes (2001) used a logarithmic scale to convert Hayes's acceptability rating data into frequencies of variants.

Hayes (2000) Acceptability Rating Data

Figure 1: Taken from Hayes (2000:102)

(21)

	as light		as dark	
light, Louanne	✓	(1.20)	*	(5.83)
gray-ling, gai-ly, free-ly	✓	(1.57)	?	(3.34)
Mailer, Hayley, Greeley, Daley	✓	(1.90)	✓	(2.64)
mail-er, hail-y, gale-y, feel-y	?	(3.01)	✓	(2.01)
mail it	??	(4.40)	✓	(1.10)
bell, help	*	(6.60)	✓	(1.12)

Ratings: 1 meaning "sounds just right, normal in my dialect"
7 meaning "sounds awful."

Stochastic OT

In Stochastic OT, constraints are ranked on a continuous scale. They have ranges, not single points.

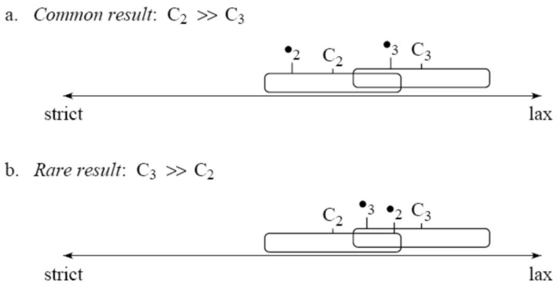


Figure 2: From Boersma & Hayes (2001:49)

The Gradual Learning Algorithm (GLA)

- A constraint ranking algorithm for learning Stochastic OT grammars.
- The GLA attempts to match the frequencies of variants in the training corpus.
- It calculates the average error per candidate, which shows how successfully it was able to match the data.
- Great advantage for Stochastic OT over other variation models (cf Antilla, 1997; Coetzee, 2004)
- Success demonstrated with data from Finnish, Ilokano and notably English /l/ in Boersma & Hayes (2001)

Outline

Boersma & Hayes (2001)

Boersma & Hayes's analysis of English /l/ (based on data from Hayes 2000) shows /l/-darkening is morphosyntactically conditioned:

Environment type	**% Light /l/**
light	99.956
free-ly	94.53
Hayley	76.69
mail-er	16.67
mail it	0.49
bell	0.0011

Table 1: Inferred frequencies of light /l/ in different morphosyntactic positions (Boersma & Hayes, 2001)

Apparent Overapplication?

- As expected, the data from Boersma & Hayes (2001) show that /l/ is almost obligatorily light in words such as *light* and obligatorily dark in words such as *bell*.
- However, the process of darkening seems to overapply in words such as *mailer* /maɪ.lə/ where the /l/ is in the onset, but is dark over 83% of the time.
- Additionally, the /l/ in *mail it* /maɪ. lɪt/ is in the onset but is dark almost obligatorily.
- This is evidence for the morphosyntactic conditioning of /l/-darkening.

Output-Output Correspondence (OOC)

- Morphosyntactically conditioned variation requires a theory of the phonology-morphosyntax interface, for which Hayes adopts Output-Output Correspondence (OOC).
- This is the idea that outputs of derived forms have a link with the output of the base on the surface.

Hayes uses this to say:

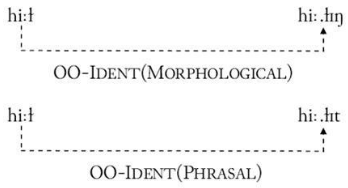


Figure 3: from Bermúdez-Otero, 2011:30

Outline

Output-Output Correspondence (OOC)

- However, Hayes's two OOC constraints predict an impossible dialect by factorial typology if: OO-IDENT(MORPHOLOGICAL) » OO-IDENT(PHRASAL)
- This ordering gives an unattested dialect displaying dark [ɫ] in *heal* and *healing*, but light [l] in *heal it*

*[hiːɫ, hiː.lɪt, hiːɫɪŋ]

The *[hiːɫɪŋ, hiː.lɪt] dialect

*[hiːɫɪŋ, hiː.lɪt]

- If a prevocalic /l/ undergoes darkening opaquely because a vowel follows after a stem-suffix boundary, it can be inferred that the same overapplication will take place before a word boundary.
- Not only is this type of dialect unattested empirically, it is also argued to be impossible theoretically...

