Variation in the French suffix *–esque*

Due Friday, Nov. 20, 2012 to my mailbox in Campbell 3125 by 4 PM

This problem is based on Plénat 1997, with additional data from Wiktionnaire (fr.wiktionary.org/wiki/-esque) and Sajous & Tanguy 2006. You're free to consult those sources if you really want to, but I don't think it will help.

The French suffix *–esque* forms adjectives from nouns, much like its English correspondent. But some interesting phonological changes can results.

Contents

Part I: Develop an OT analysis of the basic pattern

Here are some ordinary examples of the suffix. You'll notice, for the few words where the surface form of the noun is different from the underlying form, that sometimes an underlying vowel and following nasal coda consonant combine to form a nasal vowel—but sometimes they don't (*Clinton*). You don't have to account for this.

underlying form of noun[1]	**surface** form of noun, if different from underlying	**surface** form of adjective	gloss (just the French spelling)
tyb		tybɛsk	*tube*
katakɔmb	katakɔ̃b	katakɔ̃bɛsk	*catacombe*
ʃtrumf		ʃtrumfɛsk	*schtroumpf*
prydɔm		prydɔmɛsk	*Prudhomme*
sizif		sizifɛsk	*Sisyphe*
plantigrad	plãtigrad	plãtigradɛsk	*plantigrade*
klintɔn		klintɔnɛsk	*Clinton*
aligatɔr		aligatɔrɛsk	*alligator*
santɔr	sãtɔr	sãtɔrɛsk	*centaur*
animal		animalɛsk	*animal*
brynɛl		brynɛlɛsk	*Brunel*
karnaval		karnavalɛsk	*carnaval*
katedral		katedralɛsk	*cathédrale*
karavaʒ		karavaʒɛsk	*Caravage*

[1] For convenience I'm using "r" for the French rhotic consonant, which is typically a uvular fricative or approximant.

klɛrdəlyn		klɛrdəlynɛsk	*clair de lune*
aristɔfan		aristɔfanɛsk	*Aristophane*
barsəlɔn		barsəlɔnɛsk	*Barcelone*
danbrɔn		danbrɔnɛsk	*Dan Brown*
klun		klunɛsk	*clown*

These examples show that even when the VN sequence that can change to a nasalized vowel is final, nothing much happens in the *–esque* form:

aʃiltalɔn	aʃiltalɔ̃	aʃiltalɔnɛsk	*Achille Talon*
akɔrdeɔn	akɔrdeɔ̃	akɔrdeɔnɛsk	accordéon
babuin	babuɛ̃	babubinɛsk	*babouin*
lapin	lapɛ̃	lapinɛsk	*lapin*
kaiman	kaimã	kaimanɛsk	*caïman*
kameleɔn	kameleɔ̃	kameleɔnɛsk	*chaméléon*
kamjɔn	kamjɔ̃	kamjɔnɛsk	*camion*
ʃampiɲɔn	ʃãpiɲɔ̃	ʃãpiɲɔnɛsk	*champignon*
ʃarlatan	ʃarlatã	ʃarlatanɛsk	*charlatan*
tʃjoran	tʃjorã	tʃoranɛsk	*Cioran*
dantɔn	dãtɔ̃	dãtɔnɛsk	*Danton*

One more thing you don't have to analyze: sometimes an underlying consonant deletes when word-final. But it doesn't affect the *–esque* form.

fragonard	fragonar	fragonardɛsk	*Fragonard*
kanard	kanar	kanardɛsk	*canard*
koʃmard	koʃmar	koʃmardɛsk	*chauchemar*
ʃarlot	ʃarlo	ʃarlotɛsk	*Charlot*
dykrot	dykro	dykrotɛsk	*Ducrot*
sɔldat	sɔlda	sɔldatɛsk	*soldat*
abrakadabrant	abrakadabrã	abrakadabrãtɛsk	*abracadabrant*
elefant	elefã	elefãtɛsk	*éléphant*
pedant	pedã	pedãtɛsk	*pédant*

Now the fun begins. Here are some words ending in sibilants (deleting word finally and non-deleting), arranged by syllable count (in leftmost column). These words show two different behaviors. Develop an analysis of which words do what.

