Another look at English phonotactics

Péter Szigetvári

Eötvös Loránd University, Budapest
szigetvari@elte.hu

24mfm, 2016-05-28

miscategorized vowels of British English

- FLEECE and GOOSE are diphthongs (Sweet 1900:233f, Jones 1960:65f, Gimson 1989:121): ɪj, ~~ʉ~~w

miscategorized vowels of British English

- FLEECE and GOOSE are diphthongs (Sweet 1900:233f, Jones 1960:65f, Gimson 1989:121): ɪj, ʉw
- NEAR, SQUARE, and CURE are long monophthongs (Jones 1960:117, Upton 1995, Lindsey 2012): ɪː, ɛː, oː/əː/ʉː

miscategorized vowels of British English

- FLEECE and GOOSE are diphthongs (Sweet 1900:233f, Jones 1960:65f, Gimson 1989:121): ɪj, ʊw
- NEAR, SQUARE, and CURE are long monophthongs (Jones 1960:117, Upton 1995, Lindsey 2012): ɪː, ɛː, oː/əː/ʊː

lexical sets	__C	__#	__V
KIT, DRESS, TRAP, LOT, FOOT, STRUT	✓	✗	✗
NEAR, SQUARE, START, NURSE, FORCE, COMMA (CURE)	✓	✓	✗
FLEECE, FACE, PRICE, CHOICE, MOUTH, GOAT, GOOSE	✓	✓	✓

Cruttenden (2014:99f) for the diphthong analysis

1. vowels and glides do not freely combine

SHORT	LONG	DIPHTHONG	
ɪ kit	ɪː near	ɪj fleece	
ɛ dress	ɛː square	ɛj face	
a trap	aː start	aj price	aw mouth
o lot	oː force	oj choice	
ʉ foot	(ʉː cure)		ʉw goose
ə strut	əː nurse		əw goat

Cruttenden (2014:99f) for the diphthong analysis

1. vowels and glides do not freely combine

SHORT	LONG	DIPHTHONG	
ɪ kit	ɪː near	ɪj fleece	
ɛ dress	ɛː square	ɛj face	
a trap	aː start	aj price	aw mouth
o lot	oː force	oj choice	
ʉ foot	(ʉː cure)↕		ʉw goose
ə strut	əː nurse		əw goat

2. postvocalic glides are weakly articulated. . .

Cruttenden (2014:99f) for the diphthong analysis

1. vowels and glides do not freely combine

SHORT	LONG	DIPHTHONG	
ɪ kit	ɪː near	ɪj fleece	
ɛ dress	ɛː square	ɛj face	
a trap	aː start	aj price	aw mouth
o lot	oː force	oj choice	
ʉ foot	(ʉː cure)↕		ʉw goose
ə strut	əː nurse		əw goat

2. postvocalic glides are weakly articulated. . .
   just as almost any postvocalic C, eg *kick*, *lull*, compared to its prevocalic counterpart

Cruttenden (2014:99f) for the diphthong analysis

1. vowels and glides do not freely combine

SHORT	LONG	DIPHTHONG	
ɪ kit	ɪː near	ɪj fleece	
ɛ dress	ɛː square	ɛj face	
a trap	aː start	aj price	aw mouth
o lot	oː force	oj choice	
ʉ foot	(ʉː cure)↕		ʉw goose
ə strut	əː nurse		əw goat

2. postvocalic glides are weakly articulated. . .
   just as almost any postvocalic C, eg *kick, lull,* compared to its prevocalic counterpart
3. postvocalic glides have no devoiced fricative soundings, like in *tune* [tɕʉwn]/[tʃʉwn], *queen* [kʍɪjn]. . .

Cruttenden (2014:99f) for the diphthong analysis

1. vowels and glides do not freely combine

SHORT	LONG	DIPHTHONG	
ɪ kit	ɪː near	ɪj fleece	
ɛ dress	ɛː square	ɛj face	
a trap	aː start	aj price	aw mouth
o lot	oː force	oj choice	
ʉ foot	(ʉː cure)↕		ʉw goose
ə strut	əː nurse		əw goat