The Russian Doll Theorem

Let there be nested cyclic domains [γ ...[β ...[α ...] ...] ...]. If a phonological process ρ is opaque in β because its domain is α, then ρ is opaque in γ.

(Bermúdez-Otero, 2011:7)

Avoiding the *[hiː.ɫɪŋ, hiː.lɪt] dialect

- Hayes acknowledges that such grammars are impossible and avoids them by stipulating an innate ranking of OOC constraints in UG:

OO-IDENT(PHRASAL) » OO-IDENT(MORPHOLOGICAL)

- This ensures the impossible dialect is not generated.
- However, Hayes's solution has been criticised by Bermúdez-Otero (2011), who argues that this innate ranking stipulation is unnecessary in a cyclic framework.

Solution: A Stratal Approach

This innate stipulation can be avoided with Stratal OT, where the work of OOC is done by faithfulness between cycles.

Stratal OT

Phonological processes apply cyclically over a hierarchy of stem-level, word-level, and phrase-level domains; each domain is subject to stratum-specific OT grammar.

(Bermúdez-Otero, 1999, 2007, 2011; Kiparsky, 2000)

Replicating the Model with Stratal OT

- Re-run Boermsa & Hayes's tests over three cyclic domains.
- First time calculations of frequencies at each cyclic level has been attempted.
- Results can only be taken as proof of principle:
  - Calculated from well-formedness judgements. How reliable is this judgement data?
  - Many contexts missing, makes it insufficient for a stratal approach
  - Judged from broken vowel, not actual dark [ɫ]s

Outline

Guy (1991)

- Combined Labovian rules with lexical phonology
- Used a cyclic approach to T/D-deletion
- Noticed there was more deletion in monomorphemic *mist* than regular past-tense *missed*
- Uses a cyclic method to show that the monomorphemes have more chances to undergo TD-deletion.

[*mist*]

[*miss*]

- Note that Guy assumes a priori that each level has equal application of a rule.

Guy (1991)

- Combined Labovian rules with lexical phonology
- Used a cyclic approach to T/D-deletion
- Noticed there was more deletion in monomorphemic *mist* than regular past-tense *missed*
- Uses a cyclic method to show that the monomorphemes have more chances to undergo TD-deletion.

[[*mist*]]

[[*miss*]-*ed*]

- Note that Guy assumes a priori that each level has equal application of a rule.

The Life Cycle of Phonological Processes

- Considering this from a diachronic viewpoint, equal application at each level is not what we would expect given the life cycle of phonological processes (Bermúdez-Otero, 1999, 2011; Harris, 1989; McMahon, 2000).
- When a rule ceases to be gradient and becomes phonologised, it applies transparently at the phrase level and is able to see across word boundaries. Later, the rule may apply at the word level, and cannot see across word boundaries. Over time, its domain of application shrinks.

Explaining the Life Cycle

A phonological process begins by applying at the phrase level

Phrase

Word

Stem

Explaining the Life Cycle

It then works its way into the word level

Phrase

Word

Stem

Explaining the Life Cycle

Finally, it may reach the stem level...

Phrase

Word

Stem

Explaining the Life Cycle

as found in the data, linking in with the Russian Doll Theorem...

