Anastassiadis-Syméonidis Anna
Université Aristote de Thessaloniki
Grèce

L'EXPRESSION DE LIEU ET LES ADJECTIFS SUFFIXÉS EN *-in(os)* ET *-isi(os)* EN GREC MODERNE

0. Introduction

Ce travail s'inscrit dans le cadre théorique élaboré à SILEX autour de Danielle Corbin et concerne deux suffixes du grec moderne (GM), les suffixes *-in(os)* (accentué à la finale) et *-isi(os)*, accentué à la pénultième, qui construisent des adjectifs dénominaux au moyen de la RCL[1] de relation. Dans la première partie, nous allons nous occuper du suffixe *-in(os)*, dans la deuxième partie, nous allons examiner le suffixe *-isi(os)* et, dans la troisième partie, nous procèderons à une comparaison du sens locatif tel qu'il est exprimé par ces adjectifs construits, afin de déceler des différences, s'il y en a, dans la conception du lieu[2] tel qu'il est véhiculé par ces adjectifs construits. Notre corpus est constitué des adjectifs en *-in(os)* (236) et en *-isi(os)* (196) de notre *Dictionnaire Inverse du grec moderne* comportant 180.000 entrées. Par ailleurs, ces adjectifs construits sont contextualisés dans des exemples extraits de la Toile et de magazines de grande diffusion.

1. Le suffixe *-in(os)*[3]

Dans cette partie, nous allons examiner (i) la catégorie de la base et du dérivé et (ii) le sens du nom de base (Nb) et l'instruction sémantique du suffixe *-in(os).*

(i) Le suffixe *-in(os)* s'applique sur une base nominale ou adverbiale, laquelle pourrait, pourtant, être considérée comme nominale, puisque ces adverbes fonctionnent aussi comme des noms (Berthonneau 1989 : 493) ; nous proposons, par conséquent, une base nominale unifiée. Quant à la catégorie grammaticale du suffixé, nominale si l'on suit les grammaires, nous considérons qu'elle résulte d'une conversion, qui convertit l'adjectif suffixé en nom, ex. *Patra*$_{N}$ 'nom de ville'

[1] Règle de Construction de Lexème. Il faut noter que D. Corbin parlait de Règle de Construction de Mot (RCM).
[2] Pour une étude approfondie du concept de lieu et des relations spatiales en français, v. Borillo 1998.
[3] Pour une description de l'instruction sémantique temporelle de ce suffixe v. Anastassiadis-Syméonidis 2008.

.

→*patrinos*$_{A}$ ‘de Patras’→*Patrinos*$_{N}$ ‘nom de l’habitant de cette ville’. Les adjectifs construits avec le suffixe *-in(os)* sont, par conséquent, le produit de la RCL$_{REL}$, qui instaure une relation entre le nom recteur et le Nb : [[X]$_{N}$ (-in(os))$_{suf}$]$_{A}$.

(ii) Les Nb de notre corpus sont, dans la grosse majorité, des noms de période temporelle ou de lieu, ex. *proinos* ‘matinal’, *vradinos* ‘du soir’, *kalokairinos* ‘d’été’, *pashalinos* ‘de Pâques’, *aprilianos* ‘d’avril’, *simerinos* ‘d’aujourd`hui’, *pantotinos* ‘de toujours’ — *voreinos* ‘du nord’, *antikrynos* ‘d’en face’, *brostinos* ‘de devant’, *makrinos* ‘lointain’. Pourtant, il reste un 13% de constructions où le Nb ne présente pas de sens temporel ni spatial et ne semble pas constituer un groupe homogène d’un point de vue sémantique. Dans ce groupe apparemment disparate, nous distinguons (a) un ensemble de deux noms d’aliment, *kreas* ‘viande’ et *tyri* ‘fromage’, (b) cinq noms de grands animaux domestiques *alogo* ‘cheval’, *agelada* ‘vache’, *vodi* ‘bœuf’, *gaïdouri* ‘âne’ et *hoiros* ‘porc’, et (c) lun nombre très restreint de noms comme *fos* ‘lumière’, *skotos* ‘ténèbres’, *alitheia* ‘vérité’, *eleos* ‘pitié’.

En ce qui concerne le groupe (a), la relation instaurée entre ces Nb et le nom recteur, nom de temps, pourrait se ramener à une relation temporelle, puisque *tyrini evdomada* est la semaine juste avant le Carême où il est permis de manger du fromage même mercredi et vendredi. Et la même chose pour *kreatini evdomada* ‘semaine où il est permis de manger de la viande même mercredi et vendredi’. En ce qui concerne le groupe (b), nous considérons qu’il s’agit d’une relation spatiale ; ces animaux, qui constituent, au niveau référentiel, le groupe homogène du gros bétail, sont vus comme le lieu d’origine d’où l’on tire/extrait quelque chose, comme la viande, le lait, ou une qualité par ex. la patience dans *gaïdourini ypomoni* lit. “patience d’âne” ‘très grande patience’. Enfin, pour les Nb du groupe (c)[4], nous considérons qu’il s’agit aussi d’une relation spatiale entre le Nb et le nom recteur, puisque la lumière et les ténèbres étaient considérées par les Anciens comme des lieux d’où émanaient la clarté et la vie et l’obscurité et la mort respectivement (Giannakis 2001). De même pour *alitheia*, qui était considérée chez Platon comme le trait qui émanait du monde vrai.

En observant les Nb de sens spatial, nous distinguons (i) un groupe de N référant à des termes géographiques, ex. *vorras* ‘nord, vent du nord’, *oros* ‘montagne’, *thalassa* ‘mer’, (ii) des toponymes, ex. *Alexandria* ‘Alexandrie’, et c) des N/ADV qui construisent des dénominations de la référentialité énonciative qui réanalysent l’unité du niveau référentiel ICI (déixis spatiale), par exemple *antikry* ‘en face’, *konta* ‘près’, *makria* ‘loin’, *plaï* ‘à côté’, *brosta* ‘devant’, *piso* ‘derrière’.

Pour conclure provisoirement, nous souhaitons mettre l’accent sur le fait qu’avec le suffixe *–in(os)* nous retrouvons une relation entre l’espace et le temps, déjà soulignée dans la bibliographie tant philosophique que linguistique mais pour

[4] A part le nom *eleos* ‘pitié’.

d'autres phénomènes, car ce suffixe sert à construire des adjectifs dénominaux qui localisent dans l'espace ou dans le temps.

2. Le suffixe *-isi(os)*[5]

Pour commencer, nous avons éliminé du corpus les adjectifs en *–isi(os)* qui sont accentués à l'antépénultième, puisqu'ils sont des dérivés construits à l'aide d'un suffixe différent, *-isi(os)*, écrit –ήσι(ος), ex. *etisios* 'annuel', qui est construit avec l'application de ce suffixe sur le Nb *et(os)* 'an'. Nous avons aussi éliminé les adjectifs terminés en *–isios* mais construits avec le suffixe *–i(os)* ex. *paradeis-ios* 'paradisiaque', *afrodis-ios* 'vénérien', et les emprunts (non construits), ex. *markisios* 'marquis'. Il nous reste 196 adjectifs construits avec ce suffixe.

En ce qui concerne son étymologie, j'adopte la proposition de Meyer (1895), selon laquelle ce suffixe provient du suffixe du bas latin *–ēsis* < latin classique *–ēnsis*. En latin, ce suffixe construit des adjectifs dénominaux, dont le Nb réfère à un lieu, ex. *castrensis* 'du camp', ou bien à un toponyme (nom de ville ou de région), ex.
aetnensis < *Aetna, antiochensis* <*Antiochia, atheniensis* < *Athenae, babyloniensis* < *Babylon, chalcidensis* < *Chalcis, chalcidicensis* < *Chalcidica, corinthiensis* < *Corinthus, cremonensis* < *Cremona, europensis* < *Europa, herculanensis* < *Herculaneum, ithacensis* < *Ithaca, laodicensis* < *Laodicea, mediolanensis* < *Mediolanum, nicaeensis* < *Nicaea, oscensis* < *Osca, palaestinensis* < *Palaestina, peloponnensis* < *Peloponnesus, pentelensis* < *Pentele, phocaeensis* < *Phocaea, pisaurensis* < *Pisaurum, rhodiensis* < *Rhodos, sinopensis* < *Sinope, tarsensis* < *Tarsus, viennensis* < *Vienna.* Ces adjectifs convertis en noms réfèrent à l'habitant.

Ce suffixe prend aussi la forme *–iensis*. La création de cet allomorphe est due au changement de frontières dans des cas où le radical du Nb se terminait en *–i,* ex. *eretriensis* < *Eretria, macedoniensis* < *Macedonia, romaniensis* < *Romania, sardiniensis* < *Sardinia, veliensis* < *Velia* 'région de la colline Palatinus à Rome', *volsiniensis* < *Volsinii* 'ville d'Etrurie'.

En grec ce suffixe construit des adjectifs dénominaux, produits de la RCL de Relation, qui peuvent par la suite se convertir en noms :

[[$[X]_O$ $(\text{-is(ios)})_{suf}$ $]_A$ CONV$]_N$ ex. $kampos_N$ 'plaine'→ $kamp\text{-}isios_A$ 'de la plaine'→ $kamp\text{-}isios_N$ 'habitant de la plaine'

Le Nb peut référer à un [concret]:

i) animal (66), ex. *aetisios* 'd'aigle', *alepoudisios* 'de renard', *gatisios* 'de chat', *gidisios* 'de chèvre', *gourounisios* 'de cochon', *korakisios* 'de corbeau', *mosharisios* 'de veau', *skylisios* 'de chien', *fidisios* 'de serpent',

[5] Pour plus de détails sur ce suffixe, écrit avec *i* (–ίσιος) v. Anastassiadis-Syméonidis 2009.

ii) lieu (41),[6] ex. *ikarisios* ‘d’Icarie’ (île grecque), *vounisios* ‘de montagne-montagnard’, *kampisios* ‘de plaine’, *limnisios* ‘de lac’, *pelagisios* ‘de large’, *potamisios* ‘de rivière’, *spitisios* ‘de maison’,

iii) artefact (26), ex. *varelisios* ‘de baril-tonneau’, *karavisios* ‘de bateau’, *pigadisios* ‘de puits’, *pitharisios* ‘de jarre’, *prymisios* ‘de poupe’, *sakoulisios* de sac’,

iv) êtres humains – parties du corps (27), ex. *agorisios* ‘de garçon’, *gerontisios* ‘de vieillard’, *gynaikisios* ‘de femme’ – *laryggisios* ‘de larynx’,

v) plantes (23), ex. *thymarisios* ‘de thym’, *kalampokisios* ‘de maïs’,

vi) temps (9), ex. *genarisios* ‘de janvier’, *paliokairisios* ‘retro’.

Au niveau sémantique, le grec conserve des traces qu’avait le suffixe en latin : il s’agit du sens locatif, et plus spécialement de provenance. En suivant le modèle associatif de Corbin pour une instruction sémantique unique au niveau abstrait, nous proposons que ce suffixe serve à construire des adjectifs qui expriment la provenance. Par ailleurs, il y a la relation de méronymie : le Nb fonctionne comme holonyme et le nom recteur comme méronyme (v. aussi pour le français Le Pesant 2001), ex. *limnisio psari* ‘poisson de lac’ : le lac a des poissons. Toutefois, quand le Nb réfère à un [animé] (animal ou être humain), l’adjectif construit peut avoir une lecture en extension ou en intension selon les stéréotypes de la société grecque, ex. *aetisia matia* ‘yeux d’aigle’, *alepoudisia poniria* ‘ruse de renard’, *gerakisia myti* ‘nez de faucon, aquilain’, *gourounisia symperifora* ‘comportement de cochon’, *fidisio kormi* ‘corps de serpent’ – *gerontisia foni* ‘voix de vieillard’, *gynaikisia kamomata* ‘minauderies de femme’. Ce suffixe n’est plus très disponible en grec.

3. Y a-t-il synomymie entre les deux suffixes ?

Quelle sorte de sens locatif est-il véhiculé par le Nb ? Les noms locatifs sont définis et classés par les experts, les savants, au moyen des conditions nécessaires et suffisantes. Par ailleurs, ces mêmes noms ont des définitions populaires résultant de catégories populaires, structurées autour d’éléments prototypiques et créées de la manière dont le peuple perçoit le monde environnant. Nous avons appris ces définitions populaires bien avant d’entrer en contact avec les définitions des experts, et elles continuent à persister, même après l’enseignement de ces dernières (Taylor 1989).
Le suffixe *-isi(os)* sélectionne parmi les propriétés référentielles du Nb auquel il s’applique des propriétés prototypiques ou stéréotypiques (Geeraerts 1985 : 30) et

[6] Il s’agit de morceaux de terrain de taille diverse, d’étendues d’eau de taille et de forme diverses, de lieux d’habitation.

construit un adjectif qui envisage ces propriétés d'un point de vue locatif, relevant de savoirs non scientifiques. Par contre, le suffixe *-in(os)* sert à construire des adjectifs dénominaux qui localisent dans le lieu et dans le temps de façon objective, c'est-à-dire sans perceptions prototypiques ou stéréotypiques, et c'est pour cette raison que ces adjectifs font partie de la terminologie scientifique. De même, les adjectifs en *-in(os)* font partie de la langue écrite de l'Eglise, qui utilise le registre savant.

Par conséquent, on peut prédire que, si un Nb a deux allomorphes, une forme [+savante/soutenue] et une forme [-savante/-soutenue], le suffixe *–in(os)* va s'appliquer sur la forme [+savante/soutenue] et le suffixe *–isi(os)* sur la forme [-savante/-soutenue]. Cette préférence pour les formes [-savante/-soutenue] concerne aussi d'autres suffixes grecs empruntés au latin, ex. *–at(os), -(i)ar(is), -oura, -iatik(os), -oukla, -oula, -poul(os).*

Références bibliographiques

Anastassiadis-Syméonidis, Anna. 1999. "Le suffixe *-in(os)* en grec moderne" in *Greek Linguistics '97-Proceedings of the 3rd International Conference on Greek Language*. Athènes : Ellinika Grammata : 315-323 (en grec).

Anastassiadis-Syméonidis, Anna. 2002. *Dictionnaire inverse du grec moderne*. Thessaloniki : Institut d`Etudes Néohelléniques.

Anastassiadis-Syméonidis, Anna. 2008. "Les adjectifs temporels suffixés en *–in(os)* et *–iatik(os)* en grec moderne" in Fradin Bernard (éd.) *La Raison morphologique – Hommage à la mémoire de Danielle Corbin*. Amsterdam : Benjamins : 17-27.

Anastassiadis-Syméonidis, Anna. 2009. «Le morphème *–isi(os)* en grec moderne», in *Studies in Greek Linguistics* 29, Thessaloniki : 58-73 [en grec].

Berthonneau, Anne-Marie. 1989. *Composantes linguistiques de la référence temporelle. Les compléments de temps, du lexique à l'énoncé*, Thèse d'Etat, Paris VII.

Borillo, Andrée. 1998. *L'espace et son expression en français*, Paris, Ophrys.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*, 2 vol. Tübingen, Max Niemeyer Verlag, (1991[2]) Villeneuve d'Ascq, Presses Universitaires de Lille.

Corbin, Danielle. 1991. "Introduction – La formation des mots : Structures et interprétations" in *Lexique* 10 : 7-30.

Geeraerts, Dirk. 1985. "Les données stéréotypiques, prototypiques et encyclopédiques dans les dictionnaires" in *Cahiers de Lexicologie* 46, 1 : 27-43.

Giannakis, Georgios. 2001. “Light is Life, Dark is Death: An Ancient Greek and Indo-european Metaphor”, in *Dodoni-Philologia* 30 : 127-153.
Kleiber, Georges. 1990. *La sémantique du prototype. Catégories et sens lexical*, Paris, P.U.F.
Le Pesant D. (2001) *Les noms locatifs*, HDR, Paris XIII.
Melis-Puchulu, Agnès. 1991. “Les adjectifs dénominaux : des adjectifs de relation” in *Lexique* 10 : 33-60.
Taylor, J.R. 1989. *Linguistic Categorization-Prototypes in Linguistic Theory*, New York, Oxford University Press.

Lovari loan-verb adaptation markers as arguments for an analogy-based analysis of verbal systems

András Márton Baló
PhD Student, Eötvös Loránd University, Doctoral School in Linguistics, Theoretical Linguistics Programme

It has often been a question in several morphological theories and frameworks whether certain morphemes are actually morphemes or they can be broken down into even smaller meaningful units. This latter solution may appear to be more appealing in many cases as the inventory of morphemes can thus be reduced.