2. postvocalic glides are weakly articulated. . .
   just as almost any postvocalic C, eg *kick, lull,* compared to its prevocalic counterpart
3. postvocalic glides have no devoiced fricative soundings, like in *tune* [tɕʉwn]/[tʃʉwn], *queen* [kʍɪjn]. . .
   this is quite obvious: they are not after a fortis plosive

on Cruttenden's first argument

- similar constraints characterize consonant clusters

pj	*	kj
*	tw	kw

$\approx$

ɛj	*	aj
*	əw	aw

on Cruttenden's first argument

- similar constraints characterize consonant clusters

pj	*	kj
*	tw	kw

≈

ɛj	*	aj
*	əw	aw

- current developments in BrE are filling the vowel+glide space

SHORT	LONG	DIPHTHONG	
ɪ kit	ɪː near	ɪj fleece	
ɛ dress	ɛː square	ɛj face	
a trap	aː start	aj price	aw mouth
o lot	oː force	oj choice	
ʉ foot	(ʉː cure)		ʉw goose
ə strut	əː nurse		əw goat

on Cruttenden's first argument

- similar constraints characterize consonant clusters

pj	*	kj
*	tw	kw

≈

ɛj	*	aj
*	əw	aw

- current developments in BrE are filling the vowel+glide space

SHORT	LONG	DIPHTHONG	
ɪ kit	ɪː near	ɪj fleece	
ɛ dress	ɛː square	ɛj face	
a trap	aː start	aj price	aw mouth
o lot	oː force	oj choice	
ʉ foot	(ʉː cure)		ʉw goose
ə strut	əː nurse		əw goat

l-vocalization

on Cruttenden's first argument

- similar constraints characterize consonant clusters

pj	*	kj
*	tw	kw

$\approx$

ɛj	*	aj
*	əw	aw

- current developments in BrE are filling the vowel+glide space

SHORT	LONG	DIPHTHONG	
ɪ kit	ɪː near	ɪj fleece	ɪw bill
ɛ dress	ɛː square	ɛj face	ɛw bell
a trap	aː start	aj price	aw mouth
o lot	oː force	oj choice	ow ball
ʉ foot	(ʉː cure)		ʉw goose
ə strut	əː nurse		əw goat

l-vocalization

on Cruttenden's first argument

- similar constraints characterize consonant clusters

pj	*	kj
*	tw	kw

≈

ɛj	*	aj
*	əw	aw

- current developments in BrE are filling the vowel+glide space

SHORT	LONG	DIPHTHONG	
ɪ kit	ɪː near	ɪj fleece	ɪw bill
ɛ dress	ɛː square	ɛj face	ɛw bell
a trap	aː start	aj price	aw mouth
o lot	oː force	oj choice	ow ball
ʉ foot	(ʉː cure) ↕		ʉw goose
ə strut	əː nurse		əw goat

glide fronting

on Cruttenden's first argument

- similar constraints characterize consonant clusters

pj	*	kj
*	tw	kw

≈

ɛj	*	aj
*	əw	aw

- current developments in BrE are filling the vowel+glide space

SHORT	LONG	DIPHTHONG	
ɪ kit	ɪː near	ɪj fleece	ɪw bill
ɛ dress	ɛː square	ɛj face	ɛw bell
a trap	aː start	aj price	aw mouth
o lot	oː force	oj choice	ow ball
ʉ foot	(ʉː cure)	ʉj goose	
ə strut	əː nurse	əj goat	

glide fronting

on Cruttenden's first argument

- similar constraints characterize consonant clusters

pj	*	kj
*	tw	kw

≈

ɛj	*	aj
*	əw	aw

- current developments in BrE are filling the vowel+glide space

SHORT	LONG	DIPHTHONG	
ɪ kit	ɪː near	ɪj fleece	ɪw bill
ɛ dress	ɛː square	ɛj face	ɛw bell
a trap	aː start	aj price	aw mouth
o lot	oː force	oj choice	ow ball
ʉ foot	(ʉː cure)	ʉj goose	(uw bull)
ə strut	əː nurse	əj goat	əw dull

l-vocalization

on Cruttenden's first argument

- similar constraints characterize consonant clusters

pj	*	kj
*	tw	kw

≈

ɛj	*	aj
*	əw	aw

- current developments in BrE are filling the vowel+glide space

SHORT	LONG	DIPHTHONG	
ɪ kit	ɪː near	ɪj fleece	ɪw bill
ɛ dress	ɛː square	ɛj face	ɛw bell
a trap	aː start	aj price	aw mouth
o lot	oː force	oj choice	ow ball
ʉ foot	(ʉː cure)	ʉj goose	(uw bull)
ə strut	əː nurse	əj goat	əw dull

arguments against the diphthong analysis

1. vowels and glides freely combine in current BrE (nb this is not a sine qua non for treating their combinations as separate segments)

arguments against the diphthong analysis

1. vowels and glides freely combine in current BrE (nb this is not a sine qua non for treating their combinations as separate segments)
2. j-final "diphthongs" cannot be followed by j; w-final "diphthongs" cannot be followed by w ($\Leftarrow$ no geminates)

