MFM FRINGE – MAY 28TH 2025

SUBLEXICA ACROSS LANGUAGES

Abstract booklet

MFM FRINGE – MAY 28TH 2025
SUBLEXICA ACROSS LANGUAGES

Wednesday the 28th of May

Venue : University of Manchester, Samuel Alexander building, room A101.

1:00-1:30	Welcome – Opening remarks	Welcome – Opening remarks
1:30-2:00	**Bartlomiej Czaplicki**	How specific are generalizations governing allomorph distribution?
2:00-2:30	**Alireza Jaferian**	Word-internal CCCs in Persian: a Core-Periphery account
2:30-3:00	**Joseph Greeshma**	Lexical Restrictions in Malayalam Vowel Harmony: Evidence for Core-Periphery
3:00-3:30	**Renate Raffelsiefen**	Capturing contrasts via input structure
3:30-4:00	Break	Break
4:00-4:30	**Stanley Nam**	Phonotactics and sublexical structure in Korean L-Tensification
4:30-5:00	**Charles Vancaeyzeele**	The analysis of "learned" stems in French: an alternative to sub-lexica
5:00-5:30	**Ali Idrissi, Ali Nirheche**	Lexical stratification in Moroccan Arabic diglossic grammar
5:30-6:00	**Xevi Pujol Molist**	Loanword adaptation in Central Catalan: towards a characterisation of lexical strata

How specific are generalizations governing allomorph distribution?

Bartłomiej Czaplicki
University of Warsaw

In many languages the lexicon is divided into native and foreign sublexica governed by different phonological rules (Itô & Mester 1995). The division of the lexicon sometimes hinges on phonotactic properties (Becker and Gouskova 2016). However, less is known about how various sublexica interact in a single grammar. Such issues are inextricably linked to the on-going debate about the degree of specificity of generalizations that speakers track. Are constraints general and universal (Selkirk 1982) or specific and arbitrary (Kapatsinski 2013)?

Simulations in Noisy Harmonic Grammar (NHG, Boersma & Pater 2016), a probabilistic framework, were run on Polish Locative Adjectives (LAs) to establish the regularities that govern the distribution of allomorphs (OTSoft, Hayes et al. 2013). A LA refers to place names and is formed using an optional intermorph and an obligatory suffix. There are three possible intermorphs and two suffixes: *root* + {ij, aɲ, ɛɲ} + {sk, tsk}. 2,503 LAs have been extracted from a large corpus and analyzed.

Different preferences for affixes are found in foreign, compared to native LAs (foreign LAs are derived from place names located outside of Poland). Intermorphs are avoided in native LAs but commonly found in foreign LA. The avoidance in native LAs is categorical for -ij- and gradient for -aɲ- and -ɛɲ- (all three are statistically significant). Phonotactic properties also play an important role. Extrasyllabic sonorants (those that are preceded by a consonant in the base and thus become "trapped" between two consonants in a LA without an intermorph, CCsk) show preferences distinct from those exhibited by non-extrasyllabic sonorants (preceded by a vowel). Specifically, extrasyllabic sonorants select intermorphs significantly more often than non-extrasyllabic sonorants, both in native and foreign LAs. These differences between sublexica are reflected in differing weights of pertinent constraints in the NHG simulations.

How specific are the constraints governing the distribution of allomorphs? For example, can preferences for allomorphs differ for individual consonants? Simulations in NHG have been run on native and foreign LAs in three conditions with increasing specificity of constraints:

low: well-established universal markedness and faithfulness constraints, e.g. SSP (Selkirk 1982), Max, Dep.

moderate: incl. generalizations specific to each base-final consonant, e.g. g → ∅-sk (base-final /g/ selects -sk- without an intermorph),

high: incl. generalizations specific to each consonant in the extrasyllabic vs. non-extrasyllabic context, e.g. Cn → -ɛɲ-sk, Vn → ∅-sk-.

The average error per candidate, shown in Fig. 1, indicates how good a model is. Models in the *low* condition predict the data poorly. By comparison, the *moderate* and *high* conditions show low error rates, which supports the claim that speakers rely on consonant-specific generalizations. Crucially, they also rely on the affiliation with a sublexicon defined on the basis of phonotactic properties. When a sonorant is preceded by a consonant it shows different preferences for allomorphs than when it is preceded by vowel. This conclusion follows from the lower error rates in the *high* than in the *moderate* condition (0.539% vs. 1.061% for foreign LAs) and confirms the hypothesis that speakers use phonotactic properties to delimit sublexica.