Loan-verb adaptation with the help of certain markers is a very common phenomenon due to the ways and circumstances in which Lovari – and in a wider sense, Romany – is used and relies heavily on existing patterns and exemplars. Two of the most important elements of loan-verb adaptation in Lovari are the derivational markers *-sar* and *-sajv*: *zhutisar-* 'help' (from Romanian), *ašisar-* 'dig' (from Hungarian), *indulisar-* 'leave' (from Hungarian), *slobod-* > *slobodisajv-* 'be freed' (from Slavic or Romanian), *kezdődik* > *kezdődisajvel* 'begin (int.)' (from Hungarian). They can be viewed as such, that is, as one single unit, but it is possible to break them down into two parts, namely *-(V)s-* and *-ar/-ajv* on a historical basis – both possibilities can be justified. The question could be which is worth more, but apparently the changes that have happened and are happening in the language may make the strictly diachronic approach unnecessarily complicated, whereas, at the same time, it can be difficult to handle them within a traditional synchronic framework. However, if we look at factors of frequency and analogy-based processes which have taken place and are taking place as a result of those factors, the change can easily be made part of the model and the question loses its significance.

The classification of verbs itself is not without problems. Based on Matras 2002, we can say that, fundamentally, there are two separate groups formed in accordance with the final sound of the stem: the consonantal and the vocalic verbs. As for the former one, the third person singular personal concord marker in the present tense is connected to the stem by the linking vowel /e/, which is, as Baló 2008 points out, epenthetic, whereas in the case of the latter one the stem ends in the vowel /a/, therefore no linking vowels are necessary. this renders for example the third person singular present tense form *kinel* in case of the stem *kin-* 'buy, purchase' and *patjal* for the verb *patja-* 'believe'. Matras 2002 also points out that there is a third group, a mixed residual class, created through the disappearance or reduction of certain derivational markers on the one hand and through contraction on the other. Hungarian grammars (e.g. Hutterer–Mészáros 1967) split this into three further vocalic classes, the -i-, -o- and -u- stem verbs.

If we take a closer look at the -i- stem verbs in the table below, we can see that there are two variations – one with the marker *-sar* and one not containing it. Whereas all the -i- verbs are originally results of the optional reduction of the loan-verb adaptation suffix *-sar*, based on oral and written evidence it seems that the two variations now coexist within the Lovari variety spoken in Hungary, with the forms containing the marker *-sar* moving the verbs towards the consonantal class, which has got the highest type and token frequency, while the forms without it have created their own pattern. (There is also a possibility to adapt loan-verbs productively by simply placing them among the -i- verbs – *indulij* 'leave' from Hungarian *indul*, *sorakozil* 'have fun' from Hungarian *szórakozik* etc.).

A conspicuous anomaly emerges in connection with the -o- stem verbs, namely that they cannot be considered unanimously. The problem occurs in the first person forms, and the twofold nature of this verb class is traditionally explained by the different origins of the verbs

belonging to the two subgroups, the first one having been created through the partial contraction of the suffix *-uv*, the second one having been formed through the disappearance of the suffix *-sar*. This distinction, however, is obviously not recognised by native speakers. From an analogical perspective, the fact that the suffix *-uv* has remained in the first persons in the first subgroup may suggest that the first person forms have kept their original shape due to their frequency (and, on the other hand, the second and third person forms clearly bear resemblance to the conjugation of the -i- verbs).

Turning our attention to the loan-verb adaptation markers, Matras 2002 suggests that the suffixes *-sar* and *-sajv* should be broken down into a particle *-s-*, which may denote loan-verb adaptation and a transitive or an intransitive derivational suffix, respectively. On a historical basis, we can say then that the loan-verb adaptation markers in Vlax Romany look like the following: *-s-* + *-ar-* for transitive verbs and *-ajv-* for intransitive verbs. The particle *-s-* derives from Greek, the second element is of native origin and is still in use in internal verb formation. Based on the data about the verbal system presented above and additional information regarding internal verb formation and loan-verb adaptation confirmed by native informants, however, the Lovari spoken in Hungary shows a somewhat contradictory picture. Both *-sar* and *-sajv* take part in internal verb formation (e.g. *zuralo* 'strong' > *zuralosar-* 'strengthen', *phenel* 'say' > *phenosar-* 'promise', *kolo* 'soft' > *kolosajv-* 'become soft', *lolo* 'red' > *lolosajv-* 'turn red'), which means that the particle *-s-* does not denote solely loan-verb adaptation, and there is an additional suffix *-in*, probably borrowed from another Romany dialect, which is used to adapt loan-verbs but lacks the *-s-*.

These arguments support the fact that, in a synchronic aspect, they should not be broken down, which in turn provides a much more solid foundation for an analogy-based analysis of the verbal system, as the suffix *-sar* in particular has a crucial role in the derivation of consonantal verbs, which form the largest and most common verb class. In this aspect, the fact that it can be used both in internal verb formation and in loan-verb adaptation also loses its significance.

In relation to this, a parallel may be drawn between the Lovari suffix *-sar* and the Romanian suffix *-sc*, which is derived from the Latin inchoative affix. As Costanzo 2008 points out, it appears in a subclass of the fourth conjugation as a stem extension but earlier it was used in verbs borrowed from Balkan languages as well as in internal derivatives. There is also a variation as to whether the new verbs are formed with or without the *-sc*, similarly to Lovari, where, as mentioned above, new verbs can also be placed among the -i- stem verbs.

The analogy-based phenomena related to the suffix *-sar* in Lovari are also somewhat akin to the case of Hungarian linking vowels as touched upon in Kálmán (2007). In general terms we can say that there are certain, perhaps competing patterns to which the forms in the specific conjugational or declensional paradigms of words are adapted (in this case, the consonantal class and the newly formed -i- stem verbs). Thus, the point is not whether these suffixes should be considered as consisting of two parts or as individual morphemes; this is far beside the point. The forms and paradigms containing the suffixes *-sar* and *-sajv* can be considered as patterns which are both tools and bases of analogical changes in the Lovari verbal system.

present tense indicative	consonantal class	-a- stem verbs	-i- stem verbs
	kin- 'buy'	*loš-* 'be glad'	*traj-* 'live'
singular	kinav kines kinel	lošav lošas lošal	trajij/trajiv/trajisarav trajis/trajisares trajil/trajij/trajisarel

plural	kinas kinen kinen	lošas lošan lošan	trajinas/trajisaras trajin/trajisaren trajin/trajisaren
present tense indicative	-o- stem verbs		-u- stem verbs
	kerdjo- 'become'	*kino-* 'shake'	*sunu-* 'feel pity for'
singular	kerdjuvav kerdjos kerdjol	kinoj kinos kinoj/kinol	sunuj sunus sunul/sunuj
plural	kerdjuvas kerdjon kerdjon	kinonas kinon kinon	sununas/sunusaras sunun/sunusaren sunun/sunusaren

Bibliography

Baló, A. M. 2008. "A Strange Case of Defectiveness in the Lovari Verbal Paradigm". In: Kálmán, L. (ed.), *Papers from the Mókus Conference.* Budapest: Tinta.

Choli-Daróczi, J. and L. Feyér. 1988. *Cigány nyelvkönyv.* [Romany Textbook] Budapest: Magyarországi Cigányok Kulturális Szövetsége.

Costanzo, A. 2008. "On the expansion and productivity of a Balkan Romance conjugational sub-class". Poster presentation at the First Oxford Workshop on Romance Verb Morphology, 27-28/08/08.

Hutterer. M. and Gy. Mészáros. 1967. *A lovári cigány dialektus leíró nyelvtana.* [A descriptive grammar of the Lovari Gypsy dialect] Budapest: Magyar Nyelvtudományi Társaság.

Kálmán, L. "Holisztikus szemlélet a nyelvészetben" [A Holistic Approach to Linguistics]. *Szabad Változók* [Free Variables] 10th April 2010. <http://www.szv.hu/cikkek/holisztikus-szemlelet-a-nyelveszetben> 15th March 2007.

Matras, Yaron. 2002. Romani: A linguistic introduction. Cambridge: Cambridge University Press.

Complex Scales in Multiargument Agreement

Sebastian Bank & Jochen Trommer, University of Leipzig

In languages with rich inflectional morphology, verbal agreement often depends on the relative markedness of the subject and the object (*hierarchical agreement*, *direct/inverse marking*). Virtually all existing formal approaches to hierarchy effects like this (e.g. Aissen 1999, Béjar 2003) relate them to mismatches on atomic scales that rank feature values for specific categories like the traditional person hierarchy 1 > 2 > 3. In this talk, we show that the agreement pattern of the Hokan language Karuk (also Karok, cf. Bright 1957, Macaulay 1992) demands an analysis that *firstly* makes reference to complex scales wich conflate different categories (e.g. 1 > 2pl > 2sg > 3 for person and number) and *secondly* measures the strength of mismatches on such scales as two-step-effects. We develop a principle-based theory for the construction of such (language specific) complex scales from (universal) base scales. Nevins & Sandalo (to appear) abolish the usage of such complex scales and show how they can be avoided. Examining a quite similar agreement pattern in the Mataco-Guaicuru language Kadiwéu, they conclude the explanation needs to be shifted to the properties of individual exponents, and derive it from their morphotactics interacting with general markedness statements. However, as we show, our approach is empirically superior: The Kadiwéu pattern can straightforwardly be described in terms of the complex scales, but the Karuk data can't be captured analogously to their analysis, as the relevant morphotactic pecularities it relies on are not present in Karuk.

Background: Languages with the potential to agree with either argument of transitive sentences often choose whether to agree with the subject or the object on the basis of their relative prominence (hierarchical agreement): The verb agrees with the argument most prominent on a hierarchy, like one of the more common scales in (1). Hence, a language employing (1a) marks first person as soon as there is a first person argument and third person only, in case both subject and object are third person.

(1) a. 1 > 2 > 3 b. animate > inanimate c. pl > sg[1]

Additionally, those languages often exhibit direct/inverse marking: An inverse marker occurs if the subject is lower on a hierarchy than the object, while the opposite (direct) configuration typically remains unmarked (Comrie 1980). If the inverse marking is governed by a hierarchy similar to that of the other agreement markers, this may serve to at least partially reduce the ambiguity introduced by hierarchical agreement – both inflectional patterns are thus functionally related and therefore frequently occur together.

Data: In Karuk transitive agreement the person/number prefixes agree with the object by default, turning to subject-agreement if a) the subject is first or second and the object is third person (1→3, 2→3); b) the subject is first and the object is second person singular (1→2sg); c) the object is third person singular and the subject is anything different (3pl→3sg). For local (first and second) vs. non-local (third) person(s), this means there is always agreement with the local argument – independently of its grammatical function:

(2) a. (neg)

1→3	3sg	3pl
1sg	*ná-*	*ná-*
1pl	*kín-*	*kín-*

3→1	1sg	1pl
3sg	*ná-*	*kín-*
3pl	*ka-ná-*	*kín-*

[1] Abbreviations and conventions used for examples: 1/2/3 = first/second/third person, sg/pl = singular/plural, A/P = transitive subject/object, pos/neg = positive/negative order, $\alpha \rightarrow \beta$ = transitive subject α object β.

b.

2→3	3sg	3pl
2sg	*ʔi-*	*ʔi-*
2pl	*ku-*	*ku-*

3→2	2sg	2pl
3sg	*ʔi-*	*ki·k-*
3pl	*ʔi-*	*ki·k-*

(pos)

With *na-* and *kín* as first person markers (singular and generic) and *ʔi-* and *ku/ki·k-* expressing second person singular and plural, (2) thus forms a perfect instance of hierarchical agreement governed by the scale {1, 2} > 3 (Siewierska 1996). However, the agreement pattern in contexts with local persons only ('you and me' forms) is more complex:

(3)

1→2	2sg	2pl
1sg	*kín-*	*ki·k-*
1pl	*kín-*	*ki·k-*

2→1	1sg	1pl
2sg	*ná-*	*kín-*
2pl	*ka-ná-*	*kín-*

(neg)

In transitives with a first person and a second person `singular` argument (4a-d), there is always (hierarchical) agreement with the first person argument, while in contexts with a first person and a second person `plural` argument (4e-h) we are faced with (default) object agreement:

(4) *Agreement pattern for 1→2 and 2→1 forms (agreed-with argument in bold)*

a. **1sg**→2sg
b. 2sg→**1sg**
c. **1pl**→2sg
d. 2sg→**1pl**
e. 1sg→**2pl**
f. 2pl→**1sg**
g. 1pl→**2pl**
h. 2pl→**1pl**

Analysis: By comparison of (4a) and (4g), it becomes obvious, that any analysis utilizing simple (person, number, grammatical function) scales could not avoid to falsely predict agreement with the same argument for these two contexts – which then again is the case in the other mininal pair (4b) compared to (4h). In fact, Béjar (2003) admits that her analysis does not extend to these cases. To account for the differential behavoiur of second person in a hierarchical approach, the person scale (5a) needs to be subdifferentiated through the number scale (5b).

(5) a. {1, 2} > 3 b. pl > sg c. A(gent) > P(atient)

Departing from approaches where particular instantiations of universal hierarchies are captured by the language-specific parametrization of feature structure (Béjar & Rezac 2009), or the construction of optimality-theoretic constraints (Aissen 1999), we thus shift the burden of explanation to the licensing of (language-specific) scales by (universal) simplex scales. Licensing of such scales follows the Scale Composition Principle:

(6) **Scale Composition**
A complex scale $CS = C_m, C_{m-1}, \ldots, C_1$ is licensed
by the ranking of scales $SS = S_n, S_{n-1}, \ldots, S_1$ **iff:**
for every pair of categories $C_i, C_j, i > j$:
If $C_j > C_i$ for scale S_p
then $C_i \geq C_j$ for scale S_o, $o > p$

This operation only allows for sub-differentiation of atomic points on a higher ranked scale through the categories of a lower scale. It is thus restricted to only yield possibilities for more fine-grained complex scales that inherit all the restrictions of the base scales.

(7) PERSON ≻ NUMBER ≻ GRAMMATICAL FUNCTION

Crucially, *Scale Composition* allows to derive the complex scale (8) from the simplex scales in (5) under the ranking in (7).

(8) 1 > 2pl > 2sg > 3plA > 3plP > 3sgA > 3sgP

The hierarchy effect in Karuk now only occurs, if two arguments differ by at least two positions on the scale in (8). Thus the distance between first person and second person plural would not trigger a hierarchy effect, but the distance of first person and second person singular would. The distribution of the prefixes is then fully captured by the requirement that they switch from default object agreement to hierachical agreement if one argument outranks the other argument by at least two steps on (8).

Further evidence for an analysis that measures hierarchy mismatches in form of steps over positions for complex scales comes from the distribution of the suffix *-ap*: Macaulay (1992) identifies it as an inverse marker but has to admit that it has a defective distribution as it doesn't occur in typical inverse cases like 3→1. Rather it only occurs in cases with a second person object but again not in all cases and the distribution again shows an interaction of person and number hierarchies:

(9) -ap *in the X→2 forms of the positive and optative order (occurrences in bold)*

a.	1sg→2sg	1pl→2sg	b.	**1sg→2pl** **1pl→2pl**
c.	**3sg→2sg**	**3pl→2sg**	d.	**3sg→2pl** **3pl→2pl**

In our analysis, this distribution follows from the complex scale in (10) and another two-step-effect. Since first and second person are not strictly ordered under the person scale, *Scale Composition* licences to construct (10) from (5) and (7).

(10) 2pl > 2sg > 1 > 3

The inverse marker *-ap* exactly occurs, if the subject is not only lower, but also two steps lower on this scale than the object. Hence Karuk only displays inverse marking, if the strength of the inverse relation between subject and object is big enough.

Discussion: Nevins & Sandalo (to appear) argue against the employment of a complex hierarchy drawing on a quite similar pattern in Kadiwéu. Instead they attribute the observed differential behavior of first person singular vs. plural and second person subject vs. object to the properties of individual markers and restrictions on their combinability.

In Kadiwéu transitive agreement (11a), the subject markers (unshaded cells) are the same as the markers for the subject of unergatives (11b-i), and the object markers (lightly shaded cells) are the same as the markers for the subject of unaccusatives (11b-ii). While person is overtly expressed, there is a strong tendency to neutralize number distinctions. In the second person this has been fully applied, as the second person is always plural.

(11) a.

A→P	1sg	1pl	2sg	2pl	3sg	3pl
1sg	–	–	*Ga- -i*	*Ga- -i*	*j-*	*j-*
1pl	–	–	*Ga- -i*	*Ga- -i*	*j- -aGa*	*j- -aGa*
2sg	*a- -i*	*Go-*	–	–	*a- -i*	*a- -i*
2pl	*a- -i*	*Go-*	–	–	*a- -i*	*a- -i*
3sg	*i-*	*Go-*	*Ga- -i*	*Ga- -i*	*y-*	*y-*
3pl	*i-*	*Go-*	*Ga- -i*	*Ga- -i*	*o-y-*	*o-y-*

b. (i)

	sg	**pl**
1	*j-*	*j- -aGa*
2	*a- -i*	*a- -i*
3	*y-*	*n- -aGa*

(ii)

	sg	**pl**
1	*i-*	*Go-*
2	*Ga- -i*	*Ga- -i*
3	∅	*n- -aGa*

Looking again at the transitive combinations of local with non-local person(s), (11a) is a perfect instance of hierarchical agreement governed by the person scale {1, 2} > 3: There is always agreement with the local person. In non-local person only configuration (3→3), there is subject agreement which arises from A > P. The pattern for two local persons ('you and me' forms) however again deviates from this simple picture:

(12) *Agreement pattern for 1→2 and 2→1 forms (agreed-with argument in bold)*
a. 1sg→**2**(pl) b. **2**(pl)→1sg c. 1pl→**2**(pl) d. 2(pl)→**1pl**

In transitives with a second person and a first person singular argument, there is always agreement with the second person argument (12a,b) – as would be predicted by a 2 > 1 > 3 hierarchy. However, in contexts with a second person and a first person plural argument, there is object agreement.