arguments against the diphthong analysis

1. vowels and glides freely combine in current BrE (nb this is not a sine qua non for treating their combinations as separate segments)
2. j-final "diphthongs" cannot be followed by j; w-final "diphthongs" cannot be followed by w ($\Leftarrow$ no geminates)
3. VG sequences seem to be missing (eg *narwhal* náːwəl, *Iowa* ájəwə, but *Dewi* dɛ́wɪj is unique; cf Polgárdi 2015)

arguments against the diphthong analysis

1. vowels and glides freely combine in current BrE (nb this is not a sine qua non for treating their combinations as separate segments)
2. j-final "diphthongs" cannot be followed by j; w-final "diphthongs" cannot be followed by w ($\Leftarrow$ no geminates)
3. VG sequences seem to be missing (eg *narwhal* náːwəl, *Iowa* ájəwə, but *Dewi* dɛ́wɪj is unique; cf Polgárdi 2015)
4. schwa-epenthesis in *fire, file* (also cf Kristó 2015:192f)

arguments against the diphthong analysis

1. vowels and glides freely combine in current BrE (nb this is not a sine qua non for treating their combinations as separate segments)
2. j-final "diphthongs" cannot be followed by j; w-final "diphthongs" cannot be followed by w ($\Leftarrow$ no geminates)
3. VG sequences seem to be missing (eg *narwhal* náːwəl, *Iowa* ájəwə, but *Dewi* dɛ́wɪj is unique; cf Polgárdi 2015)
4. schwa-epenthesis in *fire, file* (also cf Kristó 2015:192f)
5. New Zealand English acrolect flapping only after short vowels: *matter* márə vs *mitre* májtə, *outer* áwtə (cf Bye & de Lacy 2008:195ff; an alternative account in Balogné Bérces 2015:35f)

arguments against the diphthong analysis

1. vowels and glides freely combine in current BrE (nb this is not a sine qua non for treating their combinations as separate segments)
2. j-final "diphthongs" cannot be followed by j; w-final "diphthongs" cannot be followed by w (⇐ no geminates)
3. VG sequences seem to be missing (eg *narwhal* náːwəl, *Iowa* ájəwə, but *Dewi* dɛ́wɪj is unique; cf Polgárdi 2015)
4. schwa-epenthesis in *fire, file* (also cf Kristó 2015:192f)
5. New Zealand English acrolect flapping only after short vowels: *matter* márə vs *mitre* májtə, *outer* áwtə (cf Bye & de Lacy 2008:195ff; an alternative account in Balogné Bérces 2015:35f)
6. unstressed syllables may contain STRUT, KIT, and FOOT, ie ə ɪ ʉ (*lemon* lɛ́mən, *image* ímɪʤ, *capitulate* kəpíʧʉlɛjt), **as well as** əw ɪj ʉw (*motto* mótəw, *happy* hápɪj, *value* váljʉw)

are jʉ and jo diphthongs?

CjV is constrained, CwV is less so

Cj—	Cw—	Cl—	Cr—
mute mjʉwt	swoon swʉwn	flute flʉwt	fruit frʉwt
pure pjoː	quart kwoːt	flaw floː	frog frog
pure kjəː	quirk kwəːk	blurb bləːb	Brazil brəzɪl
piano pjaːnəw	qualm kwaːm	class klaːs	grass graːs
	dwell dwɛl	bless blɛs	dress drɛs
	twin twɪn	bliss blɪs	brief brɪjf

are jʉ and jo diphthongs?

CjV is constrained, CwV is less so

Cj—	Cw—	Cl—	Cr—
mute mjʉwt	swoon swʉwn	flute flʉwt	fruit frʉwt
pure pjoː	quart kwoːt	flaw floː	frog frog
pure kjəː	quirk kwəːk	blurb bləːb	Brazil brəzɪl
piano pjaːnəw	qualm kwaːm	class klaːs	grass graːs
	dwell dwɛl	bless blɛs	dress drɛs
	twin twɪn	bliss blɪs	brief brɪjf

but *CCj, *CCw (except sCj, sCw)

that is, the size of the prevocalic string is limited: *blue* *bljʉw

are jʉ and jo diphthongs?

CjV is constrained, CwV is less so

Cj—	Cw—	Cl—	Cr—
mute mjʉwt	swoon swʉwn	flute flʉwt	fruit frʉwt
pure pjoː	quart kwoːt	flaw floː	frog frog
pure kjəː	quirk kwəːk	blurb bləːb	Brazil brəzɪl
piano pjaːnəw	qualm kwaːm	class klaːs	grass graːs
	dwell dwɛl	bless blɛs	dress drɛs
	twin twɪn	bliss blɪs	brief brɪjf

but *CCj, *CCw (except sCj, sCw)

that is, the size of the prevocalic string is limited: *blue* *bljʉw

cf F *Blois* blwa vs *Blier* blije, *blje

⇒ French wa (in *Blois*) is a diphthong, je is not one

are jʉ and jo diphthongs?