1	rɔs		rɔsɛsk	*rosse*

1	buz		buzɛsk	*bouse*
1	fars		farsɛsk	*farce*
1	dœz	dœ	dœzɛsk	*(Louis) II*
2	fidjas		fidjɛsk	*Phidias*
2	gijuz	giju	gijɛsk	*Guilloux*
2	marɛz	marɛ	marɛsk	*marais*
3	bymamys		bymamɛsk	*bumammus*
3	servantɛs	sɛrvãtes	servãtɛsk	*Cervantes*
3	klitɔris		klitɔrɛsk	*clitoris*
3	kɔsinys		kɔsinɛsk	*Cosinus*
3	djafwarys		diafwarɛsk	*Diafoirus*
3	myljebris		myliebrɛsk	*muliébris*
3	klapɔtis	klapɔti	klapotɛsk	*clapotesque*
3	ʒavanɛz	ʒavanɛ	ʒavanɛsk	*javanesque*
3	sisʒurnez	siʒurne	siʒurnɛsk	*six journées*
4	seɛnɛrɛs		seɛnɛrɛsk	*CNRS*
4	øpalinos		øpalinɛsk	*Eupalinos*
4	faraminœz	faraminœ	faraminɛsk	*faraminesque*
4	galimatias	galimatia	galimatiɛsk	*galimatiesque*
4	libidinœz	libidinœ	libidinɛsk	*libidinesque*
5	ʒyljenas		ʒyljenɛsk	*juliénas*
5	mefistɔfelɛs		mefistɔfelɛsk	*Méphistophélès*
5	jœvuzɛkompriz	jœvuzɛkõpri	jœvuzɛkõprɛsk	*je vous ai compris*

Here are some words that end with a velar stop, again arranged by syllable count. They show a new type of candidate. They also show variation. Pretend that all of the velar-final words can show the same set of variants (even though that's not what the data say), and determine what are the conflicting constraints that need to be variably ranked:

1	gag		gagɛsk, gagɛst	*gag*
1	laŋg		laŋgɛsk, laŋgɛst	*Lang*
1	mark		markɛst	*Marc*
1	blag		blagɛst	*blague*
1	blɔg		blɔgɛsk, blɔgɛsk	*blog*
2	maʒik		maʒikɛsk	*magique*
2	petrark		petrarkɛsk	*Pétrarque*
2	pipik		pipikɛsk	*Pipik*
4	krɛmlinolog		krɛmlinolɛsk	*kremlinologue*
4	panɛʒirik		panɛʒirɛsk	*panégyrique*

Here are some nouns that end in vowels; they show variation. Pretend that all these words can show the same variants, and identify the conflicting constraints whose ranking must be variable:

goja		gojɛsk, gojatɛsk	*Goya*
zola		zolatɛsk	*Zola*
kaka		kakatɛsk	*caca*
nana		nanɛsk	*nana*
kɔma		kɔmatɛsk	*coma*
ferja		ferjatɛsk	*féria*
gargantɥa	gargãtɥa	gargãtɥɛsk	*Gargantua*
gevara		gevarɛsk	*Guevara*
alibaba		alibabɛsk, alibabaɛsk	*Ali Baba*
ajatɔla		ajatɔlɛsk	*ayatollah*
imalaja		imalajɛsk	*Himalaya*
pasilina		pasilinɛsk	*Paasilinna*
bede		bedeɛsk	*BD*
kɔkto		kɔktɛsk, kɔktoɛsk	*Cocteau*
tɔro		tɔrɛsk	*taureau*
gogo		gogotɛsk	*gogo*
ulipo		ulipɛsk	*Oulipo*
bigɔrno		bigɔrnɛsk	*bigorneau*
gɔbino		gɔbinɛsk	*Gobineau*
bɛrni		bɛrniɛsk, bɛrnɛsk	*Berni*
myrfi		myrfiɛsk	*Murphy*
barbari		barbarɛsk	*barbarie*
ʃəvalri		ʃəvalrɛsk	*chevalerie*
kɔsmati		kɔsmatɛsk	*Cosmati*
polini		polinɛsk	*Pollini*
sarkozi		sarkoziɛsk	*Sarkozy*
fɛlini		fɛlinɛsk, fɛliniɛsk	*Fellini*
kaligari		kaligarɛsk	*Caligari*
ʃony		ʃonyɛsk	*Chaunu*
staty		statyɛsk	*statue*
yby		ybyɛsk	*Ubu*
pjupju		pjupjɛsk, pjupjutɛsk	*pioupiou*
vodu		voduɛsk	*vaudoue*
mɔbutu		mɔbutɛsk	*Mobutu*