In Nevins & Sandalo's analysis, it is the insertion of the first person plural object marker with second person subjects that represents the 'unexpected' case in terms of their general assumptions about the markedness conditions in Kadiwéu. In this case the competing second person subject marker is blocked, because another general local person object marker has already been inserted by a different component of the agreement system. Because this marker can't be deleted, their general ban on two-argument agreement then causes a ban on marking the other (subject) argument for this configuration. However, such an analysis does not work for Karuk, as there is so such marker that would block the expected marker for the cases, which are deviant in terms of a general markedness hierarchy.

On the other hand the Kadiwéu pattern can be straightforwardly be captured in our approach by constructing the complex scale (13) from the established base scales.

(13) {1pl, 2(pl)} > {1sg, 2sg} > 3A > 3P

As there is no second person singular in Kadiwéu, the corresponding point on this scale has to be considered inactive. The differential behavior of first person singular vs. plural now is a regular hierarchy effect arising from mismatches on (13): The verb agrees with the higher argument on this scale. In case no hierarchy mismatch can be established (1pl→2 and 2→1pl), the system falls back to the default object agreement of an ergative language.

References

Aissen, J. (1999). Markedness and subject choice in Optimality Theory. *Natural Language and Linguistic Theory*, 17(4), 673–711.

Béjar, S. (2003). *Phi-Syntax: A Theory of Agreement*. PhD thesis, University of Toronto.

Béjar, S. & Rezac, M. (2009). Cyclic agree. *Linguistic Inquiry*, 40(1), 35–73.

Bright, W. (1957). *The Karok Language*. Berkeley: University of California Press.

Comrie, B. (1980). Inverse verb forms in Siberia. *Folia Linguistica Historica* 1, 61–74.

Macaulay, M. (1992). Inverse Marking in Karuk: the Function of the Suffix *-ap*. *International Journal of American Linguistics* 58, 182–201.

Nevins, A. & Sandalo, F. (to appear). Markedness and Morphotactics in Kadiwéu [+participant] Agreement. *Morphology*.

Siewierska, A. (1996). Word order type and alignment type. *Sprachtypologie und Universalienforschung* 49, 149–176.

Causative light verbs in Mandarin Chinese

Bianca Basciano
Università degli Studi di Verona

This talk will deal with the issue of Mandarin Chinese complex V-V verbs formed by means of a causative light V_1, also in a comparative perspective, taking into account other Sinitic languages. Different languages may express causativity in different ways, according to the items available in their respective lexical inventories (cf. Ramchand 2008) and other language-specific characteristics. Mandarin Chinese has few lexical causatives (labile verbs), e.g. 开 *kāi* 'open' and 沉 *chén* 'sink'; however, even when labile verbs are available, a compound form is generally preferred to express the transitive variant. The main means to express causativity in Mandarin Chinese are periphrastics means and complex verbs formed by two verbal roots (i.e. resultative compounds and verbs formed with a light V_1). The tendency to express causativity by means of compounding seems to be linked to the analytic nature of Mandarin. This is even clearer if we consider the diachronic development of the Chinese language, which is characterized by a typological shift from a synthetic to an analytic language and by a substantial change in the lexicon. These factors apparently contributed also to the change in the ways to express causativity by means of different strategies; complex causative verbs seem to emerge out of the need to compensate for the loss of other means to express complex event structures. Old Chinese possessed morphological and lexical causatives; Middle Chinese developed means such as tonal contrast and the voiced/voiceless alternation (清浊别义 *qīngzhuó bié yì*; cf. Mei 1991, Pulleyblank 2000, among others). By the time of Late Middle Chinese, all these means were extinct; Chinese developed other means, such as the resultative construction and, then, resultative compounds, where both the causing and the result events are expressed, as e.g. 摇醒 *yáoxǐng* 'shake-awake'.

In this talk, we will not deal with resultative compounds, but, rather, we will focus only on complex V-V verbs formed with a phonetically realized light verb (带音的轻动词 *dài yīn de qīng dòngcí*), i.e. a verb that has general and abstract semantic content (see Grimshaw & Mester 1988, Feng 2005, Zhu 2005, Jie 2008), as e.g. 打 *dǎ* 'beat, strike, hit', 弄 *nòng* 'make', 搞 *gǎo* 'do'. These verbal roots, when appearing as V_1s of V-V compounds often do not represent a particular action, origin or manner, as in the case of resultatives, but are blurred verbs, with a general causative meaning, forming the transitive version of intransitive change-of-state verbs. Some complex verbs formed with a causative light verb are shown in (1):

(1)

弄暗 *nòng'àn* 'darken'	弄开 *nòngkāi* 'open'
弄沉 *nòngchén* 'sink'	搞破 *gǎopò* 'break'
弄断 *nòngduàn* 'break'	搞坏 *gǎohuài* 'ruin; destroy; break'

Light verbs do not give any semantic contribution to the whole complex verb. The compound verb just expresses the resultant state, leaving the causing event unspecified: different actions can bring about the resultant state. Moreover, these complex verbs can choose different kinds of subjects, showing that they can come about without the intervention of a volitional agent (see alternating verbs in English, e.g. *break*, which allow as external causes natural forces, as well as agents or instruments; cf. Levin & Rappaport Hovav 1995):

(2)

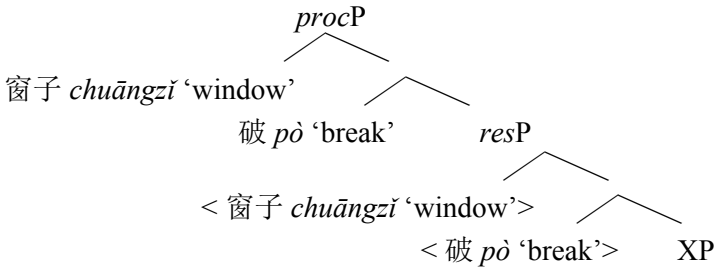
老张	/木头 / 台风	弄破	窗子。
Lǎo Zhāng	*mùtou*	*táifēng*	*nòngpò*	*chuāngzǐ*
Lao Zhang	wood	typhoon	*nòng*-break	window

'Lao Zhang / the wood / the typhoon broke the window.'
(Examples from Lin 2001:49)

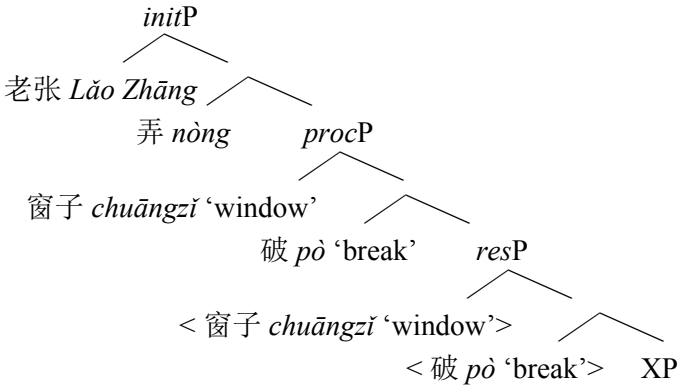
We will propose an analysis of these complex verbs adopting the framework put forth by Ramchand (2008), consisting in a syntactic decomposition of the event structure, which can be

decomposed into a maximum of three subevents: the causing (*init*P), the process (*proc*P) and the result (*res*P) subevents. Each lexical item specifies the relevant information (category labels or 'tags') which permit its insertion in the eventive structure: e.g. Eng. *push* [init, proc], *throw* [init, proc, res]. Accordingly, we argue that the light V_1 represents the spell-out of the causing projection head, forming an extra layer on top of verbs which do not possess an [init] feature in their lexical entry.

(3) a. 窗子 破 了。 破 *pò* 'break' [proc, res]
chuāngzǐ *pò* *le*
window break ASP
'The window broke.'



b. 老张 弄破 窗子。
Lǎo Zhāng *nòngpò* *chuāngzǐ*
Lao Zhang *nòng*-break window
'Lao Zhang broke the window.'



Mandarin Chinese supports the view that in the causative alternation the direction of the derivation is from inchoative to causative (see e.g. Ramchand 2008) and not vice versa (see e.g. Levin & Rappaport Hovav 1995, Reinhart 2002, Chierchia 2004). Moreover, we will argue that these complex verbs (as well as resultative compounds) are left-headed on a structural basis.
Among light verbs, we will illustrate one particular case, i.e. the root 打 *dǎ* 'hit, beat, strike'. In V-V compounds 打 *dǎ* can be used either as a full verb, forming a resultative compound, or as a light verb. In a complex verb like 打死 *dǎsǐ* '*dǎ*-die', the meaning of 打 *dǎ* could be either 'beat and kill (make die) as a result' or simply 'kill (make die)'; in the latter case, the resultant state 'die' can be reached performing different actions:

(4) 小明 开枪 打死 了 一 只 (From PKU corpus)
Xiǎo Míng *kāiqiāng* *dǎsǐ* *le* *yī* *zhī*
Xiao Ming shoot *dǎ*-die ASP one CL
'Xiao Ming killed one (bird) by shooting'

Similar roots are found in other two Sinitic languages, i.e. Taiwanese Southern Min (TSM) 拍 *phah4* 'hit' and Hakka 打 *da2* 'hit'. In TSM, a simplex intransitive verb may become transitive when preceded by the dummy verb 拍 *phah4*: 拍 *phah4* serves to change the argument structure of the simplex verb (cf. Lien 1999).

(5) 拍勼 *phah4 kiu1* 'shrink' 拍否 *phah4 phai2* 'spoil'
拍熄 *phah4 sit4* 'extinguish' 拍醒 *phah4 chhin2* 'wake'
拍破 *phah4 phoa3* 'break' 拍斷 *phah4 tng7* 'break' (Lien 1999:8)

In Hakka (Yeh 2008) the verb 打 *da2* apparently has the same function:

(6) 打壞 *da2 fai3* 'break' 打毋見 *da2 m5gien3* 'to make something disappear'
打缺 *da2 kiet4* 'chip' 打衊 *da2 met8* 'to make dirty' (Yeh 2008:67-68)

Finally, we will illustrate another particular light verb, i.e. the root 加 *jiā* 'increase'; our proposal is that its function is to form the transitive variant of change-of-state verbs based on a particular set of open-range adjectives. In Mandarin Chinese, stage-level adjectives (cf. Carlson 1977) can be used as eventive predicates, while individual-level adjectives can only occur in stative predication (cf. Gu 1992, Xiao & McEnery 2004, Liu 2010):

(7) a. 天 黑 / 亮 了。
tiān hēi / liàng le
sky black bright ASP
'It got dark / downed.'
b. 碗盘 刚 干，你 又 要 用 了。
wǎnpán gāng gàn nǐ yòu yào yòng le
dishes just dry you again want use ASP
'The dishes have just dried and you want to use them again.'
(Adapted from Tham 2009:5)
c. *张三 聪明 / 笨 了。
Zhāngsān cōngming / bèn le
Zhangsan clever stupid ASP
'Zhangsan got clever / stupid.'

We argue that these items are endowed with verbal features and such features are listed in their lexical entry along with the adjectival ones. We will also show that these items behave as degree achievement verbs and are ambiguous between being [proc] and [proc, res] verbs (having both an atelic and a telic punctual reading). Most of these change-of-state verbs can only be used intransitively; their transitive variant is formed by adding a light verb. We will show that, among these items, those based on open-range adjectives involving an increase in some property can be transitivized by means of the root 加 *jiā* 'increase':

(8) 加宽 *jiākuān* 'increase + wide = widen' 加深 *jiāshēn* 'increase + deep = deepen'
加长 *jiācháng* 'increase + long = lengthen' 加高 *jiāgāo* 'increase + high = heighten'

We will highlight the differences between these verbs and resultative compounds. Moreover, we will point out that the root 加 *jiā* 'increase' cannot be added to verbs of change of state based on closed-range adjectives. Also, we will show that to form the transitive variant of change-of-state verbs based on open-range adjectives denoting a decrease in some properties (increase in negative properties), a V_1 that marks the negative direction of the change in degree is required (cf. Steffen Chung 2006). We will propose that the root 加 *jiā* 'increase' 1) represents the causative component, forming the transitive variant of verbs of change of state based on open-

range adjectives involving an increase in the property denoted by the adjective; 2) is the spell-out of one relevant part of the logical representation, i.e. the increasing event (cf. Hay, Kennedy & Levin 1999).

Our analysis and findings will be based on a variety of data coming from different sources: the literature on the topic, dictionaries, contemporary literary texts and newspapers (mostly available on-line), corpora of Mandarin Chinese (above all the Corpus of the Center for Chinese Linguistics at Peking University – PKU corpus), Google searches.

Selected references

BASCIANO, B. (2010). Verbal compounding and causativity in Mandarin Chinese. PhD dissertation, Università degli Studi di Verona.

CHIERCHIA, G. (2004). A semantics for unaccusatives and its syntactic consequences. In A. Alexiadou et alii. (eds.), *The Unaccusative Puzzle: Explorations of the Syntax-Lexical Interface.* New York: Oxford University Press, 22-59.

FENG, S.L. (2005). 轻动词移位与古今汉语的动宾关系 'Light verb movement and the verb-object relationship in Old and Modern Chinese'. 语言科学 *Yuyan Kexue* 4 (1):3-16.

GRIMSHAW, J. & MESTER, A. (1988). Light verbs and theta-marking. *Linguistic Inquiry* 19:205-232.

GU, Y. (1992). *The syntax of resultative and causative compounds in Chinese*. Ph.D. dissertation. Cornell University.

HAY, J., KENNEDY, C., LEVIN, B. (1999). Scalar structure underlies telicity in "degree achievements". In T. Mathewsand & D. Strolovitch (eds.), *SALT IX*, CLC Publications: Ithaca, 127–144.

JIE, Z.M. (2008). 轻动词假说与词典释义 'The light verb hypothesis and dictionary definition'. 辞书研究 *Cishu Yanjiu* 4:30-44.

LEVIN, B. & RAPPAPORT HOVAV, M. (1995). *3*. Cambridge, MA: MIT Press.

LIEN, C.F. (1999). A Typological Study of Causatives in Taiwan Southern Min. *The Tsing Hua Journal of Chinese Studies,* 29 (4): 395-422.

LIN, T.H. (2001). *Light Verb Syntax and the Theory of Phrase Structure*. PhD dissertation, University of California, Irvine.

LIU, C.S.L. (2010). The positive morpheme in Chinese and the adjectival structure. *Lingua* 120:1010-1056.

MEI, T.L. (1991). 从汉代 "动、杀"、"动、死" 来看动补结构的发展——兼论中古时期起词的施受关系的中立化 'The development of verb-complement construction looking at 'V-die', 'V-kill' in the Han dinasty: a discussion of the neutralization of the agent–patient relationship in verbs since the Middle Chinese period. In 语言学论丛 *Yuyuanxue luncong* 16: 112-136.

PULLEYBLANK, E. (2000). From Archaic Chinese to Mandarin. In G. Booij et alii (eds.) G. *Morphology: A Handbook of Inflection and Word Formation.* Berlin: Mouton de Gruyter, 1730-1740.

RAMCHAND, G. C. (2008). *Verb Meaning and the Lexicon: A First-Phase Syntax*. Cambridge: Cambridge University Press.

REINHART, T. (2002). The Theta System- An Overview. *Theoretical Linguistics* 28: 229-290.

STEFFEN CHUNG, K. (2006). *Mandarin compound verbs*. Taiwan Journal of Linguistics, Book Series in Chinese Linguistics. Taipei: Crane.

THAM, X.W. (2009). Building resultatives in Mandarin (from the result). Handout, Tenth Stanford Semantic Fest, Stanford University, March 2009.

XIAO, R. & MCENERY, T. (2004). *Aspect in Mandarin Chinese: A Corpus-based Study.* Amsterdam/Philadelphia: John Benjamins.

YEH, C.S. (2008). *A Family of da2 'hit' Constructions in Hakka: An Inspection of Argument Realization and Transitivity*. MA dissertation, Taipei, National Chengchi University.
ZHU, X.F. (2005). 轻动词和汉语不及物动词带宾语现象'Light verbs and the phenomenon of Chinese intransitive verbs bearing an object'. 现代外语 *Xiandai Waiyu* 28 (3): 221-231.

Stéphanie BELIGON
Paris-Sorbonne (Paris IV)

Un- en anglais : un ou des préfixes ?

Le préfixe négatif *un-* forme aujourd'hui, en anglais, de très nombreux adjectifs, ainsi que des verbes et des noms, et se caractérise par sa grande productivité. Cette omniprésence soulève la question suivante : s'agit-il dans tous ces cas d'un même préfixe ou de préfixes différents et homonymes ? Les avis divergent mais l'origine du préfixe plaide apparemment pour une distinction entre, d'une part, un préfixe verbal et, d'autre part, un préfixe adjectival et nominal. Nous étudierons ici plus particulièrement le cas des adjectifs et des verbes.