mai[ɫ]

mai[ɫ] it

mai[ɫ]-er

Explaining the Life Cycle

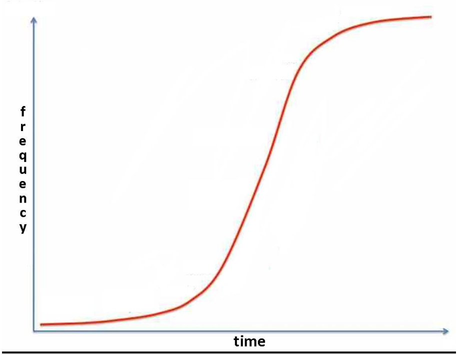
because we do not find patterns like this:

mai[ɫ]

mai[l] it

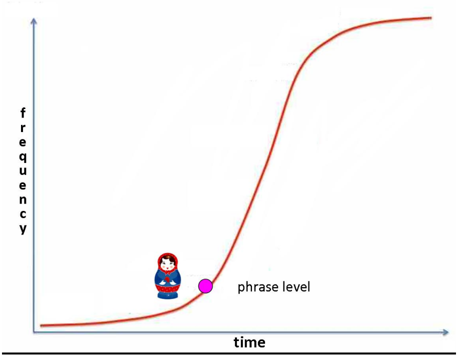
mai[ɫ]-er

Change in Progress



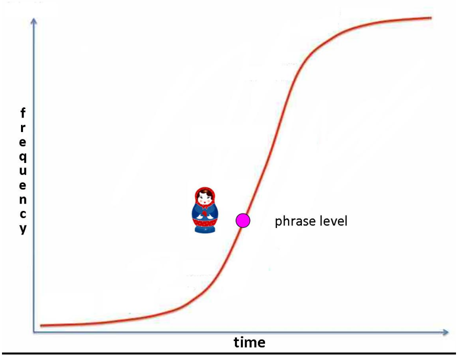
Considering that the application frequency of a change in progress increases over time PLUS the predictions made by the life cycle of phonological processes, is equal application at each level what we would predict?

Change in Progress



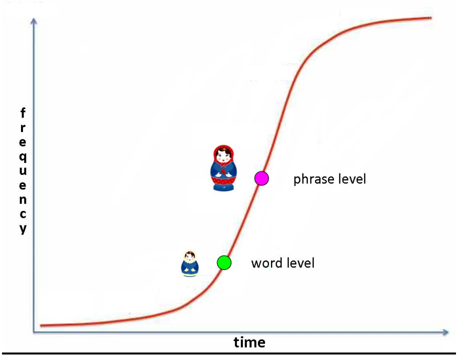
A rule enters the grammar at the phrase level...

Change in Progress



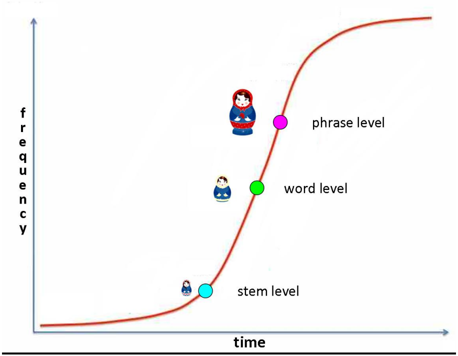
...increasing over time as a change in progress...

Change in Progress



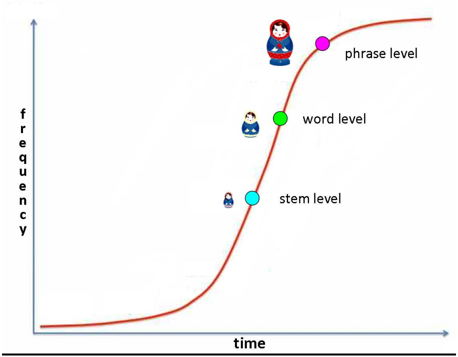
...then starts to apply at the word level...

Change in Progress



...and finally the stem level, giving us the following corollary...

The Variation Corollary of the Russian Doll Theorem



If a phonological process ρ shows a certain rate of application in a small embedded domain α, then ρ will apply at an equal or higher rate in a wider domain β.

Aims of the Present Study

- Develop a Stochastic Stratal OT approach to modelling Hayes's data.
- Calculate frequencies of darkening (or retention of light /l/) at each *individual* level of the cycle by using figures in Hayes's paper.
- Feed these frequencies into the GLA in three cycles to see if it can successfully model the data.
- If so, I argue this is an improvement as it does not involve the innate stipulation of the order of constraints **AND** it does not stipulate equal rates of application at each level, but rather seeks to learn the actual rate of application for each level.