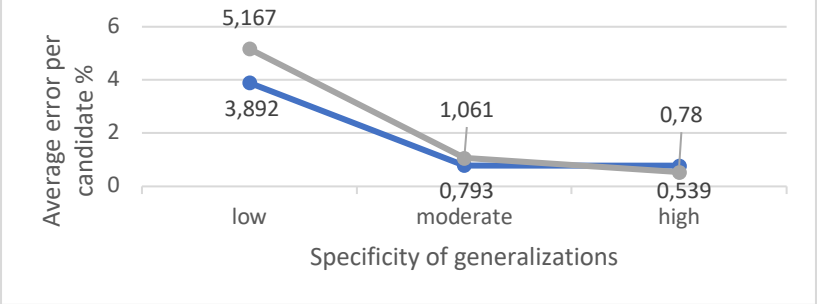


Figure 1. Average error per candidate from NHG simulations: native LAs (blue) and foreign LAs (gray)

References

Becker, Michael & Gouskova, Maria. 2016. Source-oriented generalizations as grammar inference in Russian vowel deletion. *Linguistic Inquiry* 47(3). 391–425.

Boersma, Paul & Pater, Joe. 2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In McCarthy, John & Pater, Joe (eds.), *Harmonic Serialism and Harmonic Grammar*, 389–434. Sheffield: Equinox.

Hayes, Bruce & Tesar, Bruce & Zuraw, Kie. 2013. "OTSoft 2.5" software package, http://www.linguistics.ucla.edu/people/hayes/otsoft/.

Itô, Junko & Mester, Armin. 1995. The core-periphery structure of the lexicon and constraints on reranking. In Beckman, Jill & Urbanczyk, Suzanne & Walsh, Laura (eds.), *University of Massachusetts Occasional Papers in Linguistics Vol. 18: Papers in Optimality Theory,* 181–209. University of Massachusetts, Amherst: GLSA.

Kapatsinski, Vsevolod. 2013. Conspiring to mean: Experimental and computational evidence for a usage-based harmonic approach to morphophonology. *Language* 89(1). 110–148.

Selkirk, Elisabeth. 1982. *The Syntax of Words.* Cambridge, MA: MIT Press.

Word-internal CCCs in Persian: A Core-Periphery account

Alireza Jaferian. SFL CNRS. jaferian@gmail.com

Introduction. In this talk, I propose that in colloquial Iranian Persian, a.k.a. Farsi, loans containing tautomorphemic internal CCCs pertain to a peripheral stratum of the lexicon, where violation of the *CCC markedness constraint is tolerated.

The data. The Persian lexicon has more than 700 words containing internal CCC clusters. Most of these are compounds, including morpheme boundary following the second C of the cluster: /ga**rm**+**k**on/ "warm+do =sweater". In this language, morpheme boundary has been analyzed as having similar properties as word boundary (Jaferian 2024). The remaining 34 words are exclusively loans, mostly English (e.g. /ʔe**ksp**eres/ "express") and French (e.g. /po**rtr**e/ "portrait"), but also including one Russian (/ʧo**rtk**e/ "abacus") and two Turkic words (e.g. /su**rtm**e/ "sled"). Most of these loans surface with no simplification.

An existing Core-Periphery analysis of Persian loans. Using the OT model of Core-Periphery (Itô & Mester 1995, 1999), Ariyaee (2019) proposed that the Iranian Persian lexicon consists of three strata: Core (spoken Persian-Arabic words), Middle (written Persian-Arabic words) and Periphery (unassimilated foreign words). He defined a set of constraints called PERSIANPHONOLOGY, such as *#CC (*COMPLEXONSET), which are never violated in surface forms in any of the three strata. At Core level, consisting of colloquial Persian and Arabic words, inputs surface faithfully and no constraint violations are observed, while in the Middle stratum, consisting of literary words, the pre-nasal raising of the low back vowel /ɑ/ is blocked, violating the *[ɑN] constraint: /zemɑm/ → [zemɑm] (*[zemum]) "harness". The Periphery stratum hosts "unassimilated foreign words" where *CCC# is violated, e.g. FR "décembre" ("December") surfacing as [desɑ**mbr**] (simplified [desɑmr] is as well attested). Ariyaee does not take word-internal CCCs into account in his analysis.

Updating Ariayee's analysis. Of the 34 words with apparently tautomorphemic internal CCCs, a dozen can arguably be analyzed as containing morpheme boundary. These are composed of two distinct morphemes in the source language, which is reflected in their written form in Persian: FR "force majeure" adapted as /fo**rsm**ɑʒor/. While Persian natives do not have access to the morphological structure of these words, the latter are mostly learned through their Persian written forms, including a space between the two morphemes. I therefore argue that they are acquired as compounds. As such, they present no violation of *CCC, banning tautomorphemic clusters of three consonants. The remaining words with no synchronic morpheme boundaries, such as /terɑ**nsf**er/ "transfer" and /ʔa**nst**itu/ "institute", surface with their CCC, thereby violating *CCC. I propose to integrate these into Ariyaee's analysis and to make them pertain to the Periphery level, along with other unassimilated foreign words. Note that *CCC# is a special case of *CCC and not an independent constraint.

Discussion. Lexical items hosted by the Periphery stratum are not necessarily recent loans. While they notably include 21st century words such as /ʔi**nst**ɑgerɑm/ "Instagram", they are not limited to these. /terɑ**nsf**er/ "transfer/, /ʔe**jrb**ɑs/ "Airbus", /mɑ**rks**ism/ "Marxism" (20th century or older), and /su**rtm**e/ "sled" and /ju**rtm**e/ "gallop" (probably much older) belong to this layer of sublexion as well, exclusively consisting of non-Arabic loans.