CjV is constrained, CwV is less so

Cj—	Cw—	Cl—	Cr—
mute mjʉwt	swoon swʉwn	flute flʉwt	fruit frʉwt
pure pjoː	quart kwoːt	flaw floː	frog frog
pure kjəː	quirk kwəːk	blurb bləːb	Brazil brəzɪl
piano pjaːnəw	qualm kwaːm	class klaːs	grass graːs
	dwell dwɛl	bless blɛs	dress drɛs
	twin twɪn	bliss blɪs	brief brɪjf

but *CCj, *CCw (except sCj, sCw)

that is, the size of the prevocalic string is limited: *blue* *bljʉw
cf F *Blois* blwa vs *Blier* blije, *blje
⇒ French wa (in *Blois*) is a diphthong, je is not one
*CCj suggests that English jʉ and jo are not diphthongs

word-final sequences with j

token | type frequency

81301163363	j	11001	j	860443980	jnt	104	jf	21201231	jfs	16	jðz
9774087753	jt	2940	jz	645876190	jʧ	83	jps	19956706	jgz	14	jbd
9754415803	jz	1601	jt	587405840	jnʤ	72	jnt	15583200	jʒ	13	jðd
8082333371	jn	1131	jn	491708637	jvz	67	jmd	13929132	jsdz	13	jʧt
7772555374	jl	1037	jd	318491200	jb	66	jvd	13829840	jð	12	jnʤd
7625550060	jd	831	jts	286405623	jzd	56	jʤ	10728097	jʃ	11	jnθ
4510720625	jk	552	jl	258217147	jps	43	jθ	7500600	jnθ	9	jfd
4136156910	jm	372	jnz	235684378	jvd	41	jkt	5920000	jlt	8	jnθs
4047955102	js	327	jzd	219053824	jmd	40	jnts	5053800	jtθ	8	jgz
3094689905	jts	296	js	210848931	jθ	37	jpt	3101380	jðd	7	jʒ
2532714392	jv	293	jk	208198860	jg	35	jʧ	2320170	jðz	6	jgd
2171046912	jlz	259	jdz	133120984	jnts	28	jnʤ	1919983	jgd	4	jθs
1751693571	jmz	234	jlz	121679741	jʤd	28	jg	1460000	jʃd	2	jtθ
1524390546	jf	229	jm	115963550	jldz	28	jfs	579900	jfd	2	jŋk
1414074401	jnz	224	jld	96491410	jndz	26	jb	204568	jθs	2	jlt
1337431167	jʤ	209	jnd	60913433	jpt	25	jsdz	179000	jŋ	1	jtθs
1332412698	jnd	168	jv	48971708	jnʤd	25	jʃ	113700	jŋk	1	jsg
1312067143	jp	150	jks	44000000	jlsd	25	jldz	78400	jpsd	1	jʃd
1160113292	jks	143	jp	40608621	jkt	25	jʤd	65544	jnθs	1	jpsd
1153741129	jsd	139	jsd	38419957	jbd	25	jð	25700	jnsd	1	jnsd
1054783687	jdz	132	jvz	35592053	jʧt	22	jbz	6440	jtθs	1	jŋ
1033581977	jld	118	jmz	23347265	jbz	17	jndz	2130	jsg	1	jlsd

word-final sequences with w

token | type frequency

token		type		token		type		token		type	
37546622155	w	2083	w	364516871	wmz	49	wv	9060220	wʒ	13	wð
7262980005	wz	835	wz	283111072	wʧ	47	wndz	6993914	wʧt	11	wnsd
5193787908	wn	512	wn	259528047	wð	41	wb	4808428	wfs	11	wns
4520511665	wt	398	wt	239373500	wnʤ	34	wʧ	4731850	wlts	11	wlts
3057276743	ws	362	wd	178216173	wsdz	31	wldz	4604590	wʃ	10	wʒ
2685442460	wsd	326	wl	166859920	wnts	28	wpt	1562701	wðd	10	wfd
2591978644	wd	263	ws	157487756	wps	28	wnt	1432120	wgz	7	wbd
2020084588	wm	234	wts	149012720	wvd	27	wmd	1316910	wfd	6	wθs
1911589397	wl	208	wm	148982286	wndz	26	wvz	669400	wnʤd	6	wgz
1281254094	wld	204	wnz	107444810	wʤ	26	wsdz	622880	wθd	4	wθd
1267550768	wnd	178	wlz	102065510	wnsd	24	wðz	525120	wbd	3	wnʤ
1091209500	wnt	162	wnd	97785187	wg	23	wlt	279800	wθs	3	wʤd
1053147496	wp	141	wld	96855677	wf	23	wbz	55000	wgd	2	wnʤd
1030947489	wθ	118	wp	76432840	wvz	22	wkt	46550	wʤd	2	wksd
753382138	wv	118	wmz	67695647	wks	19	wfs	43800	wʒd	1	wʒd
748518766	wlz	105	wk	52320990	wlt	17	wnts	37600	wlsd	1	wsg
735224614	wzd	103	wdz	46586815	wðz	15	wʤ	16940	wksd	1	wlθ
608453128	wb	101	wsd	44361033	wldz	15	wʧt	10400	wðdz	1	wlsd
579805589	wk	87	wzd	27624500	wns	14	wvd	8680	wsg	1	wlp
500259156	wdz	76	wps	25551464	wkt	14	wg	5680	wlθ	1	wln
475016181	wnz	65	wθ	18131090	wpt	14	wðd	4030	wlp	1	wgd
427932259	wts	64	wks	14437555	wmd	13	wʃ	585	wln	1	wðdz
385959948	wbz	52	wf								