1. Des origines différentes.

H. Marchand rappelle que le préfixe *un-* présent dans *unfair*, par exemple, a « the basic meaning "not". It goes back to the same Indo-European root as OGr *a-*, *an-*, L *in-*, G *un-* »[1], alors que l'affixe que l'on trouve dans les verbes du type *unbind* « is generally said to go back to OE *and-*, *ond-* which corresponds to G *ent*, ultimately identical with OGr *anti* and Latin *ante*, orig. "opposite" »[2]. Historiquement, il s'agit donc de morphème distincts, dont les sens divergent nettement : *unfair* est assez proche, sémantiquement, de « not fair », alors que *unbind* ne signifie pas « not bind », mais plutôt « do the opposite of *bind* » : des points de vue historique et sémantique, il ne paraît pas infondé de voir en *un-* deux morphèmes homonymes.

Cette première dichotomie, cependant, est peut-être simpliste : au sein même de chaque partie du discours émergent des nuances supplémentaires. Pour ce qui est des adjectifs, on distingue contraires[3] (*happy*/*unhappy*) et complémentaires[4] (*fair*/*unfair*, *true/untrue*, etc.)

Les verbes forment un champ encore plus contrasté : on différencie les valeurs réversative, privative et ablative. Les verbes réversatifs renvoient à un procès dont le résultat est opposé à celui de leur base : on peut citer *unzip* (a *coat*), *unroll* (*a sleeping bag*) ou encore *uncoil* (*a rope*). Les verbes privatifs correspondent à un procès consistant à séparer un élément ou composant de l'entité désignée par l'objet du verbe (ex. *uncork a bottle*). Enfin, les ablatifs renvoient à un procès consistant à enlever l'entité désignée par l'objet d'un lieu ou d'un contenant (*uncage a bird*).

Dès lors, pourquoi ne pas distinguer autant de préfixes que de significations assumées par *un-* ? Se différencieraient alors une multitude d'affixes, or cette solution nous paraît peu satisfaisante : ce souci de précision aboutit à un tableau qui manque de

---

[1] Marchand, H., *The Categories and types of present-day English word formation, a synchronic-diachronic approach*, Wiesbaden : O. Harrassowitz, 1960, p.150.
[2] *ibid*., p. 153.
[3] Nous adoptions ici les définitions de D.A. Cruse, selon lequel les contraires « do not strictly bisect a domain: there is a range of values of the variable property, lying between those covered by the opposed terms »: *cold*/*hot*, *slow*/*quick* et *cheap*/*expensive* sont des couples contraires. (Cruse : 1986, p. 198)
[4] Au sujet des complémentaires, D.A. Cruse explique « There is no "no-man's-land" […] no possibility of a third term lying between them » L'auteur précise que « if we deny that one term applies to some situation, we effectively commit ourselves to the applicability of the other term » et donne pour exemples les couples *true*/*false* ou encore *dead*/*alive*. (Cruse : 1986, p.198)

cohérence. En outre, la forme est signifiante et notre parti-pris sera que des préfixes formellement identiques sont l'indice d'opérations cognitives similaires.

2. Une communauté sémantique.

D'ailleurs, comme le rappelle Marchand, que le préfixe verbal, initialement différent formellement, prenne la forme de *un-* dès l'époque du vieil anglais n'est pas le fruit du hasard ; une forme telle que *unlocked*, par exemple, peut correspondre à la fois au participe passé du verbe *unlock* ou à l'adjectif *unlocked*, que l'on peut paraphraser par la négation du participe *locked*. Or, selon l'argumentation de Marchand, l'état désigné est identique, la seule nuance résidant dans l'idée d'un procès préalable menant à cet état dans le cas du participe de *unlock*.[5] Il serait donc possible de voir une corrélation entre ces deux *un-*.

N. Maynor ou E. Andrews considèrent d'ailleurs que les divers emplois de *un-* sont l'expression d'un même morphème :

> The fact that the *un* of *unlikely* seems to differ in meaning from that of *undress* is a consequence of the different environments – adjectival and verbal. Since the meanings do not contrast, the forms should not be considered several morphemes. Even if the forms derive from different sources, in today's English these *un*-s serve the same purpose – the indicating of opposition.[6]

Le « sens » de *un-* serait donc identique dans le cas des verbes et des adjectifs : si des nuances se dégagent, elles tiennent à la base à laquelle s'affixe *un-*, qui impose au préfixe le cadre sémantique dans lequel elle-même prend place.

3. Un cadre trop flou ?

Ces explications sont certes pertinentes mais, comme le déplore L. Horn, « without a semantics of oppositeness which generalizes across verbs, adjectives, and nouns, their [Andrews and Maynor's] one *un-* position represents more a hope or promise than an analysis. »[7] Les conclusions de Maynor et Andrews manquent en effet parfois de rigueur : le sens qu'elles attribuent à *un-* est très lâche et n'est pas nécessairement très convaincant pour les adjectifs – l'idée d'annulation alléguée par Andrews n'est pas toujours incontestable avec des lexèmes comme *unfair* ou *unhappy*, par exemple. D'autre part, elles ne rendent pas compte des limites à la productivité de *un-*, quelle que soit la partie du discours considérée. Pour féconde qui puisse être la thèse unificatrice, elle risque de faire du morphème une vaste poche de sens et de manquer de précision : de telles descriptions ne font pas état d'un bon nombre de lexèmes dans lesquels *un-* n'apparaît pas. Pour véritablement cerner le fonctionnement de ce préfixe, il est nécessaire de déterminer quelles sont ses conditions d'apparition et d'essayer de comprendre pourquoi il est incompatible des classes entières de lexèmes.

4. Conditions d'apparition du préfixe *un-*.

---

[5] Selon Marchand, l'assimilation formelle de deux prefixes initalement formellement différents « certainly does not mean a mere spelling variant. Possibly starting from second ptc forms, the prf *on-* had come to be felt connected with the negative prf *un-*. The idea of negativity is common to both […]. What distinguishes *unbound* "not bound" from *unbound* "loosened" is only the additional idea of an action preceding the state of being loosened, but the state itself is the same. » (Marchand : 1960, p.153. « Prf » signifie « prefix », « ptc » « participle » et « OE » « Old English ».)

[6] Maynor, N., « The Morpheme *Un-* », *American Speech*, Vol. 54, No. 4 (Winter, 1979), p. 311.

[7] Horn, « Morphology, Pragmatics and the *Un*-verb », ESCOL, 1988, p. 211.

Pour renforcer la crédibilité de l'hypothèse de l'unicité du préfixe, nous examinerons donc ses conditions d'apparition, d'une part, avec les adjectifs et, d'autre part, avec les verbes, et chercherons s'il est possible d'y voir un dénominateur commun.

Les adjectifs préfixés par *un-* sont extrêmement nombreux, mais il serait bien sûr faux de prétendre que le préfixe est compatible avec n'importe quel adjectif. *Un-* semble revêtir un sens « subjectif » : Jespersen remarque que les adjectifs préfixés tendent à être péjoratifs.[8] Si les remarques de Jespersen sont convaincantes, il est possible de trouver de nombreux contre-exemples, en particulier avec les adjectifs à forme de participes passés ou suffixés par *-able*/*-ible*, qui sont loin de toujours revêtir des connotations négatives : beaucoup d'entre eux sont neutres (*unaccompanied*, *unabridged*, etc.), voire positifs, au moins dans certains contextes (*undaunted*, *unbreakable*, etc.).

C'est pourquoi D.A. Cruse affine cette analyse à l'aide des paires *afraid/ unafraid*, *spoilt/ unspoilt* et *polluted/ unpolluted* et observe que, dans ces cas-ci, le préfixe forme le terme qui correspond à une évaluation positive (*unpolluted*, *unafraid*, *unspoilt*). Selon l'auteur, c'est la notion de norme qui rend compte le mieux de ces lexèmes :

> For instance, one cannot use *unafraid* to simply denote the lack of fear […]*Unafraid* is apt only in situations where it would be entirely normal to be afraid. Parallel constraints apply to the use of *unspoilt* and *unpolluted*. We may therefore postulate that when the meanings of a pair of opposites incorporate an explicit contrast of normality, the linguistically positive term is the one which embodies the notion of normality[9]

A l'idée de norme, nous préférons substituer celle d'attente de l'énonciateur, qui englobe la première, mais permet de rendre compte de l'abondance de formes participiales. Ces attentes priment sur l'évaluation ou « evaluative polarity ». Cette idée de norme et d'attente semble expliquer le recours à *un-* dans le cas des adjectifs. Dans un premier temps, nous chercherons à voir si cette hypothèse se vérifie, puis nous tenterons de la mettre à l'épreuve de verbes.

A première vue, saisir un rapport étroit entre l'apparition de *un-* dans les adjectifs et dans les verbes n'est pas évident car si la paraphrase du préfixe par *not* est souvent possible avec les premiers, dans la majorité des cas, elle est exclue avec les seconds : *undo*, *undress*, ne peuvent pas être glosés par « not do » et « not dress ». Bien qu'il puisse assumer les valeurs privative, réversative et ablative, *un-* est essentiellement un préfixe réversatif, il apparaît donc très fréquemment avec des verbes téliques et réversibles, et Horn rappelle

> As Dowty notes, the base of *un*-verbs is normally an accomplishment, and crucially involves a change of state. Thus we have *unbend*, *uncoil*, *uncurl*, *undress*, *unfold*, *unravel*, and *unwind* but not **uneat*, **ungo*, **unplay*, **unsmoke*, **unsnore*, **unswim*, or**unwalk*. [10]

Le même auteur note que le sens des verbes ainsi préfixé est relativement spécifique :

> The basic insight is that, as Covington (1981: 34) puts it, "The root verb to which *un-* attaches normally signifies putting something into a more marked or specialized state, and the derived *un*-verb

[8] « the same general rule obtains in English as in other languages, that most adjectives with *un-* or *in-* have a depreciatory sense: we have *unworthy*, *undue*, *imperfect*, etc., but it is not possible to form similar adjectives from *wicked*, *foolish*, or *terrible*. » (Jespersen, O*., Negation in English and Other Languages*. København : A. F. Høst og søn, 1917, p. 144.)

[9] Cruse : 1986, p. 250-251.

[10] Horn, L. « Uncovering the Un-Word: A Study in Lexical Pragmatics », *Sophia Linguistica*. Tokyo : Sophia University, Graduate School of Languages and Linguistics, 2002, vol. 49 : 1-64.
2002, p. 14.

> signifies returning it to normal. When the prefix attaches to a positive, goal-oriented accomplishment verb, the state-change depicted by the *un*-verb is one which in effect helps entropy along, rather than creating or restoring order.[11]

C'est, pour Horn, ce qui explique que *un-* puisse parfois venir renforcer sa base, au lieu d'opérer une négation ou une inversion, quand cette base réfère au retour d'un état E2 à un état antérieur E1, retour qui se produit naturellement, spontanément et requiert moins d'énergie que le passage de cet état initial E1 à l'état E2 : c'est pourquoi *thaw* donne *unthaw*, dont le sens est très proche, alors que *melt* ne permet pas de former de verbe **unmelt* :

> If melting is akin to thawing the transmutation from a solid (or frozen) to a liquid (or unfrozen) state, why don't we have redundant *unmelt* alongside *unthaw*? The answer is that melting is in fact crucially unlike thawing: while a thawed object, say a turkey, is a temporarily frozen theme or patient which return to its unmarked state, retaining its physical integrity, the integrity of a MELTED theme or patient may be affected or even destroyed by a change of state. A thawed turkey is still a turkey (indeed, a turkey par excellence)), a thawed lake still a lake, but a melted ice cube, snowman or wicked witch is not simply a different form of the same object, but essentially, a PUDDLE of one kind or another.[12]

Dans le cas des verbes téliques réversibles, le préfixe verbal *un-* a une valeur d'inversion et est lié à un retour à un état initial, retour qui paraît naturel, puisque les procès désignés par les verbes qui lui servent de bases sont réversibles. Les cas où il n'est pas négatif à proprement parler, mais renforce sa base correspondent à une transformation entropique, donc spontanée. De même, pour les adjectifs, le préfixe *un-* est corrélé à une attente. Il nous semble ainsi pouvoir mieux saisir l'unité de *un-* dans les adjectifs et les verbes : dans les deux cas, il correspond à non seulement à une « opposition », qu'elle soit concrète ou abstraite – mais, qui plus est, à une opposition prédictible, qui s'inscrit sur un fond d'attente.

Nous chercherons dans cette étude à déterminer si cette lecture est fondée et permet d'unifier, d'une part les incompatibilités de *un-* et, d'autre part, ses différentes valeurs sémantiques à travers les diverses parties du discours dans lesquelles il figure. Dans ce cas, un- véhciulerait non pas seulement un sens logique, mais serait également la trace d'une saisie particulière du réel.

Bibliographie :

ANDREWS, E. (1986). A Synchronic Semantic Analysis of *De-* and *Un-* in American English", *American Speech*, Vol. 61, No. 3., pp. 221-232.

BOONS, J.P. (1984). Sceller un piton du mur; desceller un piton du mur. Pour une syntaxe de la préfixation négative, *Langue Française*, 62, 95-128.

COLEN, A. (1980). On the Distribution of *un-, de-,* and *dis-* in English Verbs Expressing Reversativity and Related Concepts." *Studia Germanica Gandensia* 2 1: 127-52.

CRUSE, D.A. (1986). *Lexical Semantics*, Cambridge, London, New York, New Rochelle, Melbourne, Sydney, Cambridge University Press.

DIXON, R. M. W. (1991) *A New Approach to English Grammar, on Semantic Principles*. Oxford : Clarendon Press.

GERHARD-KRAIT, F. (2000). *La Préfixation en dé(s)-: Formes construites et interpretations*, Lille, Atelier National de Reproduction des Thèses.

[11] id., p.19.
[12] Horn : 1988, p. 218.

HAMAWAND, Z. (2009). *The Semantics of English Negative Prefixes*, London, Oakville, Equinox.

HORN, L. (1988). Morphology, Pragmatics and the *Un*-verb, ESCOL : 210-233.

HORN, L., (2001). *A Natural history of Negation*, Center for the Study of Language and Information.

HORN, L. (2002). « Uncovering the Un-Word: A Study in Lexical Pragmatics », in *Sophia Linguistica*. Tokyo : Sophia University, Graduate School of Languages and Linguistics, vol. 49 : 1-64.

JESPERSEN, O. (1917). *Negation in English and Other Languages*. København : A. F. Høst og søn.

LIEBER, R., (2004). *Morphology and Lexical Semantics*, Cambridge, Cambridge University Press.

MARCHAND, H. (1960). *The Categories and types of present-day English word formation, a synchronic-diachronic approach*. Wiesbaden : O. Harrassowitz.

MAYNOR, N., (Winter, 1979) "The Morpheme Un", *American Speech*, Vol. 54, No. 4 : 310-311.

ZIMMER, K.E. (1964). Affixal Negation in English and Other Languages. *Word*, 20, 2, Supplement, Monogrpah, 5, New York: Linguistic Circle of New York.

Persian complex predicates: Lexeme formation by itself

Olivier Bonami
Université Paris-Sorbonne & IUF
Laboratoire de Linguistique Formelle (UMR7110)
olivier.bonami@paris-sorbonne.fr

Pollet Samvelian
Université Sorbonne Nouvelle
Mondes Iranien et Indien (UMR7528)
pollet.samvelian@univ-paris3.fr

Septièmes Décembrettes, 3 décembre 2010

In Persian there is no productive morphological lexeme formation process outputting verbs. When they need to refer to a new event type, speakers resort to complex predicates (CPs). We argue that while CPs are clearly multiword expressions, CP formation has all the trappings of a lexeme formation process, and should be treated as such. Thus the class of verbal lexemes is open, but all new lexemes are multiword expressions rather than simplex words. We then propose an HPSG analysis whose key ingredient is a set of lexeme formation rules turning a noun into a verb subcategorizing for that noun.

About a dozen verbs are used in CPs, and nouns can either be predicative nouns or concrete nouns. For ease of reference we group in (1) all examples discussed in this abstract, sorted by verbs.