Simplification of some of these CCCs can optionally occur in surface forms, as a function of the foreign-language knowledge of the speaker. This constitutes another interesting question, which, time permitting, I will explore during the talk.

Greeshma Joseph
Indian Institute of Technology Delhi

Lexical Restrictions in Malayalam Vowel Harmony: Evidence for Core-Periphery

A prosody-induced vowel height harmony occurs in the Malayalam [+Dravidian][1] words wherein the short high vowels /i/ and /u/ in the initial syllable are lowered to the corresponding mid-vowels when the low vowel /a/ is in the adjacent syllable as in (1).

1		UR	SR	gloss
	[i] → [e]	*ila*	*ela*	'leaf'
		iɻam	*eɖam*[2]	'space'
	[u] → [o]	*kuɻa*	*koɖa*	'umbrella'
		cuma	*coma*	'cough'

The *high-low* vowel sequence on the left edge is marked as the initial stressed syllable has a low sonority vowel followed by a high sonority vowel. According to the stringency markedness scale (De Lacy, 2004) given in (2), the Malayalam locus compromises on the syllable well-formedness.

2	stressed syllable		unstressed syllable	
	*σˈ/[(i,u),(e,o),(a)] *σˈ/[(i,u),(e,o)] *σˈ/[(i,u)]	high vowel (low sonority) is the least preferred	*σ/[(a),(e,o),(i,u)] *σ/[(a),(e,o)] *σ/[(a)]	low vowel (high sonority) is the least preferred

The resolution of this markedness is achieved with vowel lowering (Align- L, Ident-Low >>Ident-High), making the initial syllable the target for augmentation. Moreover, it is observed that this harmony is restricted and sometimes modified in other parts of this stratified lexicon. Phonotactically, these parts occupy the periphery of the Malayalam lexicon. Sanskrit loanwords are immune to vowel lowering even though these lexical items have similar syllable composition as the core vocabulary as in (3a).

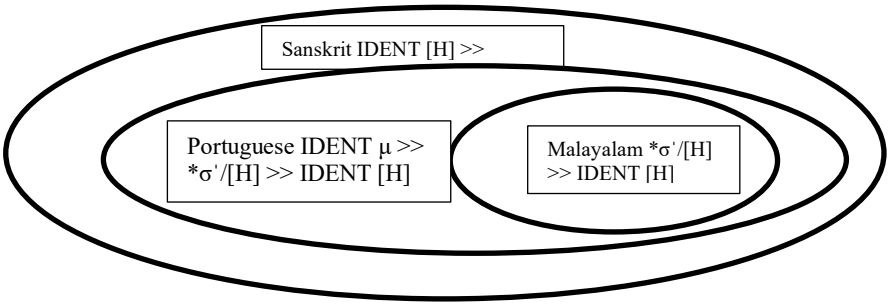
3a	Sanskrit root →Malayalam		
	sukʰ	*sugam*[3]	'well-being'
	ɖin	*ɖinam*	'day'
	ɲiʃa	*ɲiʃa*	'night'

In Portuguese loans, vowel lowering is not confined only to the initial syllable. However, both diphthongs and short vowels are lowered to corresponding long vowels to preserve the mora (μ) as in (3b).

3b	Portuguese root → Malayalam					
	leilao	*leːlam*	'auction'	*səbɔla*	*səvoːʃa*	'onion'
	kaseira	*kaseːra*	'chair'	*mesa*	*meːʃa*	'table'

The core-periphery structure for Malayalam lexicon with predominant constraints is given in (4)

(4)[4]



[1] Mohanan (1987) uses [+Dravidian] for native Malayalam
[2] Intervocalic voicing
[3] 'am' is the nominalizer in Malayalam
[4] [H] - feature 'high'

Capturing contrasts via input structure

Renate Raffelsiefen, Leibniz-Institut für Deutsche Sprache, Mannheim

Stress contrasts in simplexes demonstrate lexical stress in German. Such contrasts are characterized by an asymmetry, where for each type one pattern constitutes the default (s. 1a,b):

(1)	Type	"special" stress	default stress
a.	disyllabic ending in full vowel	/byˈʀo/ <Büro> 'office'	/ˈʒiʀo/ <Giro> 'giro'
c.	trisyllabic ending in schwa	/ˈhɛʀˌbɛʀgə/ <Herberge>	/ʃimˈpanzə/ <Schimpanse>

The respective default pattern is motivated by stability while "special" stress manifests in the emergence of variants (e.g. /byˈʀo/ ~ /ˈbyʀo/) and (diachronic) stress shifts (e.g. /kɑˈnu/ > /ˈkɑnu/ <Kanu>. The asymmetry is captured in the simple model illustrated in (2), where inputs consist of phonological forms encountered by learners. The constraint IO-FAITH(Stress) requires the stress perceived by learners in acquisition to be preserved in the output. TROCHEE stands for a set of constraints favoring leftheaded binary feet.