Outline

Coda and Foot-Based Darkening

- First issue: to disentangle coda-based and foot-based darkening.
- The /l/ in *yellow, Hayley* is never in the coda, at any cyclic level.
- This demonstrates that /l/ is vulnerable to darkening not only in codas, but when non-initial in the foot.
- The figure given by Hayes for *Hayley* is the result of three cycles of foot-based darkening:

Foot-based darkening

$$R_{Hayley} = F_s \times F_w \times F_p = 0.76691$$

Retention of light /l/ in *Hayley* is equal to the product of foot-based retention at the stem, word and phrase levels.

Coda-Based Darkening

- Isolate separate cycles by dividing results by each other:
- *mail-er*: undergoes foot-level darkening at all three levels of the cycle and coda-level darkening at the stem level only.

- At the stem level, the /l/ is in the coda, as affixation has not yet occurred.
- At the word and phrase levels, the /l/ is resyllabified into the onset and escapes coda-based darkening.
- Dividing *mail-er* by *Hayley* isolates the level of coda-based retention at the stem-level.

Coda-Based Darkening

- Isolate separate cycles by dividing results by each other:
- *mail-er*: undergoes foot-level darkening at all three levels of the cycle and coda-level darkening at the stem level only.

Stem Level

meɪɫ

- At the stem level, the /l/ is in the coda, as affixation has not yet occurred.
- At the word and phrase levels, the /l/ is resyllabified into the onset and escapes coda-based darkening.
- Dividing *mail-er* by *Hayley* isolates the level of coda-based retention at the stem-level.

Coda-Based Darkening

- Isolate separate cycles by dividing results by each other:
- *mail-er*: undergoes foot-level darkening at all three levels of the cycle and coda-level darkening at the stem level only.

Stem Level	Word Level	Phrase Level
maɪl	maɪ.lə	maɪ.lə

- At the stem level, the /l/ is in the coda, as affixation has not yet occurred.
- At the word and phrase levels, the /l/ is resyllabified into the onset and escapes coda-based darkening.
- Dividing *mail-er* by *Hayley* isolates the level of coda-based retention at the stem-level.

Coda-Based Darkening

- Isolate separate cycles by dividing results by each other:
- *mail-er*: undergoes foot-level darkening at all three levels of the cycle and coda-level darkening at the stem level only.

Stem Level	Word Level	Phrase Level
maɪl	maɪ.lə	maɪ.lə

- At the stem level, the /l/ is in the coda, as affixation has not yet occurred.
- At the word and phrase levels, the /l/ is resyllabified into the onset and escapes coda-based darkening.
- Dividing *mail-er* by *Hayley* isolates the level of coda-based retention at the stem-level.

Coda-Based Darkening

Therefore, the only difference between *Hayley* and *mail-er* is that *mail-er* undergoes one extra process of coda-based darkening at the stem level.

C_s: Stem level coda darkening

$$\begin{aligned}
\frac{R_{mail-er}}{R_{Hayley}} &= \frac{F_s C_s F_w F_p}{F_s F_w F_p} = C_s \\
\frac{R_{mail-er}}{R_{Hayley}} &= 0.2173 \therefore \\
C_s &= 0.2173
\end{aligned}$$

Coda-Based Darkening

Dividing *mail it* by *mail-er* will isolate the retention of light /l/ at the word level.

C_w: Word level coda darkening

$$
\begin{aligned}
\frac{R_{mailit}}{R_{mail-er}} &= \frac{F_s C_s F_w C_w F_p}{F_s C_s F_w F_p} = C_w \\
\frac{R_{mailit}}{R_{mail-er}} &= 0.02947 \therefore \\
C_w &= 0.02947
\end{aligned}
$$

Coda-Based Darkening

C_p: Phrase level coda darkening

To isolate coda darkening at phrase level, divide *bell* by *mail it*.