(1)

a.	*dast* hand	*andâxtan* throw	‘mock’	l.	čaqu knife	*zadan* hit	‘stab’
b.	*anjâm* accomplishment	*dâdan* give	‘realize’	m.	*dast* hand	*zadan* hit	‘touch, applaud’
c.	*dast* hand	*dâdan* give	‘shake hands’	n.	*faryâd* scream	*zadan* hit	‘scream’
d.	*tekân* movement	*dâdan* give	‘cause to move’	o.	*piano* piano	*zadan* hit	‘play the’ piano
e.	*dust* friend	*daštan* have	‘like’	p.	*qor* complaint	*zadan* hit	‘complain’
f.	*fekr* thought	*kardan* do	‘think’	q.	*sili* slap	*zadan* ‘hit	‘slap’
g.	*guš* ear	*kardan* do	‘listen’	r.	*šane* comb	*zadan* hit	‘comb’
h.	*dast* hand	*xordan* strike	‘be started on sth.’	s.	*šeyhe* neigh	*zadan* hit	‘neigh’
i.	juš joint	*xordan* strike	‘bind’	t.	*tašar* admonition	*zadan* hit	‘reprimand’
j.	*tekân* movement	*xordan* strike	‘be moved’	u.	*telefon* phone	*zadan* hit	‘phone’
k.	*xat* scratch	*xordan* strike	‘be scratched’	v.	*tohmat* slander	*zadan* hit	‘slander’

1 Evidence for multi-word status

The two elements in a CP are clearly separate syntactic atoms. All inflection occurs on the verb: the negative prefix occurs before the verb (2a)[1] , and the two elements can be separated by the future auxiliary (2b). Object pronominal affixes can attach to the noun (3a) in a CP, just as they can attach

[1]Abbreviations in glosses: DDO = definite direct object marker; EZ = *Ezafe* particle; NEG = negation.

generally to a complement (3b). The noun and verb can be separated by adverbs (3c). Both the nouns and verbs can be coordinated (4), and the noun can be extracted (5). Finally if the complex predicate is sufficiently compositional the noun can head a complex NP (6). These observations highlight the fact that the syntactic properties of complex predicates are identical to those of combinations of a verb with an object NP. While there is a tendency for the noun in a CP to be more cohesive with the verb than a bare direct object is (in terms of word order, stress, differential object marking, pronominal affix placement), there is no categorical syntactic contrast between the two types of sequences (*pace* Karimi-Doostan, 1997; Goldberg, 2003).

(2) a. Maryam Omid=râ dust na-dâr-ad
Maryam Omid=DDO friend NEG-have-3S
'Maryam does not like Omid.'

b. Maryam Omid=râ dust xâh-ad dâšt
Maryam Omid=DDO friend want-3S had
'Maryam will like Omid.'

(3) a. Dust=aš dâr-am
friend=3S have-1S
'I like her/him/it.'

b. Be bâzâr=aš bord.
to bazar=3S took
'(S)he took it to the bazar.'

c. Maryam Omid=râ dust aslan na-dâr-ad
Maryam Omid=DDO friend absolutely NEG-have-3S
'Maryam does not like Omid at all.'

(4) a. Maryam mu-hâ=yaš=râ bros va šâne zad
Maryam hair-PL=3S=DDO brush and comb hit
'Maryam brushed and combed her hair.'

b. Omid sili zad va xord.
Omid slap hit and strike
'Omid gave and received slaps.'

(5) Dust to Maryam=râ dâr-i?
friend you Maryam=DDO have-2S
'Is that Maryam whom you like?'

(6) Maryam [xabar=e marg=e Omid]=râ be mâ dâd.
Maryam news=EZ death=EZ Omid=DDO to us gave
'Maryam told us about Omid's death. (litt. gave us the news of Omid's death)'

2 Evidence for lexemic status

While complex predicates are multi-word combinations, the combination as a whole should be seen as the exponent of a single lexeme. Such an analysis is evidently needed in cases where the meaning of the complex predicate is opaque. The new idea we want to defend here is that productive complex predicate formation is a case of lexeme formation. We provide four arguments to this effect.

CPs are lexicalized

N-V combinations are subject to various levels of lexicalization, in a way that closely parallels what is seen with lexemes formed by morphological means. It is barely ever the case that the meaning of a CP is fully predictable from the meaning of its component parts—(1d, 1o) are good but rather isolated candidates. In many cases the CP meaning is a specialization of the predictable meaning of the combination (1c, 1l, 1m, 1r), but this particular specialization has to be learned. In other examples

semantic drift has taken place; the link between the compositional meaning and the lexicalized meaning is sometimes still recoverable synchronically (1g, 1h, 1u), sometimes not (1a, 1e). Analogy often plays an important role in motivating new lexicalizations: in (1n, 1p, 1s, 1t, 1v) the CP is formed by analogy with preexisting combinations such as (1l, 1q), not by specialization or drift from a nonlexicalized combination. Finally, even when the contribution of the verb to the CP meaning is clear, there is quite often no semantic justification for the choice of a particular verb (1b, 1f)—a situation familiar from support verb constructions, but also from affix rivalry situations.

CPs feed lexeme formation rules

N-V combinations serve as inputs to further lexeme formation rules. We give two examples of a very widespread phenomenon. (i) the suffix *-i* forms abilitative adjectives from verbs, e.g. *xordan* 'eat' > *xordani* 'edible' (and by further conversion > *xordani* 'food'). This suffix is found in combination with CPs, independently of whether they are compositional or not (7). (ii) perfect participles can regularly be converted to adjectives, and this process readily applies to CPs.

(7) a. dust daštan > dustdaštani
'love (1e)' 'lovely'

b. xat xordan > xatxordani
'be scratched (1k)' 'scratchable'

c. juš xordan > jušxordani
'bind (1i)' 'linkable'

(8) a. dast xordan > dastxorde
'be started on sth. (1h)' 'sullied'

b. xat xordan > xatxorde
'be scratched (1k)' 'scratched'

c. juš xordan > jušxordane
'bind' (1i) 'bound'

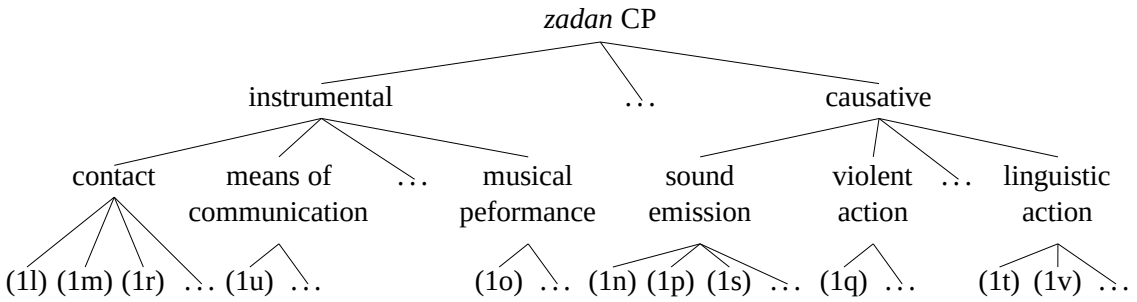
Paradigmatic groupings of CP forming verbs

Verbs used in CPs group in families with similar, if not undistinguishable, effects. *Dâdan*, *kardan* and *zadan* form instrumental or causative CPs, while *xordan*, *šodan* and *yâtan* form unaccusatives (compare 1d to 1j). Two verbs of the same family usually do not give rise to concurrent CPs, unless one of the combinations has been specialized or demotivated (compare 1c to 1m). This type of pattern closely parallels (partial) blocking effects in morphological lexeme formation (e.g. Aronoff, 1976).

Clustering of CPs based on the same verb

Complex predicates sharing the same verbal element group into clusters of related combinations with varied levels of internal coherence and of productivity, as the partial classification in (9) illustrates. However there is only a family resemblance among the clusters. Once again this is strikingly familiar to what is observed for morphological lexeme formation (e.g. Riehemann, 1998).

(9)



3 Analysis

Persian CPs are lexemes, but lexemes whose exponents take the form of combinations of two words. To borrow Gaeta and Ricca (2009)'s vocabulary, they are [+lexical,−morphological] constructions. Most existing analyses of Persian CPs are problematic because they confuse the two dimensions of analysis, and argue that PCs are words (e.g. Karimi-Doostan, 1997), phrases (e.g. Ghomeshi and Massam, 1994; Folli et al., 2005), or 'words by default' (Goldberg, 2003).

For the representation of individual PC lexemes we adapt the analysis set forth by (Müller, in press) (10): the PC enters syntax as a word form of a verbal lexeme which subcategorizes for a specific noun through the dedicated Lexical IDentifier (LID) feature (Sag, 2007; Spencer, 2004), which we

redefine as individuating lexemes via their morphological paradigm type (MP) and main semantic relation (MREL). The actual combination of verb and noun is then a matter of regular syntax.

(10) Lexical entry for the lexeme *dust dâštan* 'to like/love'.

$$\begin{bmatrix} \text{LID} & \begin{bmatrix} \text{MP} & \begin{bmatrix} \textit{dâštan} & \\ \text{STEM1} & \text{dar} \\ \text{STEM2} & \text{daʃt} \end{bmatrix} \\ \text{MREL} & \textbf{love}(i,j) \end{bmatrix} \\ \text{ARG-ST} & \left\langle \text{NP}_i, \text{NP}_j, \text{N}\begin{bmatrix} \text{LEX} & + \\ \text{LID}\,|\,\text{MP} & \textit{dust} \end{bmatrix} \right\rangle \end{bmatrix}$$

New CPs are the product of lexeme formation rules (LFRs) such as the one in (11) for intransitive instrumental CPs using *zadan*. Just like a morphological LFR, (11) turns a noun into a verb, which denotes an event type involving the use by an agent of an instance of this noun as an instrument. Two features of the new verb are unusual. First, it shares its morphological paradigm with the lexeme *zadan*, although it has a different semantics and thus a different LID. Second, it selects as a complement for a word with the same LID as the input word—thus in effect, the LFR turns a noun into a verb selecting for that noun.

(11)
$$\begin{bmatrix} \textit{lexeme} & \\ \text{CAT} & \begin{bmatrix} \textit{noun} & \\ \text{LID} & \boxed{1} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & j \\ \text{RELS} & \mathbf{P}(j) \end{bmatrix} \end{bmatrix} \mapsto \begin{bmatrix} \textit{lexeme} & \\ \text{CAT} & \begin{bmatrix} \textit{verb} & \\ \text{LID} & \begin{bmatrix} \text{MP} & \textit{zadan} \\ \text{MREL} & \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & e \\ \text{RELS} & \exists j[\mathbf{P}(j) \wedge \boxed{2}\mathbf{use}(e,i,j)] \end{bmatrix} \\ \text{ARG-ST} & \left\langle \text{NP}_i, \begin{bmatrix} \text{LEX} & + \\ \text{CAT} & \begin{bmatrix} \textit{noun} & \\ \text{LID} & \boxed{1} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}$$

Since this is a standard HPSG lexeme-to-lexeme rule, familiar analytic techniques can be applied to account for the rest of the properties. LFRs are organized in an multiple inheritance hierarchy (Riehemann, 1998), where information shared by rules based on the same verb, or having the same semantic effect, can be factored out as properties of a common supertype. Lexicalized CPs are elements of the lexical hierarchy with a frozen, non-compositional semantics (Koenig, 1999). Finally this analysis integrates readily with existing analyses of Persian morphology and syntax in HPSG.

References

Aronoff, M. (1976). *Word Formation in Generative Grammar*. Cambridge: MIT Press.

Folli, R., Harley, H., and Karimi, S. (2005). 'Determinants of event type in Persian complex predicates'. *Lingua*, 115:1365–1401.

Gaeta, L. and Ricca, D. (2009). '*Composita Solvantur*: Compounds as lexical units or morphological objects?' *Rivista di Linguistica*, 21:35–70.

Ghomeshi, J. and Massam, D. (1994). 'Lexical/syntactic relations without projection'. *Linguistic Analysis*, 23:175–217.

Goldberg, A. (2003). 'Words by default: Inheritance and the Persian complex predicate'. In E. J. Francis and L. A. Michaelis (eds.), *Mismatch : Form–function incongruity and the architecture of grammar*. Stanford: CSLI Publications, 117–146.

Karimi-Doostan, G. (1997). *Light verb constructions in Persian and Kurdish*. Ph.D. thesis, University of Essex.

Koenig, J.-P. (1999). *Lexical relations*. Stanford: CSLI Publications.

Müller, S. (in press). 'Persian complex predicates and the limits of inheritance-based analyses'. *Journal of Linguistics*, 46.

Riehemann, S. (1998). 'Type-based derivational morphology'. *Journal of Comparative Germanic Linguistics*, 2:49–77.

Sag, I. A. (2007). 'Sign-based construction grammar: An informal synopsis'. Ms., Stanford University.

Spencer, A. (2004). 'Generalized paradigm function morphology'. Ms., University of Essex.

Reading derived words by Italian children with and without dyslexia:

The effect of root length

Cristina Burani [1], Stefania Marcolini [1], Daniela Traficante [2], & Pierluigi Zoccolotti [3 4]

[1] Institute of Cognitive Sciences and Technologies (ISTC), CNR, Rome
[2] Department of Psychology, CR*Idee*, Catholic University, Milan
[3] Department of Psychology, Sapienza University of Rome
[4] Neuropsychology Unit, IRCCS Fondazione Santa Lucia, Rome

INTRODUCTION.

Italian children with dyslexia experience difficulties in reading long stimuli and show an extremely slow and analytical reading behaviour. Recently, we showed that word naming times of children with dyslexia were shorter for stimuli composed of a root and a derivational suffix (*e.g.*, CASS-IERE, 'cashier'), as compared to simple words of the same length and frequency not parsable in root + derivational suffix (*e.g.*, CAMMELLO, 'camel') (Burani *et al.*, 2008). We proposed that morphemes prove useful in processing linguistic stimuli in children with limited reading ability for whom most printed words are too large units to be processed as a whole.

A similar facilitation on naming times due to morphological composition was found in skilled readers of the same age. However, whereas skilled readers showed a reading advantage of morphological composition for new words (pseudowords) and low-frequency words only, children with dyslexia were facilitated both in new words and in words of different frequency, including high-frequency words (Marcolini *et al.*, in press). For both children with dyslexia and skilled readers, the facilitating effect on naming times was mainly driven by the root (Traficante *et al.*, 2010). Overall, the facilitating effect of word's morphological composition is larger in children with dyslexia as compared to skilled readers of the same age (see also Carlisle & Stone, 2005; Elbrö & Arnbak, 1996).

The present study addresses the question of whether the facilitating effect found on low-frequency derived word reading is moderated by the length of the root. For adult readers, shorter

affixed words are more subject to whole-word processing than longer affixed words, with increasing word length enhancing the probability of compositional processing (*e.g.*, Niswander-Klement & Pollatsek, 2006). To the best of our knowledge, the role of root length has not yet been investigated. We expected that, in the case of skilled young readers, the likelihood of relying on the root morpheme as a processing unit would depend on its perceptual salience within the word (see Kuperman *et al.*, 2010; Laudanna & Burani, 1995, for processing accounts of the perceptual salience of affixes), thus biasing reading toward morphological decomposition. In contrast, readers with dyslexia might be facilitated by the presence of a root within a derived word, irrespective of root's length, because of their difficulty in processing the word as a whole.

METHOD

Participants. Twenty children with dyslexia and 40 skilled readers, matched for gender, age and non-verbal intelligence were recruited from 6th grade classes of junior high school in Milan and from the Centre for Cognitive and Linguistic Disorders (ASL 1) in Rome.

Materials. Two sets of derived words, 30 with a short root (3-4 letters) and 30 with a long root (5-6 letters), and two sets of 30 simple words, matched to each set of derived words for length, word frequency and familiarity, bigram frequency, initial phoneme characteristics, and orthographic complexity, were selected. Words had a low frequency in a child's written frequency count. Derived words were composed of a root and a derivational suffix. They were phonologically and semantically transparent with respect to their base word, and included highly familiar roots and suffixes. Suffix length was the same in the two derived sets (mean: 3.7 letter length). Accordingly, short-root derived words (*e.g.*, DUR-EZZA, 'hardness') were shorter than long-root derived words (*e.g.*, LONTANANZA, 'distance'). Simple words (*e.g.*, DIFETTO, 'defect'; LIQUIRIZIA,

'liquorice', matched to each of the two length sets, respectively) were not parsable into root + derivational suffix. [1]

Procedure. Participants were instructed to read aloud as fast and accurately as possible the words that appeared singly on a computer screen. Responses were recorded by a microphone connected to a voice-key. Naming reaction times (RTs) were measured in milliseconds (ms). The experimenter noted mispronunciation errors.

RESULTS

ANOVAs were conducted for short-root and long-root derived words separately, comparing each set with the corresponding set of simple words. By-participants ANOVAs with group (children with dyslexia and skilled readers) as unrepeated factor, and morphological type (derived *vs.* simple) of the words as repeated factor were carried out on logarithmically transformed RTs and arcsine transformed error scores. In the by-items ANOVAs, word morphological type was the unrepeated factor and readers' group was the repeated factor.

Naming times were faster and more correct for skilled readers than for children with dyslexia. However, while children with dyslexia took advantage of the morphemic structure of the word both with short-root and long-root stimuli, naming derived stimuli faster than simple words in both sets, skilled readers benefited from the morphemic structure only with long-root stimuli. For children with dyslexia, the presence of a long root also led to higher accuracy for derived than for simple words; for skilled readers pronunciation errors were few and no morphological effect emerged.

Post-hoc regression analyses on the derived words of the two length sets all together, were also conducted on RTs (as well as a sequential analysis of variance; see Baayen, 2008) to partial out the contribution of root length from the possible confounding influence of word length. For skilled readers, root length significantly predicted naming times with faster RTs for longer roots, over and

[1] About one half of the simple words ended with a pseudo-suffix (e.g, DIFETTO and LIQUIRIZIA, in which both -ETTO and -IZIA could be an Italian suffix). The number of pseudo-suffixed words was matched in the two length sets. None of the pseudo-suffixed simple words included a real root.

above the inhibitory effect of word length. The naming times of children with dyslexia were not predicted by root length.

