

## *When and Why to Ignore Lexical Patterns in Turkish Obstruent Alternations*

Michael Becker, UMass Amherst (michael@linguist.umass.edu)  
Nihan Ketrez, Yale (nihan.ketrez@yale.edu)  
Andrew Nevins, Harvard (nevins@fas.harvard.edu)

### Highlights:

- In the lexicon, alternation rates can be predicted by the place of articulation of the stem-final stop, the size of the word, and the height and backness of the stem-final vowel
- Speakers are attuned to the place of articulation and size effects; they ignore preceding vowels – they distinguish between phonologically-motivated generalizations and accidental generalizations
- We propose that UG acts as a filter on the generalizations that humans can make; UG contains information about possible and impossible interactions between phonological elements
- We show that statistical models that do not have information about possible interactions reproduce accidental generalizations
- We propose that generalizations are expressed in terms of typologically-responsible OT constraints

## 1 The Turkish lexicon

Turkish regulates the voicing of stem-final stops, productively enforcing final devoicing and intervocalic voicing (Lees 1961, Inkelas & Orgun 1995, Vaux 2005, and others):

- |     |                 |                  |         |
|-----|-----------------|------------------|---------|
| (1) | rop ~ rob-u     | < French [rɔb]   | ‘dress’ |
|     | tüp ~ tüb-ü     | < French [tüb]   | ‘tube’  |
|     | gurup ~ gurub-u | < French [grup]  | ‘group’ |
|     | ešarp ~ ešarb-i | < French [ešarp] | ‘scarf’ |

But lexical exceptions abound:

- |     |                                  |             |        |
|-----|----------------------------------|-------------|--------|
| (2) | Failure of final devoicing:      | ad ~ ad-i   | ‘name’ |
|     | Failure of intervocalic voicing: | top ~ top-u | ‘ball’ |

Exceptions to final devoicing are fairly rare (~2% of the lexicon at best). We focused on the application of intervocalic voicing, which affects ~54% of the lexicon.

Application of intervocalic voicing is unpredictable for any given existing lexical item, but certain factors are known to affect it. We mined an electronic lexicon (TELL, Inkelas et al. 2000) for such factors:

### (3) Size

	<i>n</i>	% alternating
Monosyllabic, simplex coda (CVC)	137	12%
Monosyllabic, complex coda (CVCC)	164	26%
Poly-syllabic (CV.CVC and longer)	2701	59%

### (4) Place of articulation of the stem-final stop

	<i>n</i>	% alternating
p	294	84%
t	1255	17%
č	191	61%
k	1262	85%

### (5) Height of the stem’s final vowel:

	<i>n</i>	% alternating
–high (a, e, o, ö)	1690	42%
+high (ı, i, u, ü)	1312	72%

### (6) Backness of the stem’s final vowel:

	<i>n</i>	–back	+back	difference
p	294	90%	79%	–11%
t	1255	16%	18%	+1%
č	<b>191</b>	<b>44%</b>	<b>74%</b>	<b>+30%</b>
k	1262	84%	86%	+1%

## 2 Speakers' knowledge about voicing alternations

Inkelas & Orgun (1995) and Inkelas et al. (1997) mention **size** and **place**, but no vowel effects.

In our wug-tests, speakers replicate the **size** and **place** effect. Speakers do not replicate any vowel effects – neither **height** not **backness**.

### 2.1 Experiment 1: Size × place

We showed 10 four-year-old children (average age 4;0, age range 3;9–4;4) various toys, presenting each with a name, e.g. yuč. We asked them to name Kermit's yuč, hoping to get either "Kermit'in yuč-ü" or "Kermit'in yüj-ü". We also ran the same experiment with 12 adults.

(7) Bak bu bir yuč. Bu Kermit'in ... ne-si?  
Look this a yuč. This Kermit's ... what-poss?