(2)a.

/byˈʀo/ <Büro>	IO-FAITH(Stress)	TROCHEE
☞ /byˈʀo/		*
/ˈbyʀo/	*!	

b.

/ˈʒiʀo/ <Giro>		
/ʒiˈʀo/	*!	*
☞ /ˈʒiʀo/		

The asymmetry between "special" and default stress is captured in that for stress-final inputs there is no ideal candidate while for trochaic inputs there is one which satisfies both constraints. A candidate with "special stress" can be optimal only when that stress is already present in the input, while default stress is optimal not only for inputs with default stress (see (2b)) but also for those with no stress present in the input (assuming that IO-FAITH(Stress) is vacuously satisfied then), such as written forms. This accounts for novel acronyms such as <*ePa*> (< *elektronische Patientenakte "electronic patient file*"), which can be pronounced only /ˈepɑ/, not */eˈpɑ/.

Actual instability of special stress can come about in various ways, including unrankedness of the two constraints, which would yield a tie to be broken by lower-ranking markedness constraints. This could also favor the special stress candidate, resulting in phonologically defined "islands of stability" such as disyllabic words lacking a word-initial onset as in (3).

(3) /eˈtɑ/ <Etat>, /iˈde/ <Idee>, /eˈklɑ/ <Eklat>, /arˈme/ <Armee>, /aˈle/ <Allee>, /aˈde/ <Ade>

Positing inputs with stress (/eˈtɑ/ <Etat>) versus those with no stress (/epɑ/ <*ePa*>) allows for reconciling absolute stability of special stress in the ordinary word with absolute regularity of default stress when pronouncing the acronym by invoking a lower-ranking constraint ONSET(Σ) (A foot must have an onset). The model further predicts that special stress can arise only with prominence already present when respective words first became part of the vocabulary. Indeed special stress *can* invariably be traced to words exhibiting the respective prominence: in particular loanwords (e.g. /tɑˈbu/ <Tabu> from Engl. /təˈbu/ <taboo>) or fused compounds, with the prominence of the prosodic head preserved (/ˈhɛʀbɛʀgə/ from OHG (ˈheri)$_{\omega HD}$(berga)$_{\omega}$).

To further clarify: Inputs are here conceived not as raw acoustic data but as phonological forms perceived by learners based on their language-specific knowledge. German speakers presumably parse phrase final prominence in French /byʀo/ as foot structure (i.e. (by(ʀo)$_{\Sigma}$)$_{\omega}$) in the input and also the winning candidate. Such a model then refers to a single phonological abstractness level, (prosodically organized) phonemic structure. Accounting for learnability, fine-grained stability patterns, and diachrony in "one stratum" challenges the core-periphery model (Itô & Mester 1995), which invokes multiple strata based on diacritics linked to lexical items.

Phonotactics and sublexical structure in Korean L-Tensification

Stanley Nam (University of British Columbia)

In Korean, a lateral sound selectively tensifies the following lenis coronal obstruent in a process commonly known as L-Tensification (Kim-Renaud, 1974). As the schema in (1) illustrates, lenis obstruents /t, s, tɕ/ become [t*, s*, tɕ*] when LT applies.

(1) L-Tensification (modified from Kim-Renaud (1974))
[−son, +cor] → [+tense] / l ___
"A coronal obstruent becomes tense following a /l/."
(NB: Kim-Renaud explicitly states that it only applies to Sino-Korean words.)

LT does not categorically apply to all words that satisfy the phonological context, following cross-linguistic patterns where the lexicon exhibits subgroups with distinctive phonotactic constraints (Chae, 1999; Itô & Mester, 1999; Kang, 1998), perceptual biases (Moreton & Amano, 1999) and rule application (Bakker & Papen, 1997; Mifsud, 1995). After Kim-Renaud (1974), LT is widely understood as etymologically-grounded: only Sino words (i.e., originating from ancient Chinese or consisting of Sino-Korean morphemes) undergo LT. This study challenges this view and claims that phonotactics may be a better predictor of LT.

Etymological accounts are inadequate on both conceptual and empirical grounds. Origins of a word may not be easily accessible in Korean since Sino and non-Sino sublexica share the same phoneme inventory and orthography. Compare this to other languages with similar selective rules: Michif has different vowel inventory between French and Cree strata (Bakker & Papen, 1997), and Japanese has a distinctive orthographic system (Itô & Mester, 1999; Kaiser et al., 2013, pp. 4-7; Martin, 1972, p. 90). Besides, synchronically, there are mismatches between rule application and etymology: there are limited number of Sino words with /l/-[CORONAL] sequences that do not undergo LT and non-Sino words with /l/ tensifying the following coronal segment.