DISCUSSION.

The results showed that, for skilled readers, the more perceptual salience a root has, the more it stands out of its embedding word, and the more biased the processing is towards using the root and the suffix for reading the complex word. In contrast, readers with dyslexia, for whom full-form processing is overall difficult, always showed a benefit from morphological processing, with faster naming times of words composed of root and suffix, irrespective of root length. These findings need to be accommodated within models of morphological processing that could account both for perceptual properties of morphemes, such as root length, and differing reading abilities.

REFERENCES.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.

Burani, C., Marcolini, S., De Luca, M, & Zoccolotti, P. (2008). Morpheme-based reading aloud: Evidence from dyslexic and skilled Italian readers. *Cognition, 108*, 243-262.

Carlisle, J. F., & Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly, 40*, 428-449.

Elbrö, C., & Arnbak, E. (1996). The role of morpheme recognition and morphological awareness in dyslexia. *Annals of Dyslexia, 46,* 209-240.

Kuperman, V., Bertram, R., & Baayen, R. H. (2010). Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language, 62*, 83-97.

Laudanna, A., & Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 345-364). Hillsdale, LEA.

Marcolini, S., Traficante, D., Zoccolotti, P., & Burani, C. (in press). Word frequency modulates morpheme-based reading in poor and skilled Italian readers. *Applied Psycholinguistics*.

Niswander-Klement, E., & Pollatsek, A. (2006). The effects of root frequency, word frequency, and length on the processing of prefixed English words during reading. *Memory & Cognition, 34*, 685-702.

Traficante, D., Marcolini, S., Luci, A., Zoccolotti, P., & Burani, C. (2010). How do roots and suffixes influence reading of pseudowords: A study of young Italian children with and without dyslexia. *Language and Cognitive Processes*, DOI 10.1080/01690965.2010.496553

Non-templatic truncation: the case of vocatives

Teresa Cabré Monné(UAB) and Maria del Mar Vanrell Bosch (UAB-UPF)
teresa.cabre@uab.cat, mariadelmar.vanrell@upf.edu

Languages can display different truncation processes which involve mapping a base form onto a particular prosodic template. In positive truncation the templatic sequence is taken and what is deleted may be inconsistent. By contrast, in substractive truncation a specific prosodic shape is applied to the deleted material, and it is the prosodic shape of the present material that may be inconsistent. Vocatives in Romance languages such as Algherese Catalan, Sardinian, Corsican or Southern varieties of Italian seem to be closely related to the substractive truncation process since the surface material can present different lengths, as can be seen in (1). As we can see in (2), the deleted sequence does not obey any prosodic template either. Thus, vocatives can exhibit different lengths depending on the syllabic distance from the beginning of the word to the stressed syllable and the residue can also show varying lengths ([*Cá*][*rmine*], [*Marí*][*a*], [*Pasqualí*][*no*]).

The question that arises with the data at hand is how such non-templatic truncation can be accounted for. Vocative truncation is driven by two enhanced positions: the word-initial syllable and the stressed syllable of the word (Beckman, 1998). In Alber's words, truncated vocatives "preserve the string from the first segment up to the stressed vowel of the base name" and […] "they do not obey any template but anchor to the first as well as to the stressed syllable of the base" (Alber, 2009). Since they preserve the segmental sequence from the word-initial syllable to the stressed vowel of the word, the deleted material is unstressed and, consequently, not psycholinguistically prominent (Beckman, 1998). This phenomenon is not only limited to vocative forms, since imperative forms like the ones show in (3) are also found in Algherese Catalan, Sardinian or Southern Italian. This is related to the fact that both imperative and vocative forms, which share a similar pragmatic function, can display minimal flexive affixation, a general tendency found in Indoeuropean languages. As far as declensional languages are concerned, there is an important tendency for vocative and imperative forms to be reduced to the bare stem (Palmer, 1955).

Focusing on vocatives, we think that because of this relative minimal/absence of inflection, other non-inflectional marks are the key to expressing the vocative force. With respect to languages like Algherese Catalan, Sardinian, Corsican or Southern varieties of Italian, vocative force can be expressed by the particle *o* (see data from Algherese Catalan and Corsican in (1)). However, this particle is not solely found in these Romance languages, but it may also be shown to head the noun or noun phrase, especially in poetic or solemn rhetorical speech, from Latin to English or even Celtic languages such as Scottish Gaelic or Irish, in which this particle adopts the form *a* (see Floricic, 2009: 9).

Vocatives are characterized by a specific intonational contour. By contrast, templatic truncated words can appear with different intonational contours depending on the pragmatic intention of the speaker. Hence, in the specific case of Algherese Catalan, the non-emphatic vocative is characterized by a falling pitch accent (a H+L* accent in the Cat_ToBI system – Prieto, P., Aguilar, L. et al., 2009) aligned to the stressed syllable of the word (see Figure 2), while emphatic vocatives present a rising pitch accent (L+H* in the Cat_ToBI system) associated with the first syllable in the word (see Figure 2). Notice that both intonational patterns are associated with two prominent positions, namely, the initial syllable and the stressed syllable of the word, which coincide with the relevant boundaries for non-templatic truncated vocatives.

In this paper, we will show how the atemplatic shape of these vocatives is reinterpreted as a prosodic sequence guided by the two strong positions of the word. These positions are reinforced by the alignment of the two pitch accents: the falling one is associated with the stressed syllable in the case of non-emphatic vocatives, whereas the rising one is aligned to the first syllable in emphatic vocatives. To sum up, vocatives are characterized by a fixed intonational pattern and the non-templatic truncation process is optional. On the contrary, templatic truncated words are only characterized by a fixed prosodic pattern (foot, syllable or mora) and the morphological process is obligatory. According to that, we suggest that the intonational pattern acts as the template of vocatives.

Examples and figures

(1)

Data from Kuen (1932) and, Prieto and Cabré (2008) for Algherese Catalan:

Pàuru! (proper name) > *Pa!*	*Ròsa!* (proper name) > *Rò!*
Barbarína! (proper name) > *Barbarí!*	*Tarésa!* (proper name) > *Taré!*
Antòni! (proper name) > *Antò!*	*Ríta!* (proper name) > *Arrí!*
Fabio! (proper name) > *o Fà!*	*Juàn!* (proper name) > *o Juà!*

Data from Pittau (1952) for Sardinian:

Préddu (proper name) > *Pre'*	*Mariánzela (proper name)* > *Maria'*
Tzú, -a (uncle/aunt) > *tzi'*	*María (proper name)* > *Mari'*

Data from Floricic (2002) for Corsican:

Babbu (father) > *O Ba'*	*Nipoti* (nephew)> *O Nipó*
Fiddolu (godson) > *O Fiddó*	*Frateddu* (brother) > *O Fraté*

Data from Alber (2009) for southern varieties of Italian:

Bárbara > *Bá*	*Salvatóre* > *Salvató*
Carmé > *Carméla*	*Antonélla* > *Antoné*

(2)

Data from Prieto and Cabré (2008) for Algherese Catalan:

Doménico [Domé] [nico]	*Fábio* [Fa] [bio]	*Pasqualíno* [Pasqualí] [no]
Francésco [Francé] [sco]	*María* [Marí] [a]	*Cármine* [Cá] [rmine]

(3)

Examples obtained by personal communication from Luca Scala and Filippo Melis:

Data from Algherese	Data from Sardinian Italian	Data from Sardinian
escólta (listen to me) > *escó*	*fái* (do that) > *fá*	*nára* (say) > *ná*
míra (look)> *mí*	*vái* (go) > *vá*	*trúbba* (hurry up)> *trú*
dóna (give me) > *dó*	*stái* (be) > *stá*	*míra* (look) > *mí*

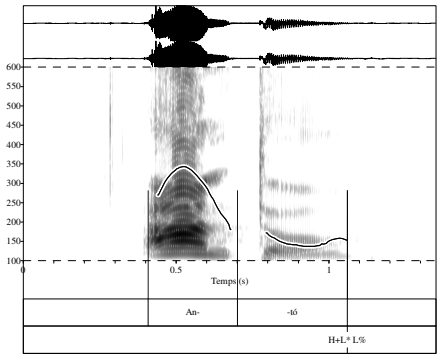


Figure 1. Waveform display, spectrogram, F0 contour, and prosodic labeling of the vocative *Antó* < *Antóni* with a falling pitch accent aligned to the nuclear syllable of the noun.

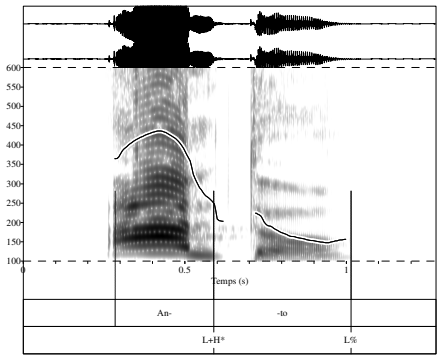


Figure 2. Waveform display, spectrogram, F0 contour, and prosodic labeling of the vocative *Anto* < *Antóni* with a rising pitch accent aligned to the first syllable in the noun.

References

Alber, B. (to appear). An Exploration of Truncation in Italian. Rutgers Working Papers in Linguistics.

Beckman, J. 1997. Positional Faithfulness. Doctoral dissertation, University of Massachusetts,

Floricic, F. (manuscript). Le Vocatif et la périphérie du système des cas: exemples tirés des domaines roman et slave.

Palmer, L.R. 1955. *Introducción al latín*, Barcelona: Ariel, 1988.

Prieto, P; Aguilar, L.; Mascaró, I.; Torres, F.J.; Vanrell, M.M. 2009. "L'etiquetatge prosòdic Cat_ToBI ". *Estudios de Fonética Experimental*, vol. XVIII. 287-309. ISSN 1575-5533.

Prieto, P. and T. Cabré (coords.). 2008. Atles interactiu de l'entonació del català. Web site: http://prosodia.uab.cat/atlesentonacio/

Phonological detail for accessing morphological structures.
Human and artificial responses in comparison

Basilio Calderone (Modyco, CNRS & Université Paris Ouest – la Défense)
Chiara Celata (Scuola Normale Superiore, Pisa)
Fabio Montermini (CLLE-ERSS, CNRS & Université de Toulouse)

1. *Introduction. The morphological impact of the quantitative properties of the lexicon*

Within the most recent models of morpholexical processing, morphemes are not recognized in isolation but rather relationally in the context of other phonologically similar material (Luce, Pisoni & Goldinger 1990, among others). In this view, morpholexical processing is to a great extent affected by statistic properties of the lexicon (or sub-parts of the lexicon), and primarily by quantitative properties of affixes, such as neighborhood density and relative frequency of morphemes (Baayen 2003; Goldsmith 2001). For example, function morphemes and lexical morphemes compete with each other for recognition, but since function words are much more frequent than their nearest lexical neighbours, they escape the inhibitory effects of high neighborhood density. Therefore, the processing of function words is predicted to be relatively efficient thanks to relative frequency (which is high) and in spite of lexical neighborhood (which can be high or low regardless) (Segalowitz & Lane 2000).

In Italian, a language which shows high morphological complexity while being transparent for orthography-to-phonology conversion, morpholexical reading proved an efficient and frequently activated routine for both word recognition and - more recently - word naming. It has been demonstrated that "pronunciation is obtained through activation of morphemic units both in the input and in the output lexical components", and "[in those components] the formal representations of morphemes are stored, with no necessary involvement of a semantic component as well" (Burani & Laudanna 2003: 260). As many other Indo-European languages, particularly of the fusional type, Italian shows bound inflectional morphemes predominantly inserted by suffixation; most grammatical relations and relational categories are overtly expressed by morphological *endings* more often than by other types of affixes. Consequently, bound function morphemes tend to occupy the right edge of the word. From a quantitative point of view, a contrast between the left and the right edge of a word may be trivially set up by the different statistical properties of morphemes that tend to occur in either position of the word. One and the same phonological sequence will define a set of different quantitative properties (absolute 'token' frequency; frequency of the lexical forms in which it appears; number of neighbours etc.) depending on its position in the word. These differences will necessarily impact over lexical processing altogether.

2. *Micro- and macro-phonotactics: previous SOM-based simulations*

A quite basic generalization such as the one stated above can be seen as a working hypothesis to empirically test the emergent nature of morpholexical processing of complex words in natural (inflecting) languages (Bybee 2007; McClelland et al. 2002). Statistical distributions for word-final vs. non-word-final phonotactic regularities may be conceived of as a prerequisite for the emergence of paradigms and analogical rules for speakers of languages where morphology is shaped by different affixation preferences.

Previous research has shown that positional variables (i.e., the occurrence of the same sound sequence in initial vs. final position within the words, all other things equal) constitute psychological and computational significant preconditions for morphological parsing in Italian (Celata & Calderone 2010). In particular, it was verified that the salience of the right side of

morphological complex words (i.e., the portion usually occupied by function morphemes) emerges as a by-product of *micro-phonotactic* preferences (sequential information among segments) and sub-lexical frequency effects (or *macro-phonotactics*: positional information within the word). This hypothesis was tested on a behavioural and a computational ground, within an experimental protocol aimed at correlating speakers' responses with computational outputs obtained over one and the same linguistic data set. Morphologically complex pseudo-words were used to elicit similarity values from both native Italian subjects and an unsupervised topographic map (*Self-Organising Map*, Kohonen 2002) trained with a phonologically encoded corpus of spoken Italian. SOMs are plausible models of neural computation and learning given their sensitivity to frequency patterns in the input data and the incremental (i.e., adaptive) organisation of stimuli (see Pirrelli et al. 2004). The SOM operates on a phonotactic level by mapping similar input tokens (defined in terms of tri-gram scanning, e.g. #,T,H; T,H,E; H,E,#) onto adjacent output neurons. To obtain a final vector representation of the word, the system performs a generalization process by summing the activation values of each tri-gram. The cumulative action of tri-grams' activations gives a graded and distributed representation of the word in which both phonological similarity (at the string level) and token frequency effects (at the word level) are taken into account.

Morphologically complex pseudo-words were created by associating a non-root to an Italian inflectional or derivational affix, which was placed in either initial or final position (e.g., *fera**sto*** vs. ***sto**fera*). Three associated items (made up of the same affix + a different non-root) were created for each pivot item (e.g., *milu**sto***, *lu**sto**mi*, ***sultimo*** were associated to *fera**sto***, and ***sto**milu*, *lu**sto**mi*, ***sultimo*** to ***sto**fera*) (see Table 1). The three associates of each set were exactly equivalent to each other with respect to the segmental composition, but different to the extent that the affix could be placed in either the same, or a different position with respect to the affix contained in the pivot. Both the artificial system and the pool of native Italian subjects were asked to judge the similarity of each pivot item with respect to the three associated items. Results showed that the condition in which the affix position coincided between pivot and associated item crucially elicited higher similarity values, with respect to the non-coincidence conditions, for pivots with word-final affixes more than for pivots with word-initial affix (*fera**sto*** much more similar to *milu**sto*** than to *lustomi* and *sultimo*, yet ***sto**fera*-***sto**milu*** not so much different from ***sto**fera*-*lustomi* and ***sto**fera*-*sultimo*) These data were taken as evidence that morphotactic salience, preliminary to any morphological analysis, may emerge as a by-product of distributional information at the string level and positional regularities at the word level, derived from generalizations over the inflecting nature of the language. The Pearson's correlation coefficient between the observed and simulated behaviour (r = 0.508) reported a statistically significant correlation ($p < .001$), thus confirming the psychological plausibility of the SOM-based simulation.

3. *Phonological specifications for emergent morphotactics and morphology: lexical stress*

In this study, we replicate the experimental framework to the extent that human and artificial data are elicited and compared in response to the same set of Italian pseudo-words, but some relevant changes are introduced in the domain of (1) the nature of the phonological information coded in the corpus used for SOM training, and (2) the generalization process used to derive the system's word-level representation. The two domains are strictly interconnected, inasmuch as a modification in the phonological representation of data requires specific amendment in the patterned sampling operated by the SOM-based network.

(1) In the previous experiment, words were phonologically coded following a grid of place, manner of articulation and voicing specifications (Celata & Calderone 2010 for details). In particular, vowels were specified for height and anteriority. In the present experiment, vowels are additionally specified for stress, thus distinguishing stressed vs. unstressed vowels. Given the distinctive value of lexical stress in Italian, stress specifications provide the system with a more detailed representation of the input, which is supposed to mirror the

representation of pseudo-words in native speakers' phonological competence. We consider in fact that stress pattern is part of speakers' lexical knowledge.

(2) Stress, however, is a property of (phonological) words, not just of syllables or, even less so, single phonemes. The addition of a [± stressed] feature to the set of specifications for vowels does not cope indeed with the supra-segmental dimension of lexical stress codification. For this reason, the phonotactic information previously recovered by the system by means of a tri-gram sampling of the input forms is substituted here by an algorithm of full-word memorisation achieved through sampling of larger portions of the stimulus. This associative-like lexical memorisation is then used by the system in order to produce a vector word-level representation (Figure 1).