word-final sequences with l

token	type	frequency					
32198251120	l	4358	l	12610010	ldz	7	lpt
2897482779	lz	1155	lz	6396010	ln	6	ln
1051876592	ld	504	ld	5170590	lb	6	lkt
797839778	lf	98	lt	4501200	lʤ	6	lfs
678247960	lp	48	lts	2636602	lks	5	lsd
520007759	lt	47	lf	1299330	lbz	5	lʤ
490543315	lts	26	lvz	1210000	lfθ	5	lfd
471063500	lm	22	lk	644200	lnz	4	lnz
448619400	lθ	21	lv	436908	lfd	4	lʤd
149304280	ls	17	lm	420780	lfs	4	lʧt
133691824	lvz	14	lvd	380560	lkt	4	lb
122061670	lʃ	14	lks	368762	lsd	3	lmd
118664280	lvd	13	lp	299030	lʤd	3	lbz
86486230	lv	12	ls	229540	lʧt	2	lθs
75608300	lk	11	ldz	112000	lg	1	lʃd
57906368	lps	10	lps	69200	lpts	1	lpts
39010503	lmz	10	lmz	4890	lfθs	1	lkts
33861670	lpt	9	lʧ	4240	lʃd	1	lg
26052000	lmd	7	lθ	3624	lθs	1	lfθs
15006750	lʧ	7	lʃ	2320	lkts	1	lfθ

word-final sequences with nasals

token	type	frequency											
49605623759	n	7344	n	314522930	nʤ	52	ŋkt	583000	ŋg	3	ŋkts		
26595588726	ŋ	6789	ŋ	281765410	nθs	49	nsd	580600	nsg	3	nks		
13616787610	m	2227	nz	270013470	nʧ	48	mpt	482000	ntθs	2	ŋsd		
11236898141	nt	1904	m	159184900	ŋθ	39	nʤ	346600	nʃ	2	mt		
9199231534	nd	1276	nt	86676746	mps	29	nθ	304640	nks	2	mpf		
5652524672	nz	712	nd	78533642	mpt	25	ŋd	280053	mfd	2	mfd		
4408205824	ns	590	nts	44570163	md	25	nʧt	157200	nzd	2	mb		
2021860188	nts	536	ns	32538790	mpts	20	nʤd	136300	mt	1	nʒ		
1766483158	ŋk	414	ŋz	27868061	mf	12	nθs	91100	mlz	1	ntθs		
1433136194	mz	370	mz	12736758	nʧt	8	mf	76100	mpf	1	ntθ		
1207608955	ŋks	198	ndz	10918270	ŋθs	7	ŋθ	55000	msg	1	ŋg		
1180775972	ŋz	162	ŋk	10252220	nʤd	5	ŋθs	52800	ntθ	1	nf		
1160953893	ndz	118	ŋks	8732800	ŋsd	5	mpts	27400	mpsd	1	msg		
469672112	mp	87	mp	7711995	ŋd	4	nsg	9290	nʒ	1	mpsd		
415453225	nθ	71	md	1480000	nf	4	mfs	1178	mb	1	mlz		
373161873	ŋkt	69	mps	1174908	mfs	3	nzd						
368013667	nsd	53	nʧ	795200	ŋkts	3	nʃ						