To investigate whether phonotactics can be a better predictor of LT, I conducted a nonce word production experiment using phonotactic minimal sets consisting of three classes of non-existing words: (i) ambiguous, (ii) Sino-like, and (iii) non-Sino-like. Ambiguous base words do not have any phonotactic characteristics reported in previous studies in Korean phonology (Chae, 1999; Kang, 1998; Park, 2014, 2020; Park, 2023; Park et al., 2013). Minimal changes were made to each base word to create two biased words: one geared toward Sino, and the other toward non-Sino.

The unvoiced duration and closure duration around the target (i.e., the obstruent segment after /l/) were measured in the native speakers' productions. Since tense sounds exhibit longer durations, the application of LT would manifest as a long duration. The observations were statistically analyzed using a linear mixed-effects model.

Tables 1 and 2 present the experimental results. In both cases, non-Sino phonotactics were associated with significantly reduced durational measures, indicating a lower likelihood of applying LT. These results support the phonotactic hypothesis: LT applies to some, but not all, nonce words, and is less likely in those with non-Sino phonotactics. In contrast, the findings challenge etymological accounts, which predict no LT in nonce words that lack etymology.

Table 1 Fixed-effects estimates for unvoiced duration

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.0760	0.0065	53.32	11.78	< 2e-16
phntcs2*	−0.0053	0.0020	23.12	−2.66	0.0140
phntcs3*	0.0019	0.0032	38.82	0.58	0.5623
man_loc	0.0055	0.0050	30.78	1.10	0.2806
LT_loc	0.0481	0.0046	22.08	10.43	5.39e-10
phntcs2:man_loc	0.0069	0.0038	31.12	1.81	0.0798
phntcs3:man_loc	0.0006	0.0056	28.74	0.11	0.9103

*** Note:** phntcs2 = Neutral → Non-Sino, phntcs3 = Neutral → Sino

Table 2 Fixed-effects estimates for closure duration

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.0639	0.0055	51.20	11.57	6.67e-16
phntcs2*	−0.0045	0.0028	35.22	−1.65	0.1081
phntcs3*	−0.0007	0.0023	33.16	−0.29	0.7725
man_loc	0.0092	0.0040	29.08	2.27	0.0308
LT_loc	0.0230	0.0040	30.03	5.71	3.15e-06
phntcs2:man_loc	0.0010	0.0050	27.86	0.20	0.8433
phntcs3:man_loc	−0.0064	0.0042	29.67	−1.52	0.1380
phntcs2:LT_loc	0.0002	0.0050	29.12	0.04	0.9722
phntcs3:LT_loc	0.0007	0.0043	30.92	0.16	0.8738
man_loc:LT_loc	0.0033	0.0081	28.83	0.41	0.6871
phntcs2:man_loc:LT_loc	−0.0157	0.0100	27.59	−1.58	0.1264
phntcs3:man_loc:LT_loc	0.0212	0.0084	29.30	2.53	0.0172

*** Note:** phntcs2 = Neutral → Non-Sino, phntcs3 = Neutral → Sino

Bibliography

Bakker, P., & Papen, R. A. (1997). Michif: A mixed language based on Cree and French. In S. G. Thomason (Ed.), *Contact languages: a wider perspective* (pp. 295-363). John Benjamins.

Chae, S.-y. (1999). Umun pyenhwaey nathanan hankwuke ehwiuy chungwikwuco [The core-periphery structure in the Korean lexicon reflected in a phonological variation and change]. *Studies in Phonetics, Phonology and Morphology*, *5*, 217-236.

Itô, J., & Mester, A. (1999). The Phonological Lexicon. In N. Tsujimura (Ed.), *The Handbook of Japanese Linguistics*. Wiley-Blackwell. https://doi.org/10.1002/9781405166225.ch3

Kaiser, S., Ichikawa, Y., Kobayashi, N., & Yamamoto, H. (2013). *Japanese: a Comprehensive Grammar*. Taylor & Francis Group. http://ebookcentral.proquest.com/lib/ubc/detail.action?docID=1114712

Kang, Y. (1998). Hankwuke ehwipu kwuco [The organization of lexicon in Korean]. *Studies in Phonetics, Phonology and Morphology*, *4*, 55-67.

Kim-Renaud, Y.-K. (1974). *Korean consonantal phonology* [PhD Dissertation, University of Hawai'i]. Honolulu, Hawai'i.

Martin, S. (1972). Nonalphabetic writing systems: some observations. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading* (pp. 81-102). MIT Press.

Mifsud, M. (1995). *Loan verbs in Maltese: A descriptive and comparative study*. Brill.

Moreton, E., & Amano, S. (1999). Phonotactics in the perception of Japanese vowel length: Evidence for long-distance dependencies. 6th European Conference on Speech Communication and Technology, Budapest, Hungary.