Part II: Modeling the variation

Plénat reports that the rates of the three variants in vowel-final words depend on the quality of the vowel and the number of syllables in the noun form. Here is a simplified version of his table, counting how many examples he found of each type:

example			/i/	/y/	/u/	/e,ɛ,ə/	/o/	/a/
nanɛsk	2 sylls	delete V	3	0	1	3	7	5
pjupjutɛsk		insert C	1	0	1	0	2	8
voduɛsk		normal	10	3	3	1	4	0
mɔbutɛsk	3 sylls	delete V	15	0	2	7	26	27
ʔ		insert C	0	0	0	0	2	2
fɛliniɛsk		normal	8	1	3	0	1	0
pasilinɛsk	4 sylls	delete V	12	1	0	7	13	16
none		insert C	0	0	0	0	0	0
alibabaɛsk		normal	1	2	0	0	0	0[2]

Inspect the table to understand the trends and think about what constraints you might need to capture them.

Open the OTSoft input file 01FrenchVariation.txt (download from course web page). You'll see that it already has inputs (including one imaginary one, /faramino/) and output candidates for all the crucial cases. In the third column is the frequency of each output, which I estimated from Plénat's data. Add the constraints that you've devised, and how many times each is violated.

Feel free to add more examples and/or more candidates, if your analysis calls for them.

Run the GLA and take a look at your results. See if you can get a better match to the input frequency by increasing the number of iterations.

Then run MaxEnt and do the same.

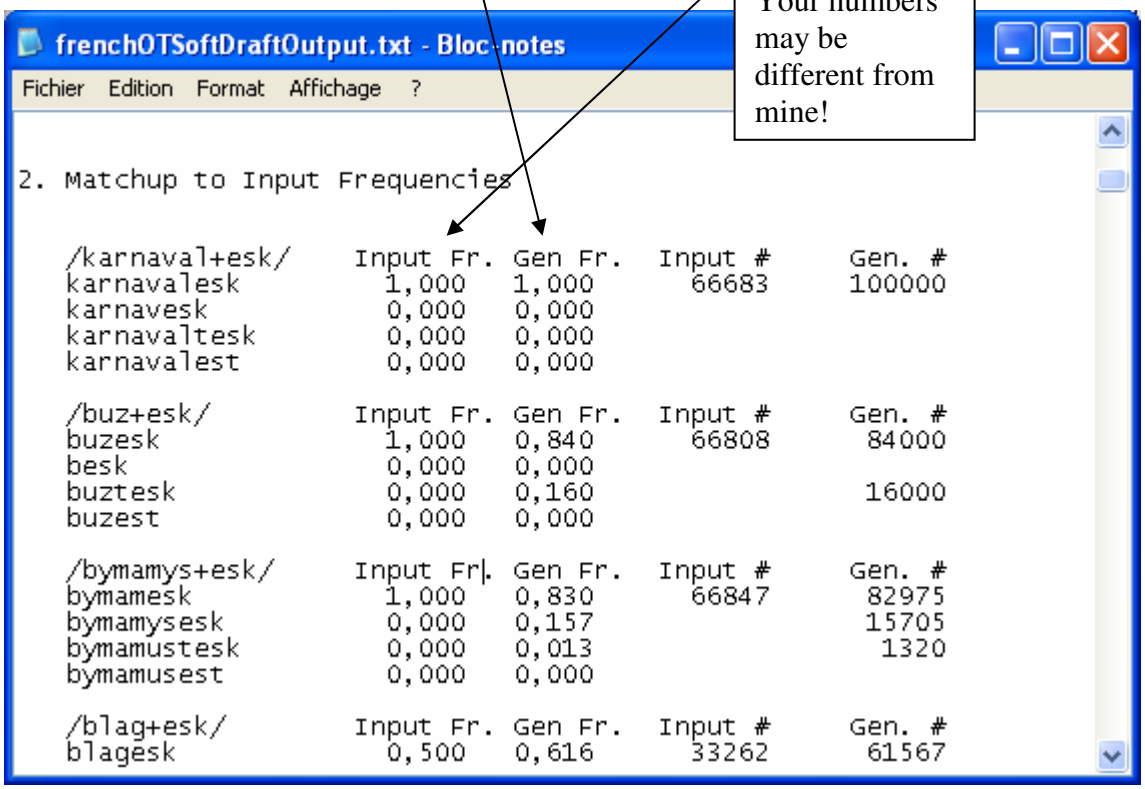
Try changing your constraint set if you're not getting at least the trends in the data—it's OK if not all the numbers match exactly though.