word-final sequences after a long vowel

token | type frequency

25785065457	ː	1534	ː	300074201	ːkt	47	ːkt	11437768	ːfdz	9	ːnsd
5777697543	ːt	581	ːz	188190840	ːʒ	44	ːsdz	10956664	ːðz	8	ːsb
4531190679	ːl	578	ːd	125329632	ːmd	44	ːnt	8596755	ːlt	8	ːpt
3383057289	ːk	380	ːn	122195190	ːv	43	ːʧ	7841110	ːθs	6	ːsgd
2745071793	ːd	320	ːt	118881960	ːsg	42	ːfs	6850000	ːmθ	6	ːndz
2674341007	ːʧ	300	ːl	92680570	ːg	35	ːʒ	4195048	ːbd	6	ːgz
2567287240	ːz	269	ːk	88625900	ːnʧ	35	ːns	3726593	ːsb	5	ːsbz
2282237036	ːsd	200	ːdz	82064100	ːsgd	32	ːv	3645270	ːpt	5	ːldz
2084188002	ːn	176	ːs	82008209	ːb	32	ːb	1691906	ːps	4	ːʒd
1537277682	ːts	158	ːm	68107960	ːʤd	25	ːvz	1155440	ːgz	4	ːθd
1475666952	ːs	155	ːts	67513150	ːnsd	24	ːnts	710500	ːθd	4	ːsbd
1167032450	ːmz	152	ːsd	67375773	ːfd	23	ːʤd	688000	ːln	4	ːnʧt
1120891379	ːθ	141	ːlz	65200440	ːfs	22	ːvd	322300	ːsbz	3	ːzd
1094957837	ːm	128	ːks	61929836	ːp	21	ːfdz	278400	ːʒd	3	ːlts
1028976851	ːld	116	ːf	57415905	ːvz	20	ːbz	207000	ːls	2	ːln
901551391	ːks	96	ːθ	54483561	ːsdz	18	ːsg	191300	ːsbd	1	ːʃd
775476482	ːdz	91	ːnz	49966600	ːsgz	18	ːðz	149000	ːnʃ	1	ːpts
686151756	ːʤ	91	ːmz	47669695	ːnts	17	ːθs	127000	ːnʤ	1	ːpsd
570673190	ːvd	91	ːg	46699430	ːnʧt	16	ːʧt	111000	ːpts	1	ːnʒ
529237660	ːf	69	ːfd	36164000	ːzd	15	ːps	59340	ːlts	1	ːnʃ
412662183	ːnz	63	ːld	32674430	ːʃ	14	ːʃ	55600	ːnʒ	1	ːnʤ
388255337	ːnt	48	ːp	24080940	ːldz	13	ːsgz	29100	ːð	1	ːmθ
356212905	ːlz	48	ːnd	21905960	ːndz	13	ːnʧ	25800	ːksd	1	ːls
334963320	ːns	48	ːʤ	19190750	ːbz	11	ːbd	19500	ːʃd	1	ːksd
318795690	ːnd	47	ːmd	17290559	ːʧt	10	ːlt	4960	ːpsd	1	ːð

word-initial sequences involving prevocalic sonorants

token | type frequency

17281773674	j	757	j	8313403	dw	11	pw	1053700	vr	3	vr
3392208529	nj	275	nj	5520400	hw	9	θw	29910	ʃbr	3	ʃbr
1602784290	vj	135	dj	4222000	bw	8	ʃw	25800	nkr	1	sfr
1161242205	dj	132	mj	2700171	ʃw	6	zw	3870	sfr	1	nkr
691076793	mj	122	kj	1309100	θw	5	hw	32413214941	l	3471	l
597462882	hj	115	pj	680000	ʒw	3	bw	4147212366	pl	799	kl
437740969	fj	104	tj	267200	mw	2	tsw	3810715600	kl	674	fl
434670417	sdj	99	fj	258000	vw	2	mw	2003588964	bl	629	bl
388154458	kj	98	hj	191000	blw	1	ʒw	1320501911	sl	588	pl
366755003	tj	70	bj	122210	zw	1	vw	794060193	fl	501	sl
227791607	bj	43	sdj	52140	tsw	1	nw	572201900	gl	360	gl
145093569	pj	20	sgj	44700	nw	1	klw	192164700	sbl	73	sbl
15901440	θj	15	lj	44400	fw	1	fw	55977400	vl	12	sgl
6660000	ʃj	14	vj	25800	klw	1	blw	2277849	sgl	8	ʃl
6241930	lj	14	sj	36961439862	r	5457	r	1156600	zl	6	vl
2506616	sj	13	sbj	13705578264	pr	1910	pr	211300	ʃl	2	zl
2137222	sgj	11	θj	7550827745	tr	1267	tr	8750	tl	1	tl
954590	sbj	2	zj	4067538920	fr	1067	kr	41919828360	m	5748	m
231030	gj	2	gj	3927827517	br	968	br	19164955823	n	1966	n
113900	zj	1	smj	2829702833	gr	839	gr	1036239303	sm	312	sn
20700	smj	1	ʃj	2531071281	kr	653	fr	542873989	sn	231	sm
36991428022	w	2851	w	2296912459	sdr	509	dr	5627621	ʃn	12	ʃm
1883200328	kw	547	kw	1699562374	θr	431	sdr	695367	ʃm	8	ʃn
1789605565	tw	363	sw	1158557368	dr	220	sgr	469000	km	4	kn
1059046415	sw	146	sgw	369553707	sgr	171	θr	438972	kn	1	tm
306806061	sgw	143	tw	255381502	sbr	127	sbr	28800	tm	1	pn
132612450	gw	44	gw	79450655	ʃr	89	ʃr	19700	pn	1	km
53844100	pw	26	dw	22798500	sr	5	sr				