Park, N. (2014). Machine learning of phonotactic constraints in Korean nouns. *Studies in Phonetics, Phonology and Morphology*, *20*(3), 297-322. https://doi.org/10.17959/sppm.2014.20.3.297

Park, N. (2020). *Hankwuke umsopayyelceyyakuy thongkyeycek haksupkwa cekhyengseng phantan* [PhD dissertation, Seoul National University]. Seoul, Korea. https://s-space.snu.ac.kr/handle/10371/167803

Park, S. (2023). Machine learning classification of native Korean, Sino-Korean and loanwords based on phonotactics. *Studies in Phonetics, Phonology and Morphology*, *29*(3), 329-349. https://doi.org/10.17959/sppm.2023.29.3.329

Park, S., Hong, S.-h., & Byun, K. (2013). Hankwukeuy ehwikyeychungkwa umunloncek pokcapseng [Lexical strata in Korean and phonological complexity]. *Studies in Phonetics, Phonology and Morphology*, *19*(2), 255-274. https://doi.org/10.17959/sppm.2013.19.2.255

The analysis of "learned" stems in French: an alternative to sub-lexica

Charles Vancaeyzeele – LLL, University of Orléans (France)

1. Background and data. Since the 9th century AD, French has extensively borrowed from Latin, progressively saturating its lexicon with etymological doublets (Bertrand 2011). These doublets combine, among other things, a "learned" (L) with a "non-learned" (¬L) stem, both of which belong to the same morphological paradigm if they are semantically related (cf. Mel'čuk 1994). A few derivational paradigms (out of 218 collected) are illustrated in (1). Each one of them exhibits alternations between an L-stem (1c) and an ¬L-stem (1a, 1b), revealing vowel alternation sites. This study proposes a synchronic account of the existence of L-stems.

1)

	[ɛ]~[a]		[œ]~[o]		[ɛ/e]~[i] & Ø~[e]	
a.	[klɛʁ]	"bright"	[œʁ]	"hour"	[lɛtØʁ]	"letter"
b.	[klɛʁ-jɛʁ]	"glade"	[œʁ-ɛt]	"short hour"	[letØʁ-e]	"scholar"
c.	[klaʁ-te]	"brightness"	[oʁ-ɛʁ]	"hourly"	[litɛʁ-atyʁ]	"literature"

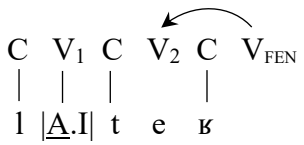
2. Proposition. Some could claim that L-stems inhabit a sub-area of the lexicon, or at least its periphery, in which some exceptional phonological activity is at hand (Itô & Mester 1995, Dabouis & Fournier 2022). This is justifiable, as L-stems do exhibit peculiar phonotactic and segmental properties (e.g. *CC#, *[œ]/[ø]). However, this approach entails that L-stems and ¬L-stems constitute two separate lexical entries, thereby burdening the lexicon tremendously. Rather, it is argued that they share the same lexical representation (cf. Schane 1968, Schane & Boulakia 1973, Dell & Selkirk 1978, Tamine 1982). The emergence of one over the other is attributed to different phonological computations, depending on the morphological context.

3. Framework. L-derivatives (as in 1c) are taken to be formed from the concatenation of an ¬L-stem and a suffix that is idiosyncratically specified for triggering the emergence of an L-stem, i.e. an L-suffix. This process is referred to as "learned derivation" (*ibid.*). Compare L-derivatives in (1c) (e.g. /klɛʁ/ + /-te/$_L$ → [klaʁ-te]) with ¬L-derivatives in (1b) (e.g. /klɛʁ/ + /-jɛʁ/ → [klɛʁ-jɛʁ]). It is suggested that the emergence of L-stems is triggered by the application of a (learned) co-phonology (cf. Orgun 1996, Sande *et al.* 2020), labeled "Φ_L", which gathers specific phonological settings. Two of these settings, which are defined accordingly with the principles of Strict CV (Lowenstamm 1996, Scheer 2004) and Element Theory (Kaye *et al.* 1985, Backley 2011), are shown in (2). Fundamentally, Φ_L applies to an ¬L-stem whenever it is selected by an L-suffix.[1]

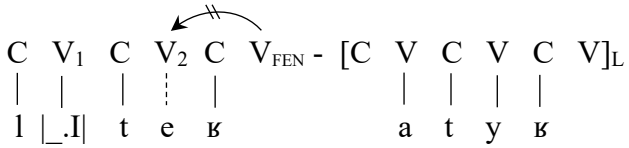
2) Φ_L { V_{FEN} may not govern/license; No phonological headedness }

4. Analysis. (3) displays the representations of [lɛtʁ] and [litɛʁ-atyʁ]. The former is illustrated in (3a), to which the L-suffix [-atyʁ]$_L$ is added in (3b). Both vowel alternations, i.e. [ɛ/e]~[i] and Ø~[e], are accounted for on the basis of Φ_L. On one hand, the head element |A| in V_1 gets deleted in (3a), yielding surface [i] at this position in (3b). On the other hand, the governing potential of the final empty nucleus (V_{FEN}) inside the stem is recalibrated. It is no longer able to govern V_2 in (3b); /e/ must therefore emerge so that V_2 satisfies the ECP.

3) a.



b.