$$\frac{R_{bell}}{R_{mailit}} = \frac{F_s F_w F_p C_s C_w C_p}{F_s F_w F_p C_s C_w} = C_p$$

$$\frac{R_{bell}}{R_{mailit}} = 0.03008 \therefore$$

$$C_p = 0.03008$$

Coda-Based Darkening

Final results for retention of light [l] in codas

$$C_p = 0.0298$$
$$C_w = 0.0298$$
$$C_s = 0.2173$$

Problems with Foot-Based Darkening

- Hayes doesn't look at the environments needed to calculate foot-based darkening in each cycle.
- Using *free-ly* does not work
- We could work out more accurate rates if Hayes had included:
  - Words like *mail-ee* where /l/ is not foot-initial at the stem level, but comes to be followed by a stressed vowel at the word and phrase levels
  - Phrases like *mail Ann*, where /l/ is not foot-initial at the stem and word levels, but initial in the phrase
- For foot-based darkening, I am forced to resort to the Guy-an model and take the cube root of the surface retention rate for *Hayley*:

Foot-based figure

$$F = \sqrt[3]{0.76691} = 0.9153$$

Frequencies

Table 2: Frequencies of light/dark allophones per million tokens

Word Type	**Stem Level Frequency**		**Word Level Frequency**		**Phrase Level Frequency**	
	light	*dark*	*light*	*dark*	*light*	*dark*
light	999,853	147	999,706	294	999,559	441
Hayley	915,341	84,659	837,849	162,151	766,917	233,082
mail-er	198,922	801,078	182,082	817,918	166,667	833,333
mail it	198,922	801,078	5,422	994,578	4,963	995,037
bell	198,922	801,078	5,422	994,578	148	999,852

Outline

Constraints

Hayes's Constraints

DARK [ɫ] IS POSTVOCALIC
PREVOCALIC /l/ IS LIGHT
PRETONIC /l/ IS DARK
/l/ IS DARK
OO-IDENT(PHRASAL)
OO-IDENT(MORPHOLOGICAL)

My Constraints

*CODA [l]
*[ɫ]
*[l] Σ-MEDIAL ONSETS
IDENT[l]
IDENT[ɫ]

- Hayes's constraints are inappropriate for a Stratal OT analysis
- In Stratal OT, the work of OOC is done by input-output correspondence between faithfulness in cycles.

Richness of the Base

- All possible inputs must be considered, for example, the output of *light*, which is [laɪt] in effect 100% of the time, must have a potential input /ɫaɪt/ which brings out surface [laɪt].
- The fact that tokens of /l/ that darken in the coda remain dark when resyllabified into the onset in a later cycle indicates that IDENT-[ɫ] must be high at the word and phrase levels. The GLA successfully learns this ranking on the basis of output frequencies for each stratum.

Outline

Stem Level Results

Constraint	Ranking Value
*[l] CODA	9,992
*[ɫ]	9,989
*[l] Σ-MED-ONSETS	9,982
IDENT[ɫ]	1,067
IDENT[l]	1,056

Input to level (Rich Base)	Light [l] in output of level			
	%		cumulative total per million	
	projected	*generated*	*projected*	*generated*
light *ɫight*	99.99	100	999,853	1,000,000
Hayley *Hayɫey*	91.53	99.33	915,341	993,250
mail-er *maiɫ-er*	19.89	14.07	198,922	140,700
mail it *maiɫ it*	19.89	14.07	198,922	140,700
mail *maiɫ*	19.89	14.07	198,922	140,700

Average error per candidate: **0.041**

Word Level Results

Constraint	Ranking Value
IDENT[ɫ]	10,008
*[l] CODA	9,962
IDENT[l]	9,952
*[l] Σ-MED-ONSETS	9,948
*[ɫ]	9,944

Input to level	Light [l] in output of level			
	%		cumulative total per million	
	projected	*generated*	*projected*	*generated*
light	99.97	100	999,705	1,000,000
Hayley	91.53	91.91	837,849	912,846
mail-er	91.53	91.91	182,082	129,310
mail it	2.73	0.01	5,422	14
mail	2.73	0.01	5,422	14