conclusions

- all clusters follow sonority seuqencing acc. to the following hierarchy: obstruents $<$ nasals $<$ l r h $<$ j w ː $<$ vowels

conclusions

- all clusters follow sonority seuqencing acc. to the following hierarchy: obstruents $<$ nasals $<$ l r h $<$ j w ː $<$ vowels
- r h (and l) only occur prevocalically, ː only postvocalically

conclusions

- all clusters follow sonority seuqencing acc. to the following hierarchy: obstruents $<$ nasals $<$ l r h $<$ j w ː $<$ vowels
- r h (and l) only occur prevocalically, ː only postvocalically
- initial clusters show a tendency for inhibiting homorganicity

conclusions

- all clusters follow sonority seuqencing acc. to the following hierarchy: obstruents $<$ nasals $<$ l r h $<$ j w ˑ $<$ vowels
- r h (and l) only occur prevocalically, ˑ only postvocalically
- initial clusters show a tendency for inhibiting homorganicity
- final clusters show a tendency for homorganicity of nasals

conclusions

- all clusters follow sonority seuqencing acc. to the following hierarchy: obstruents < nasals < l r h < j w ː < vowels
- r h (and l) only occur prevocalically, ː only postvocalically
- initial clusters show a tendency for inhibiting homorganicity
- final clusters show a tendency for homorganicity of nasals
- number of sequences (tokens/types/cluster types):

#C*j	26,761,521,824	2043	21	jC*#	163,082,882,034	23438	66
#C*w	42,240,293,465	4172	21	wC*#	80,490,153,352	7452	67
#C*r	77,457,316,267	13721	18	ːC*#	66,529,714,056	6601	75
#C*l	45,313,292,474	7124	13	lC*#	40,448,385,874	6473	40
#C*N	62,671,176,935	8284	10	NC*#	132,064,639,027	23282	49

conclusions

- all clusters follow sonority seuqencing acc. to the following hierarchy: obstruents < nasals < l r h < j w ː < vowels
- r h (and l) only occur prevocalically, ː only postvocalically
- initial clusters show a tendency for inhibiting homorganicity
- final clusters show a tendency for homorganicity of nasals
- number of sequences (tokens/types/cluster types):

#C*j	26,761,521,824	2043	21	jC*#	163,082,882,034	23438	66
#C*w	42,240,293,465	4172	21	wC*#	80,490,153,352	7452	67
#C*r	77,457,316,267	13721	18	ːC*#	66,529,714,056	6601	75
#C*l	45,313,292,474	7124	13	lC*#	40,448,385,874	6473	40
#C*N	62,671,176,935	8284	10	NC*#	132,064,639,027	23282	49

- the size of both the pre- and postvocalic sequences is constrained, “closed syllable shortening” only in “doubly closed” syllables

conclusions

- all clusters follow sonority seuqencing acc. to the following hierarchy: obstruents < nasals < l r h < j w ː < vowels
- r h (and l) only occur prevocalically, ː only postvocalically
- initial clusters show a tendency for inhibiting homorganicity
- final clusters show a tendency for homorganicity of nasals
- number of sequences (tokens/types/cluster types):

#C*j	26,761,521,824	2043	21	jC*#	163,082,882,034	23438	66
#C*w	42,240,293,465	4172	21	wC*#	80,490,153,352	7452	67
#C*r	77,457,316,267	13721	18	ːC*#	66,529,714,056	6601	75
#C*l	45,313,292,474	7124	13	lC*#	40,448,385,874	6473	40
#C*N	62,671,176,935	8284	10	NC*#	132,064,639,027	23282	49

- the size of both the pre- and postvocalic sequences is constrained, “closed syllable shortening” only in “doubly closed” syllables
- maximal clusters: #CCC, CCC# and CCC#C

conclusions

- all clusters follow sonority seuqencing acc. to the following hierarchy: obstruents < nasals < l r h < j w ː < vowels
- r h (and l) only occur prevocalically, ː only postvocalically
- initial clusters show a tendency for inhibiting homorganicity
- final clusters show a tendency for homorganicity of nasals
- number of sequences (tokens/types/cluster types):