5. Conclusion. French phonology seems to be divided into two components: a "default" component and a "learned" component (Φ_L), the latter being active only in some morphological contexts. Most importantly, Φ_L is not a primitive device, but merely a by-product of the pressure that Latin lexical items have exerted on the French lexicon for several centuries.

[1] Φ_L should not be confused with a set of phonological constraints, as used in computational approaches like Optimality Theory (Prince & Smolensky 1993, McCarthy & Prince 1993). Instead, these settings are more suited to a representational approach, directly regulating the interaction between positions and the substance of segments.

Lexical stratification in Moroccan Arabic diglossic grammar

Ali Nirheche (UMass Amherst) and Ali Idrissi (Qatar University)

The total assimilation of the Arabic definite marker /l-/ is typically triggered by stem-initial coronal consonants, except for the alveopalatal fricative [ž], which never assimilates in Standard Arabic (StA) and shows variability in Moroccan Arabic (MA) (see (1)). Heath (1987), Harrel (1962) and Freeman (2016) argue that StA borrowings into MA, such as (1f), resist l>ž-assimilation. In a corpus study, Nirheche (2025) argues that l>ž-assimilation is phonologically conditioned but shows that it is variable: occurring in 96% of cases when [ž] precedes a consonant (1b), 84% before a schwa (1d), and 37% before a full vowel (1f).

(1)	Standard Arabic				Moroccan Arabic			Gloss
a.	/l-žamal/	[lžamal]	*[žžamal]	b.	/l-žməl/	[žžməl]	*[lžməl]	‘camel (DEF.)’
c.	/l-žarab/	[lžarab]	*[žžarab]	d.	/l-žərba/	[žžərba]	*[lžərba]	‘scabies (DEF.)’
e.	/l-žanuub/	[lžanuub]	*[žžanuub]	f.	/l-žanub/	*[žžanub]	[lžanub]	‘south (DEF.)’

Building on these findings and departing from these authors, we argue for a stratified MA lexicon distinguishing a native/nativized (MA) ‘core’ from a non-nativized (etymologically close to StA) ‘periphery.’ Core stems systematically undergo l>ž-assimilation (2a-b), while peripheral ones do not (2c-d). Crucially, both strata are parts of the synchronic MA lexicon, whose architecture is shaped by diglossia (Idrissi et al. 2021).

(2)	Core (MA)			Peripheral (StA)			
a.	/l-žəbha/	[žžəbha]	‘forehead’	c.	/l-žabha/	[lžabha]	‘frontline’
b.	/l-žayħa/	[žžayħa]	‘misfortune; disaster’	d.	/l-žaaʔiħa/	[lžaaʔiħa]	‘pandemic’

In addition to the fact that it predicts the existence in the language of such near-cognates as in (2), this view of the MA lexicon also predicts that core vs. peripheral lexical affiliation would influence the grammatical behavior of a word relative to assimilation and beyond. French borrowings, for example, integrate into the core and fully assimilate (e.g., French *jaquette* /l-žakita/ > [žžakita] ‘jacket’). Additionally, core and peripheral cognates exhibit distinct phonological traits, e.g., native MA words lack glottal stops (e.g., [mumən] vs. StA [muʔmin] ‘believer’) and allow initial CC clusters (e.g., [ktuba] vs. StA [kutub] ‘books’). Morphologically, plural formation differs: the ‘core’ item in (2c), for example, selects the broken plural form [žbuh] ‘foreheads’ while its cognate in the ‘peripheral’ sublexicon selects the sound plural form [žabah-aat] ‘frontlines’. Semantically, it is predicted that cognates diverge in meaning, proving that they are stored in separate sub-lexica. A predication borne out by the data (see (2)). Finally, we also expect a different syntactic behavior depending on core vs. peripheral lexical affiliation. In this regard, we examined the distribution of the two possible negative markers (particle *maši* and circumfixal *ma--š*) among adjectives in three corpora of MA and observed that the adjectives that do not obey l>ž-assimilation tends not to select the negative circumfixal form (e.g., /NEG + žunħi/ ‘unlawful’> [maši žunħi], *[ma žunħiš]), a typically MA construction. Under our view, these adjectives would belong to the peripheral sub-lexicon.

Importantly, because the borders of the core and peripheral sub-lexica in a diglossic grammar are fluid, we expect to find cases of free variation vis-à-vis l>ž-assimilation. Our data reveals that such cases abound in MA, and we expect that their occurrence is likely to be modulated by such factors as level of education/literacy, mastery and use of StA, register of speech, among others.