24 stimuli	CVC	CVCVC
p	tup, dap, zip	yıyap, jisip, kunup
t	nut, hit, fet	gevit, muyut, niket
ç	niç, yuč, peç	bölüç, heveç, mariç
k	sük, pik, vek	banik, nönük, mesek

Results: Kids and adults only replicated the **size** and **place** effects from the lexicon (see also Nevins & Kamali 2005, Nakipoğlu & Üntak 2006)

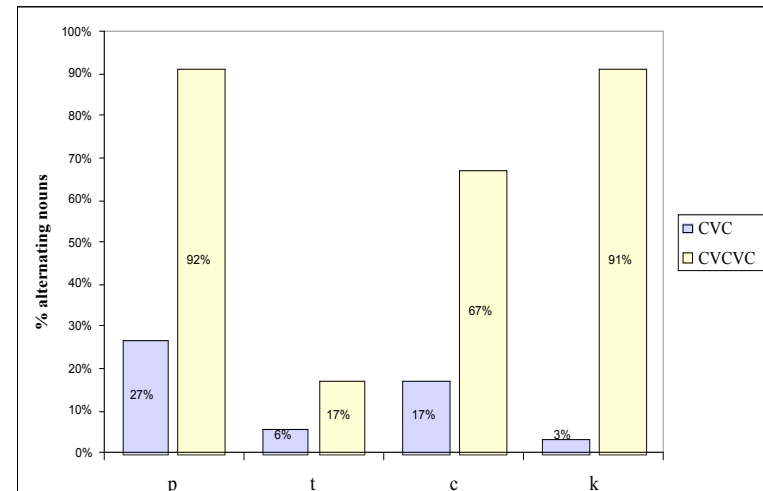
(8) Logistic regression, using the enter method:

	Lexicon	Kids (n=10)
size	p < .001	p < .001
place	p < .001	p = .004
high	p < .001	ns
back	p < .001	ns
round	ns	ns

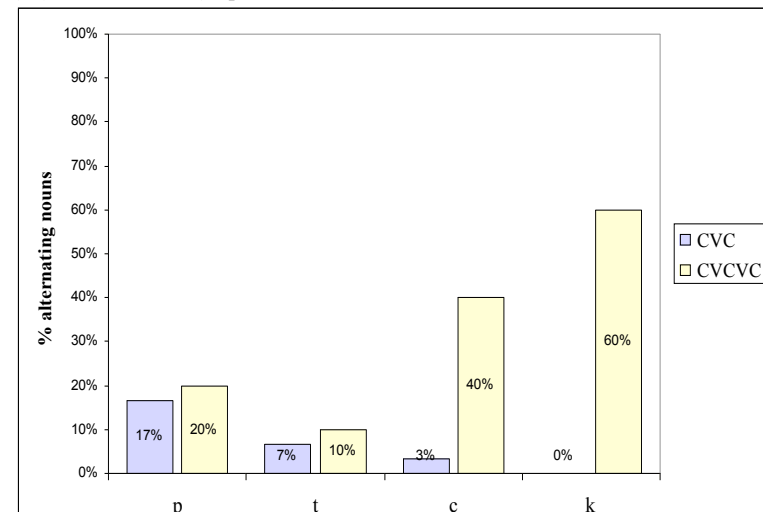
(9) ANOVA on the experimental results with **size** and **place**:

	Adults (n=12)	Kids (n=10)
size	p < .001	p < .001
place	p = .013	p = .021

(10) The lexicon (according to TELL)



(11) Kids' choices (Experiment 1)



(See Appendix A for detailed results)

## 2.2 Experiment 2: Size × place × vowel quality

We showed 24 adults nouns, e.g. kőč. They were shown a possessor, and asked to choose between two vocal renditions of the possessed noun, i.e. “kőč-ü” or “kőj-ü”.

- (12) Sizes: CVC, CVCC and CVCVC  
 Places: p, t, č, k  
 Vowels: a, i, e, i, o, u, ö, ü (high, back, round)

Total 72 stimuli

(See Appendix B for detailed results)

- (13) Logistic regressions on the lexicon and on the experiments results:

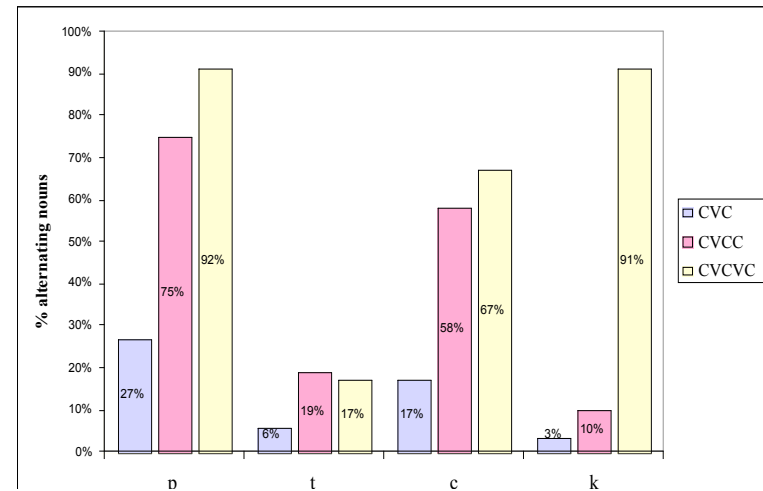
	Lexicon	Experiment 2
size	<b>p &lt; .001</b>	<b>p &lt; .001</b>
place	<b>p &lt; .001</b>	<b>p = .005</b>
high	<b>p &lt; .001</b>	ns
back	<b>p &lt; .001</b>	ns
round	ns	ns