#C*j	26,761,521,824	2043	21	jC*#	163,082,882,034	23438	66
#C*w	42,240,293,465	4172	21	wC*#	80,490,153,352	7452	67
#C*r	77,457,316,267	13721	18	ːC*#	66,529,714,056	6601	75
#C*l	45,313,292,474	7124	13	lC*#	40,448,385,874	6473	40
#C*N	62,671,176,935	8284	10	NC*#	132,064,639,027	23282	49

- the size of both the pre- and postvocalic sequences is constrained, “closed syllable shortening” only in “doubly closed” syllables
- maximal clusters: #CCC, CCC# and CCC#C
- the neatest account is that j w ː ∈ C

issues

	L ACCENTS	L-VOC ACCENTS
flu	flʉw	flʉw
fool	fuwl∼foːl	fow∼foː
foot	fʉt	fʉt
full	ful∼foːl	fow∼foː
doe	dəw	dəw
dole	dɔwl	dɔw
dot	dɔt	dɔt
doll	dɔl∼dɔwl	dɔw
dust	dəst	dəst
dull	dəl	dəw∼dɔw

- in l accents: the change occurs across w

issues

	L ACCENTS	L-VOC ACCENTS
flu	flʉw	flʉw
fool	fuwl∼foːl	fow∼foː
foot	fʉt	fʉt
full	ful∼foːl	fow∼foː
doe	dəw	dəw
dole	dɔwl	dɔw
dot	dɔt	dɔt
doll	dɔl∼dɔwl	dɔw
dust	dəst	dəst
dull	dəl	dəw∼dɔw

- in l accents: the change occurs across w
- in l-voc accents: ʉw > ow, əw > ɔw only before w < l (rule ordering?)

issues

	L ACCENTS	L-VOC ACCENTS
flu	flʉw	flʉw
fool	fuwl∼foːl	fow∼foː
foot	fʉt	fʉt
full	ful∼foːl	fow∼foː
doe	dəw	dəw
dole	dɔwl	dɔw
dot	dɔt	dɔt
doll	dɔl∼dɔwl	dɔw
dust	dəst	dəst
dull	dəl	dəw∼dɔw

- in l accents: the change occurs across w
- in l-voc accents: ʉw > ow, əw > ɔw only before w < l (rule ordering?)
- why is this optional in *dull* dəw? (Wells 1982:317)

thank

- you
- the audience at the 2016-05-19 session of ALFFA, who forced very significant changes
- Geoff Lindsey, for help with the l-vocalization data and anyway
- OTKA, #104897

links

- source of data: seas3.elte.hu/cube
- this slideshow is at seas3.elte.hu/szigetva/papers.html#24mfm
- some background at seas3.elte.hu/phono/notes/nodiphthong.html

references

- Balogné Bérces, Katalin. 2015. A hangsúlyfüggő mássalhangzó-gyengülés alrendszerei [Subsystems of stress-dependent consonant lenition]. In Katalin É. Kiss, Attila Hegedűs, and Lilla Pintér (eds.) Nyelvelmélet és dialektológia 3. Pázmány Péter Catholic University, Piliscsaba. 25–42.
- Cruttenden, Alan. 2014. Gimson's Pronunciation of English (8th ed.), London & New York: Routledge.
- Bye, Patrick and Paul de Lacy. 2008. Metrical influences on fortition and lenition. In Joaquim Brandão de Carvalho, Tobias Scheer, and Philippe Ségéral (eds.), Lenition and Fortition, Berlin/New York: Mouton de Gruyter, 173–206.
- Gimson, Charles Arnold. 1989. An introduction to the pronunciation of English. London: Edward Arnold.
- Jones, Daniel. 1960. An outline of English phonetics (9th ed.). Cambridge: CUP.
- Kristó, László. 2015. Az /r/ disztribúciója és annak eredete az angol nyelv akcentusaiban [The distribution of /r/ and its origin in accents of English]. In É. Kiss et al. 181–195.
- Lindsey, Geoff. 2012. The British English vowel system. Retrieved on 2016-05-23 from englishspeechservices.com/blog/british-vowels
- Polgárdi, Krisztina. 2015. Vowels, glides, off-glides and on-glides in English: A Loose CV analysis. Lingua 158: 9–34.
- Sweet, Henry. 1900. A New English Grammar: Logical and Historical. Oxford: Clarendon Press.
- Upton, Clive. 1995. Concise Oxford English Dictionary (9th ed.), Oxford: Oxford University Press, pronunciation editor: Clive Upton.
- Wells, John C. 1982. Accents of English. Cambridge: CUP.