Part III: Comparing GLA and MaxEnt

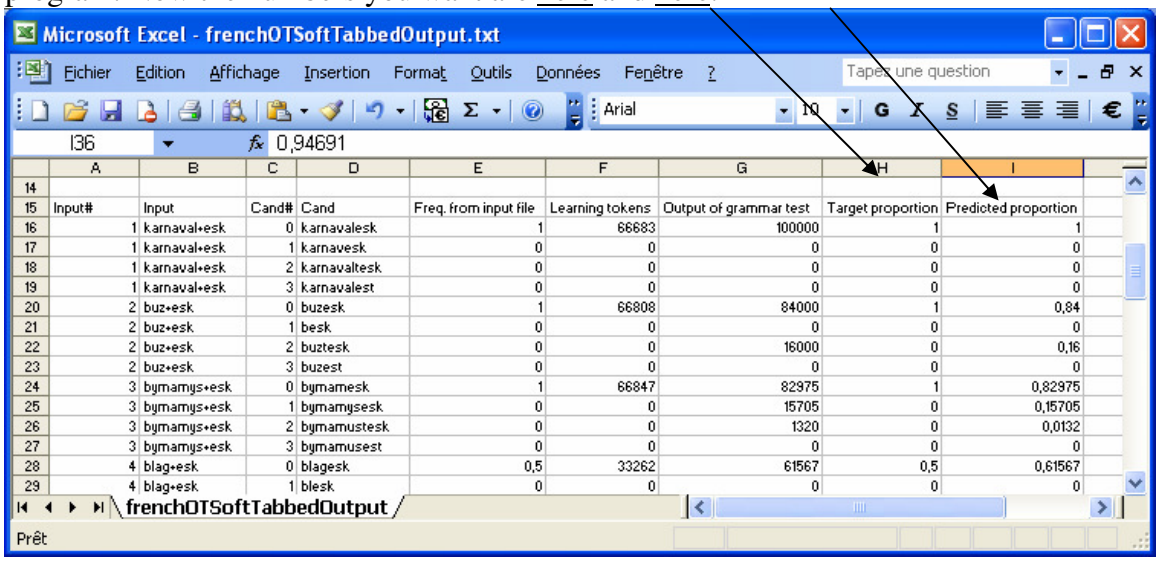
Run the GLA again, with your final constraint set. You will notice that in the folder where you saved 01VariationFrench.txt, OTSoft has created a folder called FilesFor01VariationFrench.txt. In that folder is now a file 01VariationFrenchDraftOutput.txt that contains your GLA results. Open that file (with Notepad or whatever):

[2] How can there be a zero when we have the example [alibabaɛsk]? Because that example doesn't come from Plénat's paper.

You're going to create a plot showing how well the frequencies given to the GLA in the input file match the frequencies generated by the grammar it learned. These numbers are in columns side by side in your results file.

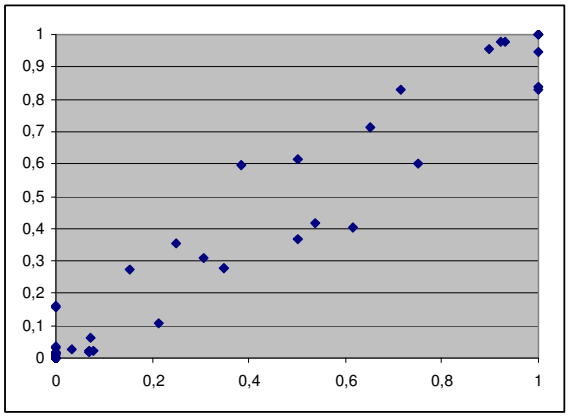


To start plotting the correlation, open the file that OTSoft produced called 01VariationFrenchTabbedOutput.txt in MS Excel, OpenOffice Calc, or any other spreadsheet program. Now the numbers you want are here and here:



	A	B	C	D	E	F	G	H	I
14									
15	Input#	Input	Cand#	Cand	Freq. from input file	Learning tokens	Output of grammar test	Target proportion	Predicted proportion
16	1	karnaval•esk	0	karnavalesk	1	66683	100000	1	1
17	1	karnaval•esk	1	karnavesk	0	0	0	0	0
18	1	karnaval•esk	2	karnavaltesk	0	0	0	0	0
19	1	karnaval•esk	3	karnavalest	0	0	0	0	0
20	2	buz•esk	0	buzesk	1	66808	84000	1	0,84
21	2	buz•esk	1	besk	0	0	0	0	0
22	2	buz•esk	2	buztesk	0	0	16000	0	0,16
23	2	buz•esk	3	buzest	0	0	0	0	0
24	3	bymamys•esk	0	bymamesk	1	66847	82975	1	0,82975
25	3	bymamys•esk	1	bymamysesk	0	0	15705	0	0,15705
26	3	bymamys•esk	2	bymamustesk	0	0	1320	0	0,0132
27	3	bymamys•esk	3	bymamusest	0	0	0	0	0
28	4	blag•esk	0	blagesk	0,5	33262	61567	0,5	0,61567
29	4	blag•esk	1	blesk	0	0	0	0	0

To make the plot, select the numbers in columns H and I, then, depending on your software, do something like Insert>Chart. You may have to fiddle a bit (come see me for help!), but you want a picture like this:



To get a numerical measure of the correlation, use Excel's CORREL() function. For the numbers that went into the plot above, the result is 0.978 (that's pretty good—the highest possible is 1).

Now do the same for the MaxEnt grammar. Save your TabbedOutput file under a different name, because it will get overwritten next time you run OTSoft. Run OTSoft again, this time using Maximum Entropy. When you open the TabbedOutput.txt file, this time the numbers of interest will be here and here:

Microsoft Excel - frenchOTSoftTabbedOutput.txt

Fichier Edition Affichage Insertion Format Outils Données Fenêtre ?

B1 fx Candidate

	A	B	C	D	E
1	Input	Candidate	Freq. in input file	Target proportion	Predicted proportion
2					Weights:
3	karnaval•esk	karnavalesk	1	1	0,999985
4	karnaval•esk	karnavesk	0	0	0,00001
5	karnaval•esk	karnavaltesk	0	0	0
6	karnaval•esk	karnavalest	0	0	0,000006
7	buz•esk	buzesk	1	1	0,86449
8	buz•esk	besk	0	0	0
9	buz•esk	buztesk	0	0	0,135505
10	buz•esk	buzest	0	0	0,000005
11	bymamys•esk	bymamesk	1	1	0,843252
12	bymamys•esk	bymamysesk	0	0	0,135507
13	bymamys•esk	bymamustesk	0	0	0,02124
14	bymamys•esk	bymamusest	0	0	0,000001
15	blag•esk	blagesk	0,5	0,5	0,614785
16	blag•esk	blesk	0	0	0
17	blag•esk	blagtesk	0	0	0,016353
18	blag•esk	blagest	0,5	0,5	0,368862
19	pipik	pipikesk	0,75	0,75	0,599615
20	pipik	pipesk	0	0	0,024676
21	pipik	pipiktesk	0	0	0,015949
22	pipik	pipikest	0,25	0,25	0,35976

frenchOTSoftTabbedOutput

Prêt

Once again, make a plot of these two columns, and check the correlation.

What your write-up should include

- An analysis of the data in Part I (which may have changed after you did parts II and III). Include suitable examples and tableaux, and make clear which constraints must be variably ranked and why.
- A discussion of how well the GLA and MaxEnt did at matching the data. Include a scatterplot for each of the two models (like the one illustrated above) and report the correlation results. Are there items that both models do badly on? If so, discuss why that might be (e.g., there doesn't seem to be any high-ranked constraint favoring the more-common output). Are there items that just one model does badly on?
    - Correlation coefficient is a rather crude measure of how close the match was. If you'd like to try some additional measure(s) and know how, feel free.

References

Plénat, Marc. 1997. Analyse morpho-phonologique d'un corpus d'adjectifs dérivés en -esque. *Journal of French Language Studies* 7. 163-179.

Sajous, Franck & Ludovic Tanguy. 2006. Répérage de créations lexicales sur le web francophone.. Paper presented at the ATALA, Le web comme ressource pour le TAL.