**High and back** are significant in the lexicon, not in speakers' choices.

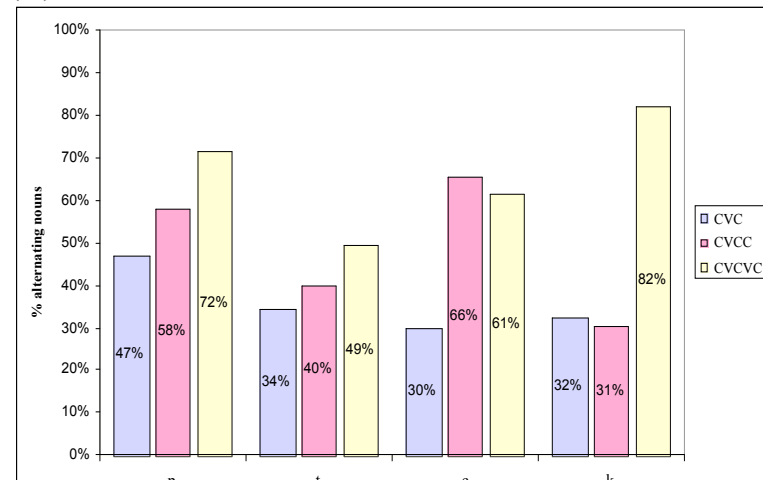
- (14) ANOVA on the experimental results:

	Adults (n=24)
size	p < .001
place	p = .006
high	ns
back	ns

(15) The lexicon (according to TELL)



(16) Adults' choices



### 2.3 Experiment in the works:

We will show four-year-old children various toys, presenting each with a name, e.g. niket. We will ask them to name Kermit's niket, hoping to get either "Kermit'in niket-i" or "Kermit'in niked-i". We will also run the same experiment with adults.

12 stimuli	
CVC[+high]t	varit, ponut, hekit, čölüt
CVC[-high]t	yiyat, sorat, fimet, nüšet
CVCVč	seyeč, moruč, kasič, hinač

In the lexicon, the height effect is strongest for poly-syllabic t-final nouns. If the effect is out there, this will be its best chance.

### 3 UG-less modeling of speakers' knowledge

Speakers' knowledge is tightly correlated with the lexicon; we assume that they project their knowledge from the lexicon. We used the Minimal Generalization Learner (henceforth MGL, Albright & Hayes 2002), which can learn generalizations from a corpus and project them onto novel words.

We gave the MGL the stop-final words of TELL and their possessive forms, annotated for mono-/poly-syllabicity of the stem, and a table of segments and their features. The MGL starts with creating a rule for each word of the lexicon, and collapses phonologically similar rules to form generalizations. Ex.:

- (17) šarap<sub>2</sub> ~ šarabi<sub>2</sub> rule: p → bi / š a r a \_\_<sub>2</sub>  
 kebab<sub>2</sub> ~ kebab<sub>i</sub> rule: p → bi / k e b a \_\_<sub>2</sub>
- 
- Generalization: p → bi / X a \_\_<sub>2</sub>

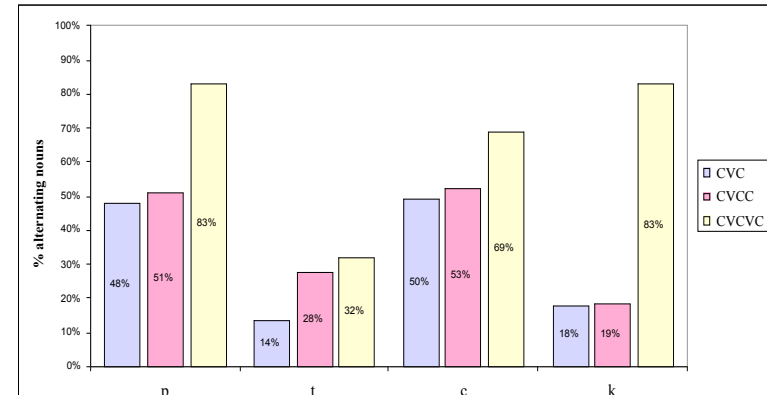
- (18) To make the MGL results comparable to the experimental results, we simulated 24 "learners", mapping the MGL results (with some random noise) onto zeros and ones.