A corpus of written Italian from the *Leipzig Corpora Collection* (Quasthoff et al. 2006) phonologically transcribed is used for the SOM training phase (word types: nearly 80,000; word tokens: nearly 5 millions). Words are phonologically coded following the specifications in (1) above. After the training, the map is able to spatially organize phonotactic sequences defined in terms of whole-word *N*-grams. Similar sequences are found in adjacent areas of the map. Each sequence is identified by a pair of coordinates in the bi-dimensional map and an activation value roughly corresponding to the frequency of occurrence in the corpus. Then the final representation of the word is performed through the cumulative action of *N*-grams' activations, allowing a graded and distributed representation of the word where both phonological similarity and token frequency effects are taken into account.

Similarly to the previous experiment (see above, §2), an activation-based representation is derived for each experimental pseudo-word, and the similarity between pivot and associates is calculated in terms of the cosine distance between the two output values. The values are then directly correlated to subjects' performances on the word similarity judgment task reported on in Celata & Calderone (2010) (Figure 2).

The results support the hypothesis that phonological specifications at the supra-segmental level improve the system's performance in recovering the phonological similarity of stimuli as shaped by positional, i.e., morphotactic regularities at the word level, thus providing a more accurate simulation of native speakers' performance on the same task. Differently from the previous experiment, where vowels were unspecified for stress, we find here a significant interaction between association type and affix position ($F = 5.566$, $p < .05$), thus confirming the system's ability to recover word-level 'paradigmatic' regularities, besides string-level phonotactic information. Moreover, the Pearson's correlation coefficient changes from $r = 0.508$ of the previous experiment, to $r = 0.569$, thus indicating that supra-segmental information allows the system to overlap to a larger extent human generalizations in morpholexical processing.

References

Baayen H. (2003) Probabilistic approaches to morphology. In R. Bod, J. Hay & S. Jannedy (eds.) *Probabilistic linguistics*. MIT Press, Cambridge MA, 229-287.

Burani C. & A. Laudanna (2003) Morpheme-based lexical reading: Evidence from pseudo-word naming. In E. Assink & D. Sandra (eds.), *Reading complex words: Cross-language studies*, Dordrecht, Kluwer: 241-264.

Bybee J. (2007) *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Calderone B., C. Celata & I. Herreros (2008) Recovering morphology from local phonotactic constraints, abstract edited after review for Laboratory Phonology 11th Conference (LabPhon11) «Phonetic Detail in the Lexicon»,Wellington, New Zealand 30 June - 2 July 2008.

Celata C. & B. Calderone (2010) Restrizioni fonotattiche, pattern lessicali e recupero delle regolarità morfologiche. Evidenze computazionali e comportamentali. In *Linguaggio e cervello / Semantica,* Atti del XLII Convegno della

Società di Linguistica Italiana (Pisa, Scuola Normale Superiore, 25-25 settembre 2008), Roma, Bulzoni: 47-72. Volume 2 (CD ROM).

Goldsmith J. (2001) Unsupervised Learning of the Morphology of a Natural Language. *Computational Linguistics* 27(2): 153-198.

Kohonen T. (2001) *Self-Organizing Maps*. Heidelberg: Springer-Verlag.

McClelland J. & K. Patterson (2002) 'Words Or Rules' cannot exploit the regularity in exceptions (Reply to Pinker and Ullman). *Trends in Cognitive Science* 6: 464-465.

Luce P.A., D. B. Pisoni & S.D. Goldinger (1990) Similarity neighborhoods of spoken words. In G. Altmann (ed.), *Cognitive Models of Speech Processing*, Cambridge, MIT Press: 122-147.

Pirrelli V., B. Calderone, I. Herreros & M. Virgilio (2004) Non-locality all the way through: Emergent Global Constraints in the Italian Morphological Lexicon. *Proceedings of the 7th Meeting of the ACL Special Interest Group on Computational Phonology* (SIGPHON), Barcelona: 11-19.

Segalowitz S. & K. Lane (2000) Lexical access of function versus content words , *Brain and Language* 75: 376-389.

Quasthoff U., M. Richter, C. Biemann (2006) Corpus Portal for Search in Monolingual Corpora, *Proceedings of the fifth international conference on Language Resources and Evaluation*, LREC 2006, Genoa: 1799-1802.

Table 1. Example of pseudo-words.

		Position	
		Final position	Initial position
	Pivot	*fera**sto***	***sto**fera*
Association Type	Association 1	*milu**sto***	***sto**milu*
	Association 2	*lu**sto**mi*	*lu**sto**mi*
	Association 3	***sul**timo*	***sul**timo*

Figure 1. Architecture of he SOM-based input-output mapping function.

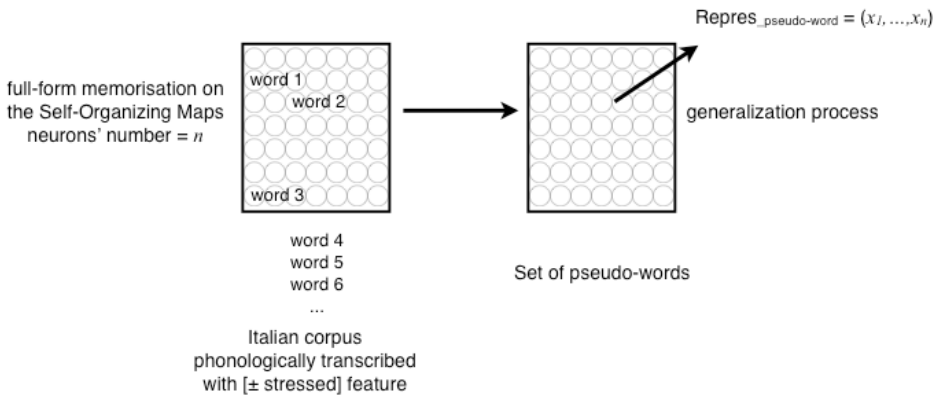
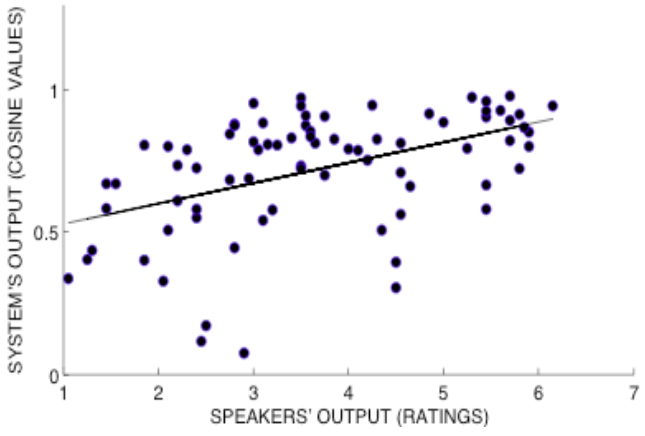


Figure 2. Global correlation between speakers' similarity ratings and computational cosine values.



Adjectifs toponymiques et ethniques en *–ský* en slovaque : construction morphologique et interprétation.

Iveta Chovanová
ATILF (Nancy-Université & CNRS)

Abstract : L'objectif de cette proposition est de fournir un traitement de la construction morphologique des adjectifs toponymiques et ethniques slovaques marqués par le suffixe *–ský*. Cette étude de corpus permet de rendre compte des aspects sémantiques et formels des adjectifs relationnels en tant que constructions paradigmatiques mettant en jeu trois types de lexèmes.

Mots clés : *morphologie du slovaque – dérivation – adjectifs toponymiques et ethniques – constructions paradigmatiques*

1. Objet. Cette proposition traite de la construction morphologique des adjectifs toponymiques et ethniques slovaques (désormais $A_{Topo/Ethno}$) en *–ský* (Horecký et al. 1989 ; Sekvent 2005 ; Makišová 2006) dans le cadre lexématique de la morphologie (Aronoff 1994 ; Fradin 2003). Ces adjectifs sont construits, à première vue, sur noms propres (NPr) de lieu (1)[1] :

(1) LIBANON$_{NPr}$ > *–ský* LIBANONSKÝ$_{ATopo/Ethno}$
Liban Liban-AZR
'libanais'

Le $A_{Topo/Ethno}$ dénominal comme en (1), du fait qu'il est issu d'une règle 'non-marquée' (Nábělková 1993), est considéré comme un adjectif relationnel (Bartning 1980 ; Mélis-Puchulu 1991 ; Roché 2006 ; Fradin 2008) : son interprétation précise dépend du nom auquel il s'applique dans un syntagme, i.e. le nom recteur (Nr). En (2), le Nr nous conduit à interpréter le $A_{Topo/Ethno}$ comme relatif à un territoire géographique, représenté par le nom de base (Nb) *Liban*. Le référent de Nr est localisé dans celui de Nb.

(2) *libanon-sk-é* *hor-y*
Liban-AZR-FLX:PL-F-NOM montagne-FLX:PL-F-NOM
'montagnes de Liban'

Le $A_{Topo/Ethno}$ en slovaque, ainsi que dans d'autres langues telles que le français (Kim 1995 ; Nowakowska 2006 ; Roché 2006, 2008), est susceptible de référer également aux habitants d'un lieu, en fonction du Nr. Dans l'exemple (3), le Nr dénote une propriété humaine :

(3) *libanon-sk-á* *pohostinnosť*
Liban-AZR-FLX:SG-F-NOM hospitalité-FLX:SG-F-NOM
'hospitalité des Libanais'

L'interprétation exclusivement toponymique ou ethnique est observée dans les exemples (4)-(5). Concernant l'énoncé en (4), il n'existe pas de nom d'habitant morphologiquement apparenté au NPr *Seine*. La seule interprétation de l'adjectif en *–ský* est la relation au lieu.

(4) *sein-sk-e* *breh-y*
Seine-AZR-FLX:PL-M-NOM rive-FLX:PL-M-NOM
'rives de la Seine'

[1] Dans les gloses, nous utilisons les conventions de Leipzig et les abréviations suivant Fradin (2008) : AZR=adjectiviseur (i.e. suffixe constructionnel qui marque la catégorie adjectivale), NZR=nominaliseur (i.e. suffixe constructionnel qui marque les noms), FLX=suffixe flexionnel.

Symétriquement, certains noms d'ethnies ne sont pas morphologiquement reliés à un NPr de lieu, pour des raisons extralinguistiques. En conséquence, l'adjectif en (5) a une interprétation uniquement ethnique.

(5) *slovan-sk-á* *keramik-a*
Slave-AZR-FLX:SG-F-NOM céramique-FLX:SG-F-NOM
'céramique slave (=produite par les membres de l'ethnie des Slaves)'

Nous proposons une analyse morphologique qui rend compte de l'apparente polyréférentialité de ces $A_{Topo/Ethno}$ ainsi que de leurs particularités formelles.

2. Problème. Parmi les 4 583 adjectifs en *–ský* provenant du SNK[2], 1 695 sont considérés comme $A_{Topo/Ethno}$ puisqu'ils sont sémantiquement apparentés à des noms propres de lieu, i.e. toponymes ($N_{Pr\ Topo}$) ou à des noms d'habitants, i.e. ethnonymes (N_{Ethno}), cf. (2)-(3). Dans cette étude sont également pris en compte les A_{Topo} et les A_{Ethno}, cf. (4)-(5). Les 1 695 $A_{Topo/Ethno}$ se répartissent en 6 groupes selon le type formel de la base, présentés dans le Tableau 1 :

	$N_{Pr\ Topo}$	**N_{Ethno}**	**$A_{Topo/Ethno}$**	**%**
CAS1	$PARÍŽ_{NPr}$ Paris	$PARÍŽAN_{N}$ Paris-NZR	$PARÍŽSKY_{A}$ Paris-AZR	81%
CAS2	$DÁNSKO_{NPr}$ Danemark-NZR	$DÁN_{N}$ Danois	$DÁNSKY_{A}$ Danois-AZR	2%
CAS3	$MODRA_{NPr}$ Modra (ville slovaque)	$MODRAN_{N}$ Modra-NZR	$MODRANSKÝ_{A}$ Modra-NZR-AZR	2,5%
CAS4	$LOTRINSKO_{NPr}$ Lorraine-NZR	$LOTRINČAN_{N}$ Lorraine-NZR	$LOTRINSKÝ_{A}$ Lorraine-AZR	3%
CAS5	$DUNAJ_{NPr}$ Danube	-	$DUNAJSKÝ_{A}$ Danube-AZR	8,5%
CAS6	-	$KELT_{N}$ Celte	$KELTSKÝ_{A}$ Celte-AZR	3%

Tableau 1. Toponymes, ethnonymes et adjectifs correspondants en slovaque.

D'un point de vue formel, la base est immédiatement identifiable soit comme un $N_{Pr\ Topo}$ (CAS1, CAS5), soit comme un N_{Ethno} (CAS2-3, CAS6). Le CAS4 pose quelques problèmes d'analyse. Nous nous demandons si l'on peut envisager un seul mécanisme prenant en compte les diversités illustrées dans le Tableau 1, et prédisant la formation de 1 695 $A_{Topo/Ethno}$, ainsi que leur interprétation.

3. Hypothèses. Contrairement aux CAS5-6, où l'identification de la base ne pose aucun problème, 2 hypothèses sont possibles pour les données relevant du CAS1-4 :

H1 Deux règles distinctes servent à former les $A_{Topo/Ethno}$: l'une sélectionne le $N_{Pr\ Topo}$ et l'autre s'applique sur le N_{Ethno}.

H2 Les $A_{Topo/Ethno}$ sont toujours construits sur le $N_{Pr\ Topo}$ mais le radical est, si nécessaire, emprunté à la forme du N_{Ethno}, cf. Booij (1997), (2002).

L'hypothèse **H1** est vite écartée, car elle ne peut pas être appliquée aux $A_{Topo/Ethno}$ relevant du CAS4. L'hypothèse **H2** inspirée de Booij (1997 : 45-48) semble être plus plausible. G. Booij analyse les données du néerlandais (6)-(8), comparables aux CAS2-4 slovaques, comme des cas d'allomorphie du radical du toponyme de base :

[2] i.e. Slovenský Národný Korpus (Corpus national slovaque). SNK contient 255 012 lexèmes adjectivaux qui correspondent à ca. 37 millions mots-formes. http://korpus.juls.savba.sk.

(6)	a. RUSLAND$_{\text{NPr Topo}}$ Russie-NZR	b. RUS$_{\text{N Ethno}}$ Russe	c. RUSSISCH$_{\text{A Topo/Ethno}}$ Russe-AZR	(cf. CAS2)
(7)	a. MOSCOU$_{\text{NPr Topo}}$ Moscou	b. MOSKOVIET$_{\text{N Ethno}}$ Moscou-NZR	c. MOSKOVIETISCH$_{\text{A Topo/Ethno}}$ Moscou-NZR-AZR	(cf. CAS3)
(8)	a. VLISSINGEN$_{\text{NPr Topo}}$ Flessingue-NZR	b. VLISSINGER$_{\text{N Ethno}}$ Flessingue-NZR	c. VLISSINGS$_{\text{A Topo/Ethno}}$ Flessingue-AZR	(cf. CAS4)

En (6), le $N_{Pr\ Topo}$ de base (6a) possède, suivant Booij (1997), (2002), un radical court qui est formellement identique à celui du N_{Ethno} (6b). C'est ce radical qui est sélectionné par la règle formant le $A_{Topo/Ethno}$ (6c). Pour les cas comme (7), le $N_{Pr\ Topo}$ disposerait d'un radical long, s'identifiant encore une fois avec le radical du N_{Ethno} (7b). Enfin, les $N_{Pr\ Topo}$ en (8a) possèdent un radical court, dépourvu de la séquence graphiquement réalisée comme *–en*, et qui est sélectionné dans la formation du N_{Ethno} (8b) et de l'$A_{Topo/Ethno}$ (8c). *« The formal basis for the coining of such adjectives is not the corresponding name of country, but the inhabitant name, although semantically the adjectives express the notion 'relating to the country', and not 'relating to the inhabitant' »* (Booij 2002 : 181).
Or, l'adaptation au slovaque de la proposition de Booij (1997), (2002) ne permet pas de rendre compte de la potentielle détermination du sens adjectival lors de la combinaison de $A_{Topo/Ethno}$ avec un Nr, ni d'envisager le CAS6.

4. Proposition. Nous proposons une structure morphologique partiellement non-orientée, qui relie 3 lexèmes (L1-3), et permet d'accéder théoriquement pour le $A_{Topo/Ethno}$ (L3) à la sémantique du $N_{Pr\ Topo}$ (L1), mais aussi à celle du N_{Ethno} (L2) pour les CAS1-4. Ce modèle est motivé lexicalement, car il permet de rendre compte des adjectifs exclusivement toponymiques (CAS5) ou ethniques (CAS6).