Average error per candidate: **0.015**

Phrase Level Results

Constraint	Ranking Value
IDENT[ɫ]	10,008
*[l] CODA	9,934
IDENT[l]	9,924
*[l] Σ-MED-ONSETS	9,918
*[ɫ]	9,916

Input to level	**Light [l] in output of level**				
	%		**cumulative total per million**		
	projected	*generated*	*projected*	*generated*	*% error*
light	99.99	100	999,559	1,000,000	0.04
Hayley	91.53	98.08	766,917	895,320	12.84
mail-er	91.53	98.08	166,667	126,828	3.98
mail it	91.53	98.08	4,963	14	0.49
mail	2.726	0.005	148	0	0.01

Average error per candidate: **0.068**

Conclusion

- Shown an SSOT grammar which models Boersma & Hayes's predicted frequencies of [l] and [ɫ] with a very small amount of error.
- The SSOT model preserves the advantages of of Stochastic OT with respect to the modelling of variation, AND of Stratal OT with respect to the analysis of morphosyntactic conditioning, without innately stipulated constraint rankings.
- The model demonstrates an easy method for working out the rate of application of a phonological process at each level. When applied to data from /l/-darkening, this method corroborates the predictions of the Variation Corollary of the Russian Doll Theorem.

Conclusion

BUT this study needs refining, as Hayes's data was not collected with such an analysis in mind.

- How reliable are conjectured frequencies projected from acceptability ratings?
- Based on vowel breaking, not on the realisation of the /l/ itself.
- No phonetic measurements
- Doesn't have all environments needed, e.g. *mail-ee, mail Ann*. This is the likely cause of relatively large error for *Hayley* in the current simulation.

Future Study

- Use this to model my own data in future study.
- This will allow me to improve the model, as I will have actual phonetic data and I can choose which environments I study.
- Also apply it to /ŋg/ simplification, e.g. in Manchester
- Linking and intrusive /r/

References

Anttila, Arto. 1997. Deriving variation from grammar: A study of Finnish genitives. In V*ariation, change and phonological theory,* ed. Frans Hinskens, Roeland van Hout, and Leo Wetzels. Amsterdam: John Benjamins. Rutgers Optimality Archive ROA-63, http://ruccs.rutgers.edu/roa.html.

Bermúdez-Otero, Ricardo. 1999. *Constraint interaction in language change: quantity in English and Germanic [Opacity and globality in phonological change].* Doctoral dissertation, University of Manchester and Universidad de Santiago de Compostela, Manchester and Santiago de Compostela. Available at http://www.bermudez-otero.com/PhD.pdf.

Bermúdez-Otero, Ricardo. 2011. Cyclicity, in Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice (eds). *The Blackwell companion to phonology.* Malden, MA: Wiley-Blackwell. Available from ling.auf.net/lingBuzz/000965

Boersma, Paul, and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm, *Linguistic Inquiry* 32, 45-86.

Coetzee, Andries. 2004. *What it means to be a loser: non-optimal candidates in optimality theory.* Doctoral dissertation, University of Massachusetts, Amherst.

References

Guy, G. 1991. Explanation in variable phonology: an exponential model of morphological constraints. *Language Variation and Change* 3(1): 1-22.

Harris, John. 1989. Towards a lexical analysis of sound change in progress. *Journal of Linguistics* 25:35-56

Hayes, Bruce P. 2000. Gradient well-formedness in Optimality Theory, in Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer (eds), *Optimality Theory: phonology, syntax, and acquisition*, 88-120. Oxford: Oxford University Press.

Kiparsky, Paul. 2000. Opacity and cyclicity, in Nancy A. Ritter (ed.), A review of Optimality Theory. Special issue, *The Linguistic Review* 17 (2-4): 351-67.

McMahon, April M. S. 2000. *Lexical phonology and the history of English.* Cambridge: Cambridge University Press.

Sproat, Richard, and Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21:291-311 .

Turton, Danielle. 2010. *The darkening of English /l/.* MA Research paper, The University of Manchester

Thanks!

MANCHESTER 1824

Arts & Humanities Research Council