- (19) Logistic regressions on the lexicon, on the experimental results, and on the MGL results:

	Lexicon	MGL	Experiment 2
size	p < .001	p < .001	p < .001
place	p < .001	p < .001	p = .005
high	p < .001	p = .007	ns
back	p < .001	p = .003	ns
round	ns	ns	ns

The MGL replicates the **high** and **back** effects of the lexicon, unlike the humans.

- (20) MGL results with the 72 stimuli from experiment 2:



The MGL modeled the lexicon *too well*, generating data that fit the lexicon more closely than humans. Most importantly, it failed to ignore the vowel effects that are in the lexicon.

We conclude that the MGL reproduced both the phonologically-motivated generalizations (**size** and **place** effects) and the accidental generalizations (**high** and **back** effects) that were found in the lexicon.

What's missing from the MGL is a theory of possible and impossible interactions between phonological elements.

## 4 The ingredients of a UG-based analysis

We assume that UG acts as a filter on learning the lexicon. UG constrains the learning process, making speakers notice phonologically-motivated generalizations and ignore others.

When speakers derive novel forms, they **do not access their lexicon**. They only use their grammar, which has the phonologically-motivated aspects of the lexicon built into it.

### 4.1 Lexical statistics are kept following Inconsistency Detection

We propose a learning model in which speakers detect inconsistency in the grammar (Pater 2006) and then start keeping track of the behavior of individual items:

(21)	/jɛp+V/	OO-IDENT(voice)	*VT]V
	→ jɛb-i	1	
	jɛp-i	L <sub>0</sub>	W <sub>1</sub>

(22)	/top+V/	OO-IDENT(voice)	*VT]V
	→ top-u		1
	tob-u	W <sub>1</sub>	L <sub>0</sub>

(23) IDENT(voice)<sub>top</sub> » \*VT]V » IDENT(voice)<sub>jɛp</sub>

As more words are learned, each instance of OO-IDENT(voice) will accumulate “weight”, and this “weight” is projected onto novel words:

(24) IDENT(voice)<sub>top,ip,sop,bap,ɕap,hep,kip...</sub> » \*VT]V » IDENT(voice)<sub>jɛp,tɪp,kap...</sub>

Thus, the ratio of alternating and non-alternating nouns is built into the constraint ranking. A novel word like *zɪp* will be attracted by the heavier top-ranking IDENT, so *zɪp-ɪ* is more likely than *zɪb-ɪ*.

### 4.2 Generalizations in terms of constraints

Inconsistency Detection is done for each constraint in CON separately:

- (25) The place effect follows from the existence of place-specific constraints: OO-IDENT(voice)&\*COR, etc.
- (26) Initial syllables are protected by positional faithfulness, allowing generalizations over mono-syllabic bases to be kept separately from generalizations over poly-syllabic bases: OO-IDENT(voice)&\*COR<sub>01</sub>, OO-IDENT(voice)&\*LAB<sub>01</sub>, etc.

### 4.3 Lack of effect from lack of constraints

No language is known to change obstruent voicing based on the quality of a neighboring vowel.

Therefore, OT has no constraints relating any voice specification and any neighboring vowel quality, such as:

(27) \*[+back][−voice]

Any relationship between vowel quality and obstruent voicing is necessarily accidental; speakers cannot be attuned to it.

## 5 Conclusion

- Turkish speakers extract information about voicing alternations from their lexicon, relativized to size and place; they ignore the effect of the stem-final vowel.
- When speakers derive novel words, they do not access the lexicon. They access their grammar, which is constrained by UG.
- UG contains information about possible and impossible interactions between phonological elements.
- We propose an OT model that can learn the phonologically-motivated aspects of the lexicon.