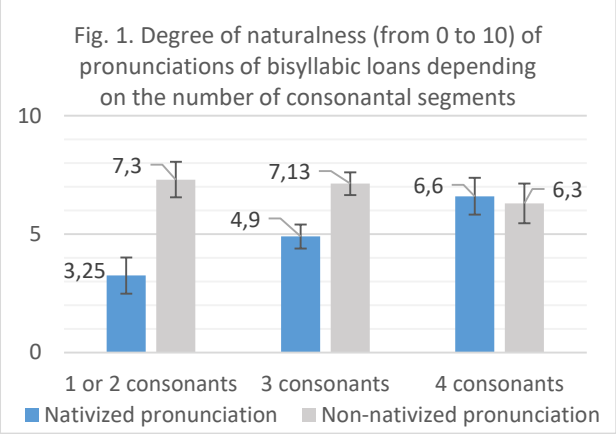
Loanword adaptation in Central Catalan: towards a characterisation of lexical strata

Xevi Pujol Molist, University of Barcelona

The core-periphery structure of a stratified lexicon with successive inclusion of layers has been shown to be a suitable model for representing the degree of adaptation of loanwords in various languages (Itô & Mester, 1999; Smith, 2018). In the case of Central Catalan, there are some insights showing this model's utility, especially with regard to the native vocalic processes applied to loans (Cabré, 2009; Pons-Moll et al., 2019; Pons-Moll & Torres-Tamarit, 2021).

The main claim of this work is that a crucial factor that **a)** decides the order of the lexical strata and **b)** identifies the layer to which a loanword belongs (i.e., the depth of penetration of a loanword into the native phonological system) is a concept that may be termed *identifiability*, referring to a loanword's capability not to be mistaken for any other lexical item. Three phonological factors which might hypothetically capture this notion have been analysed: **i.** preservation of phonological features of the most prominent segments; **ii.** preservation of phonological features of the leftmost segments of the word (first syllable); **iii.** preservation of phonological features in words with fewer segments.

According to this, the structures that would adapt phonologically in more superficial layers would be those related to the less prominent segments of the loanword and those further from the left margin, and adaptation would be greater in longer loanwords. Factors *i*, *ii* and *iii* were tested by means a survey answered by 10 participants aged 37 to 71. Participants assessed the degree of naturalness (on a Likert scale of 5 levels, spanning from 0 to 10) of different pronunciations of 30 loanwords, most of which are bisyllabic. Regarding vowel nativization of loans, two highly productive processes in Central Catalan were considered: vowel reduction of unstressed mid vowels and open pronunciations of stressed mid vowels. This gives rise to four possible patterns: taking the English loan *poster* as an example, [ˈpɔstər] (nativized), [ˈposter] (non-nativized), and [ˈpostər] and [ˈpɔster] (hybrid patterns).



Regarding vowel adaptation, results show that *iii* holds (Fig. 1), whereas *ii* holds only for stressed mid vowels, which display a higher tendency to preserve the closed feature when they appear at the beginning of the word. No significant contrast was found regarding unstressed mid vowels in terms of faithfulness (lack of vowel reduction) between pretonic or post-tonic positions in bisyllablic loans. This difference in performance between stressed and unstressed mid vowels could be due to factor *i*. On the other hand, the comparison between results of pairs of loans differing by only one sound, such as *latte vs. làtex*, *core vs. Corel* or *tote vs. tòtem*, reveals, in all cases, a higher degree of nativization for those pair-members with an extra consonantal segment (supporting *iii*). A morphophonological reason could also be adduced in these cases: a word-final reduced *e* could likely be confused with the morpheme of feminine gender /a/ in Central Catalan as both surface as [ə], diminishing the loan's *identifiability*.

Regarding consonantal adaptation, results show that the native process of final /n/ deletion is preferred to be applied to a longer loan, *orangutan*, than to shorter one, *divan* (which is consistent with *iii*). Finally, regarding the combination of vowel and consonantal adaptations, results are consistent with factor *i*, as vowels are more prominent than the foreign consonant [θ] in loans such as *cruce*, yielding the following layer ranking: [ˈkɾuθe] (outermost layer), [ˈkɾuze] (2nd layer) and [ˈkɾuzə] (nuclear layer); [ˈkɾuθə] is never produced.

REFERENCES

Cabré, Teresa (2009). Vowel reduction and vowel harmony in Eastern Catalan loanword phonology, in Vigário, Frota i Freitas (eds.), *Interactions in Phonetics and Phonology*, Amsterdam/Philadelphia: John Benjamins.

Itô, Junko, & Armin Mester (1999). The phonological lexicon, in TSUJIMURA (ed.), *A Handbook of Japanese Linguistics*, Oxford: Blackwell.

Pons-Moll, Clàudia, & Francesc Torres-Tamarit (2021). Catalan nativization patterns in the light of Weighted Scalar Constraints, in Sergio Baauw, Frank Drijkoningen, Luisa Meroni (eds.), *Romance Languages and Linguistic Theory. Selected papers from 'Going Romance' Utrecht 2018*, Amsterdam/Philadelphia: John Benjamins.

Pons-Moll, Clàudia, Francesc Torres-Tamarit & Vlad Martin-Diaconescu (2019). Catalan (im)possible nativizations in the light of Weighted Scalar Constraints, *27th Manchester Phonology Meeting*, Manchester: Manchester University.

Smith, Jennifer S. (2018). Stratified faithfulness in harmonic grammar and emergent core-periphery structure, in Bennett et al. (eds.), *Hana-bana: A festschrift for Junko Ito and Armin Mester*, Santa Cruz: UCSC Linguistics.