## 6 Acknowledgments

For their valuable comments and suggestions on several versions of this work, we are greatly indebted to Adam Albright, Nicoleta Bateman, Kathryn Flack, Lyn Frazier, Karen Jesney, Shigeto Kawahara, Michael Key, John Kingston, Joe Pater, Kathryn Pruitt, John McCarthy, Anne-Michelle Tessier, Bert Vaux, and Matt Wolf. We would also like to thank the audience at the 2004 HUMDRUM, especially Luigi Burzio, Jane Grimshaw, Alan Prince and Ellen Woolford, and the audience at the Yale 2006 Turkish Linguistics Workshop, especially İlhan Çağrı, Atakan İnce, and Jaklin Kornfilt. Thanks to the many speakers of Turkish who helped with various aspects of this project, including Mine Doğucu, Anıl and Alper Duman, Beste Kamali, Leyla Zidani-Eroğlu, several students at Boğaziçi university, and other kind people who filled our questionnaire. Thanks also to the children and parents at the Yeni Çilek and Ayışığı kindergartens in Levent, İstanbul. Any remaining errors, if any, would be ours.

## References

- Albright & Hayes (2002). Modeling English Past Tense Intuitions with Minimal Generalization. In Maxwell, Michael (ed) *Proceedings of the Sixth Meeting of the ACL Special Interest Group in Computational Phonology*. Philadelphia, July 2002. ACL.
- Inkelas, Sharon, Aylin Kuntay, John Lowe, Orhan Orgun & Ronald Sprouse (2000). Turkish electronic living lexicon (TELL). Website, <http://socrates.berkeley.edu:7037/>.
- Inkelas, Sharon & Cemil Orhan Orgun (1995). Level ordering and economy in the lexical phonology of turkish. *Language* 71. 763–793.
- Inkelas, Sharon, Cemil Orhan Orgun & Cheryl Zoll (1997). The implications of lexical exceptions for the nature of the grammar. In Iggy Roca (ed.) *Derivations and Constraints in Phonology*, Oxford: Clarendon. 393–418.
- Nakipoğlu, Mine and Aslı Üntak (2006) What does the acquisition of stems that undergo phonological alternation reveal about rule application? Talk given at ICTL 2006, Uppsala
- Nevins, Andrew & Beste Yolcu-Kamali (2005). Constructing underlying representations based on informed guesses: Turkish evidence. Talk given at the conference of the Central Eurasian Studies Society, Boston University.
- Pater, Joe (2004). Exceptions and optimality theory: Typology and learnability. Talk given in the Conference on Redefining Elicitation: Novel Data in Phonological Theory. New York University.
- Vaux, Bert (2005). Formal and empirical arguments for morpheme structure constraints. Talk given at LSA Annual Meeting, San Francisco.

## Appendix A: materials and results for experiment 1

	CVC			CVCVC		
		kids	adults		kids	adults
p	tup	0%	17%	yiyap	20%	42%
	dap	20%	42%	şisip	20%	50%
	zip	30%	42%	kunup	20%	67%
t	nut	10%	17%	gevit	20%	25%
	hit	0%	0%	muyut	10%	42%
	fet	10%	0%	niket	0%	25%
ç	niç	0%	25%	bölüç	50%	50%
	yüç	10%	42%	heveç	40%	58%
	peç	0%	17%	maiç	30%	50%
k	sük	0%	0%	banık	40%	100%
	pik	0%	0%	nönük	80%	92%
	vek	0%	8%	mesek	60%	92%

**Appendix B: materials and results for experiment 2**

			CVC		CVCC		CVCVC	
			-high	+high	-high	+high	-high	+high
p/č	-round	-back	gep 63%	yič 33%	telp 67%	ginč 75%	heveč 46%	jisip 54%
		+back	dap 42%	nič 29%	panč 54%	dirp 50%	yyiap 71%	maič 63%
	+round	-back	köč 33%	züp 33%	yönč 58%	kürp 63%		bölüč 63% türüč 75%
		+back	poč 25%	tup 50%	solp 54%	munč 75%		kunup 88% goyup 75%
t	-round	-back	pet 46%	hit 25%	zelt 38%	čint 29%	niket 33%	gevit 46%
		+back	fat 29%	mit 21%	hant 54%	širt 46%	yaat 54%	pisit 54%
	+round	-back	söt 42%	jüt 46%	gönt 42%	nürt 54%		sölüt 50% bünüt 66%
		+back	yot 29%	nut 38%	jolt 21%	bunt 38%		čorut 33% muyut 59%
k	-round	-back	vek 33%	zik 25%	helk 21%	tink 42%	mesek 79%	perik 83%
		+back	jak 21%	pik 33%	vank 38%	nirk 29%	tatak 79%	banik 92%
	+round	-back	hök 33%	šük 29%	sönk 50%	pürk 13%		nönük 92% düyük 79%
		+back	mok 33%	nuk 50%	bolk 21%	dunk 33%		zorum 75% yuluk 79%