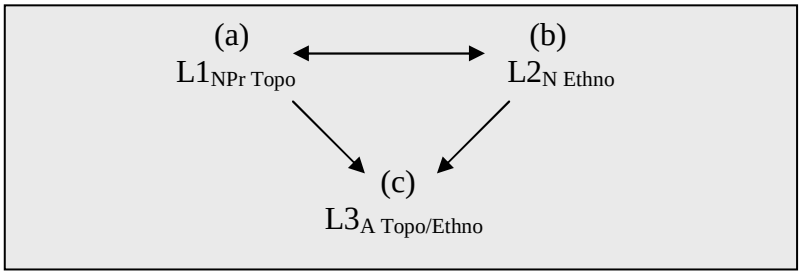


Figure 1. Modèle ternaire pour l'analyse de la construction des $A_{Topo/Ethno}$.

Le schéma à structure ternaire s'applique de la manière suivante : la flèche à double sens met en relation **(a)** et **(b)** parce que $N_{Pr\ Topo}$ et N_{Ethno} sont nécessairement reliés morphologiquement dès lors que les deux lexèmes sont attestés (CAS1-4). La relation morphologique est non-orientée, ce qui rend compte de l'existence des N_{Ethno} construits formellement sur le $N_{Pr\ Topo}$ (CAS1), ou l'inverse (CAS2-3), ou des formations qui évoquent une troncation mutuelle (CAS4). Les relations vers le **(c)** indiquent que les $A_{Topo/Ethno}$ sont dérivés soit de $N_{Pr\ Topo}$ (CAS1, CAS5), soit de N_{Ethno} (CAS2-3, CAS6), suivant la structure formelle la plus adéquate. La forme de l'adjectif relevant du CAS4 valide ces faits : la base formelle est indifféremment le $N_{Pr\ Topo}$ ou le N_{Ethno}. La forme choisie pour le radical, dans les CAS1-4, ne présage en rien de l'interprétation sémantique [+LOC], [+HUM] de l'adjectif. Ce ne sera qu'au moment de la mise en discours où la valeur du Nr va révéler l'interprétation de la base. Les $A_{Topo/Ethno}$ sont impliqués, avec les Nr, dans au moins 6 types de relations sémantiques que nous avons pu répertorier, et qui seront présentées dans notre exposé. À titre d'exemple, il s'agit de la relation de 'provenance' ou 'inclusion spatiale' pour l'interprétation [+LOC], ou de la 'possession d'une propriété' ou 'production d'artefacts' pour l'interprétation [+HUM]. Chacune de ces relations met clairement en jeu la nature toponymique ou ethnique de la base.

En résumé, le $A_{Topo/Ethno}$ accède à la sémantique de la base qui se réalise comme le toponyme ou l'ethnonyme (CAS1-4), uniquement comme le toponyme (CAS5), et uniquement comme l'ethnonyme (CAS6).

5. Conclusion. L'analyse proposée pour les relations morphologiques entre $N_{Pr\ Topo}$, N_{Ethno} et $A_{Topo/Ethno}$ dans le but de la modélisation du sens construit, consiste à prévoir un modèle ternaire reliant les 3 lexèmes en jeu. Cette analyse, qui est applicable à l'ensemble des 1 695 $A_{Topo/Ethno}$ slovaques examinés, a au moins deux avantages.

D'une part, le schéma ternaire sert à formaliser la contrainte d'instanciation de $N_{Pr\ Topo}$, et/ou N_{Ethno}. En effet, la valeur des $A_{Topo/Ethno}$ en discours justifie la relation entre 3 lexèmes, plutôt que l'emprunt par le $N_{Pr\ Topo}$ du radical du N_{Ethno} suivant Booij (1997), (2002) puisque cette solution prive l'adjectif de la possibilité d'hériter de la sémantique [+HUM] dans les CAS1-4. Nous allons démontrer que le type de Nr va impliquer le $A_{Topo/Ethno}$ dans l'une des relations sémantiques possibles, ce qui va révéler l'interprétation [+LOC] ou [+HUM] de la base.

D'autre part, le schéma proposé devrait permettre de modéliser les différentes options qui s'offrent aux locuteurs pour construire les $A_{Topo/Ethno}$. Pour cela, il est nécessaire de prouver le bien-fondé de notre hypothèse formelle par une expérience, dont l'objectif sera d'examiner la stratégie de formation d'un nouveau $A_{Topo/Ethno}$ de la part des locuteurs.

Références bibliographiques

ARONOFF, M. (1994). *Morphology by Itself*. Cambridge : MIT Press.

BARTNING, I. (1980). *Remarques sur la syntaxe et la sémantique des pseudo-adjectifs dénominaux en français*. Stockholm : Almqvist & Wiksell International.

BOOIJ, G. (1997). Autonomous morphology and paradigmatic relations. BOOIJ, G., VAN MARLE, J. (éds.). *Yearbook of Morphology*, 1996. Dordrecht/Boston : Kluwer, 35-53.

BOOIJ, G. (2002). *The Morphology of Dutch*. Oxford : OUP.

FRADIN, B. (2003). *Nouvelles approches en morphologie*. Paris : PUF.

FRADIN, B. (2008). On the semantics of denominal adjectives. RALLI, A. et al. (éds.). *Morphology and Dialectology. On-line proceedings of MMM6.* Ithaca, 27-30 Sept 2007, 84-98.
http://www.philology.upatras.gr/LMGD/el/research/downloads/MMM6_Proceedings.pdf

HORECKÝ, J. et al. (1989). *Dynamika slovnej zásoby súčasnej slovenčiny*. Bratislava : Veda.

KIM, M. S. (1995). Sur les adjectifs de relation dérivés de noms propres. *Le français moderne*, 1, 42-54.

LECOLLE, M. (2007). Polysignificance du toponyme, historicité du sens et interprétation en corpus. Les cas d'*Outreau*. *Corpus*, 6, 101-125.

MAKIŠOVÁ, A. (2006). *Prídavné mená v slovenčine a srbčine*. Thèse de Doctorat. Université de Novi Sad (Serbie), Faculté des Lettres.

MÉLIS-PUCHULU, A. (1991). Les adjectifs dénominaux : des adjectifs de « relation ». *Lexique*, 10, 33-60.

NÁBĚLKOVÁ, M. (1993). *Vzťahové adjektíva v slovenčine. Funkčno-sémantická analýza desubstantívnych derivátov*. Bratislava : Veda.

NOWAKOWSKA, M. (2006). Les adjectifs ethniques sont-ils des adjectifs de relation ? *Neophilologica*, 18, 59-69.

ROCHÉ, M. (2006). Comment les adjectifs sont sémantiquement construits. *Cahiers de grammaire*, 30, 373-387.

ROCHÉ, M. (2008). Structuration du lexique et principe d'économie : le cas des ethniques. DURAND, J. et al. (éds.). *Congrès Mondial de Linguistique Française - CMLF 2008*. Paris : ILF, 1 559-1 573.

SEKVENT, K. (2005). Morfematické aspekty slovenských adjektív na *–sky/–ský* odvodených od francúzskych zemepisných názvov. *Slovenská reč*, 70, 4, 228-234.

Louise ESHER
Balliol College & Research Centre for Romance Linguistics, University of Oxford

La morphologie autonome l'est-elle toujours ?

Objectif de l'étude

Dans cette communication j'étudie les rapports formels ainsi que fonctionnels qui existent entre le futur et le conditionnel dans les parlers occitans, et les conséquences que les rapports constatés peuvent avoir pour la notion théorique de *morphome* proposée par Aronoff (1994).

Le concept du morphome

En réponse à des régularités formelles à l'intérieur du paradigme qui s'avèrent systématiques mais qui ne se laissent pas expliquer par d'éventuelles motivations sémantique, syntaxique ou bien phonologique, Aronoff pose l'existence d'un niveau abstrait de structure linguistique qui serait celui de la morphologie autonome. Selon le schéma établi par Aronoff, le lien entre traits morphosyntaxiques et réalisation phonologique ne serait pas direct, mais passerait par ce niveau indépendant, dit morphomique. L'hypothèse du niveau morphomique permet de rendre compte des régularités formelles et systématiques sans motivation autre que morphologique, qui semblent pourtant avoir une réalité psychologique pour le locuteur.

Le rôle de la divergence fonctionnelle dans l'émergence d'un morphome

La notion de morphome a été reprise notamment par Maiden (ex. 2005) dans le cadre de ses travaux sur la morphologie historique des langues romanes. Parmi les schémas de distribution morphomiques identifiés par Maiden figure le morphome baptisé PYTA (acronyme du terme *perfecto y tiempos afines* qui sert à désigner en castillan l'ensemble de formes concernées), qui comprend essentiellement les réflexes des formes perfectives en latin.

Le système verbal latin est fondé sur une opposition aspectuelle — fonctionnelle, donc — entre perfectif et imperfectif, qui est diagrammée par une opposition formelle : pour un lexème donné, les formes à valeur aspectuelle imperfective se construisent sur une base alors que les formes à valeur aspectuelle perfective se construisent sur une base différente. Il y a donc une corrélation assez nette entre forme et fonction.

Or dans les langues romanes modernes on retrouve des séries de formes manifestement issues de formes perfectives latines, mais qui sont maintenant associées à des fonctions divergentes. Dans la plupart des parlers occitans, les formes perfectives latines ont donné le prétérit, l'imparfait du subjonctif, et le participe passé, tandis que certains parlers présentent aussi un conditionnel issu du plus-que-parfait de l'indicatif. Pourtant, malgré cette divergence fonctionnelle massive, les formes issues de formes perfectives en latin conservent presque toujours leur identité formelle : prétérit et imparfait du subjonctif présentent de façon systématique la même base, qui est dans la plupart des cas partagée par le participe passé.

Quoiqu'il ne constitue pas une classe naturelle, cet ensemble paraît avoir une forte réalité psychologique pour le locuteur ; l'hypothèse d'un schéma de distribution présent au niveau morphomique permet d'en rendre compte.

Le cas du futur et du conditionnel

Vu le rôle joué par la divergence fonctionnelle dans l'émergence du morphome PYTA, il serait légitime de s'interroger sur le statut du futur et du conditionnel synthétiques des langues romanes. En effet, le futur et le conditionnel synthétiques proviennent de deux périphrases parallèles en latin, soit l'infinitif + le présent de l'indicatif du verbe HABERE 'avoir' dans le cas du futur, et l'infinitif + l'imparfait de l'indicatif de HABERE dans le cas du conditionnel. Il y a donc à l'origine identité formelle entre futur et conditionnel, ainsi qu'identité fonctionnelle car la périphrase infinitif + HABEBAM sert initialement de futur dans le passé. Par la suite, on peut remarquer une certaine divergence fonctionnelle entre futur et conditionnel, en ceci que ce dernier acquiert des valeurs modales que le futur ne partage pas, et qui tendent à primer sur sa fonction originale de futur dans le passé, même si celui-ci peut être conservé.

L'identité fonctionnelle entre futur et conditionnel se retrouve par conséquent affaiblie, au point qu'on pourrait envisager plusieurs types d'évolution. D'un côté, la divergence fonctionnelle pourrait conduire à une divergence formelle, ce qui préserverait la corrélation forme-fonction. D'un autre côté, étant donné que dans le cas du futur et conditionnel on a affaire à une corrélation initiale forme-fonction suivi par une certaine divergence fonctionnelle — évolution qui rappelle celle du morphome PYTA — on pourrait tout aussi bien s'attendre à ce que l'identité formelle entre futur et conditionnel soit conservée en tant que morphome.

Pour la plupart des langues romanes, c'est cette dernière hypothèse qui semblerait primer : malgré leur divergence fonctionnelle, le futur et le conditionnel continuent à présenter la même base (voir ex. Maiden 2010:464-6). Le cas de l'italien est particulièrement révélateur à cet égard : depuis au moins le XVIIe siècle, le conditionnel synthétique ne sert plus de futur dans le passé, pourtant l'identité formelle avec le futur est conservée. On pourrait en conséquence poser l'existence d'un morphome qui regrouperait le futur et le conditionnel.

Une telle hypothèse se heurte cependant à plusieurs problèmes qu'il convient d'examiner de plus près.

Premièrement se pose la question de la valeur sémantique du futur et du conditionnel. Si la valeur du conditionnel dans les parlers occitans est de nos jours essentiellement modale, il n'en demeure pas moins capable de servir de futur dans le passé. Par surcroît, la catégorie linguistique du conditionnel a parfois été analysée comme comportant de façon générale un élément de futurité (cf. ex. Iatridou 2000). Dans ces conditions, on peut se demander si la conservation de l'identité formelle ne relèverait pas plutôt d'une identité sémantique résiduelle, ce qui pourrait infirmer l'hypothèse d'un cas de morphologie autonome.

Il faut également considérer la question des asymétries constatées entre la base du futur et celle du conditionnel dans les parlers occitans, à savoir des cas où l'identité formelle n'est pas maintenue. Si les parlers occitans semblent quasiment seuls parmi les langues romanes à présenter des asymétries, celles-ci sont en revanche attestées sur la plupart de l'aire linguistique occitane et ne sauraient être mises de côté comme accidents de la phonologie

ou d'une description inadéquate. L'hypothèse d'une divergence fonctionnelle qui motiverait une divergence formelle pourrait paraître attrayante comme explication des asymétries ; mais elle se heurte à une difficulté majeure, à savoir le fait que les asymétries sont partout minoritaires, le schéma majoritaire de chaque parler demeurant celui d'identité formelle entre futur et conditionnel, ce dont il faudrait alors fournir une explication.

Il semblerait difficile de soutenir en même temps que la présence d'une valeur sémantique commune assure une identité formelle entre le futur et le conditionnel *et* que la présence de valeurs sémantiques non partagées conduise à une différentiation formelle du futur et du conditionnel.

Une solution possible

Il convient pourtant de se rappeler que les exemples présentés par Aronoff sont choisis à dessein comme étant 'the clearest examples of the purely morphological' (1994:167), et qu'il est dit de façon explicite dans la présentation initiale du niveau morphomique que

> not all mappings from syntax to morphology are necessarily so indirect. It is possible to have a singleton morphosyntactic set mapped onto a singleton morphomic set, which itself is mapped onto a singleton morphophonological set [...]. It is also possible to have cases intermediate in complexity between the two extremes. (1994:25).

Cette mention ouvre la voie vers une conception plus nuancée de la morphologie autonome, qui permettrait l'interaction du niveau morphomique avec les niveaux sémantique et phonologique, et qui accessoirement permettrait une analyse plus satisfaisante du cas futur-et-conditionnel.

Nous avons vu que le futur et le conditionnel ne semblent pas entièrement dénués de valeurs sémantiques communes ; cependant ces valeurs partagées ne suffisent pas à garantir la conservation de l'identité formelle entre futur et conditionnel en occitan. De même, les valeurs sémantiques divergentes n'entraînent pas forcément une différentiation du futur et du conditionnel sur le plan formel. Ce dernier point est d'autant plus frappant en italien, où l'identité formelle entre futur et conditionnel persiste de façon systématique malgré un rapport sémantique beaucoup plus atténué que ce qui existe en occitan. Une approche qui ne ferait appel qu'à des facteurs sémantiques ou bien qu'à des facteurs morphologiques s'avérerait donc insuffisante pour rendre compte des faits observés ; c'est en tenant compte des deux plans qu'on peut arriver à une analyse plausible.

Conclusion

Cette étude part du concept théorique du morphome et des exemples canoniques de ce dernier pour aborder l'évolution du futur et du conditionnel des langues romanes occidentales et en particulier des parlers occitans. Or, si le futur et le conditionnel en occitan se comportent en général comme un ensemble présentant de nombreuses similitudes avec le morphome canonique, il faut noter deux différences majeures : les deux tiroirs ne sont pas entièrement sans cohérence sémantique, et la cohérence formelle du futur et conditionnel peut être brisée.

Les données ici étudiées invitent à une conception plus nuancée du morphome — il n'est pas indispensable d'écarter toute motivation fonctionnelle pour que la morphologie en tant que telle joue un rôle —, tout aussi bien qu'à une approche fonctionnelle plus rigoureuse — une motivation fonctionnelle ne suffit pas forcément à elle seule.

Réferences

Aronoff, Mark (1994), *Morphology by Itself; Stems and Inflectional Classes*, Cambridge: MIT

Iatridou, Sabine (2000), 'The Grammatical Ingredients of Counterfactuality', *Linguistic Inquiry* 31, pp. 231-270

Maiden, Martin (2005), 'Morphological autonomy and diachrony', *Yearbook of Morphology* 2004, pp. 137-75

Maiden, Martin. 2010. 'Morphophonological innovation'. *The Cambridge History of the Romance Languages*, ed. Martin Maiden, J. C. Smith & Adam Ledgeway. Cambridge: CUP

Inflection-Inflection vs. Inflection-Derivation competition in loanword integration

Francesco Gardani
Vienna University of Economics and Business

The goal of my contribution is to enlighten the mechanisms of rules competition which underlie and govern the process of loanword integration and, in particular, of morphological integration. More specifically, my investigation aims to show that competition occurs both within the inflectional module of morphology and between inflection and derivation. As a case-in-point, I examine instances of integration of loanwords into the nominal systems of Latin and of Old Italian.

In the literature, loanword integration is mostly treated from the mere perspective of phonological approximation (e.g., Chang in press, Dohlus 2005, LaCharité & Paradis 2005, Paradis & LaCharité 2008) and of individual lexical analogies (e.g., André 1971, Biville 1990, 1995). Even when morphological integration has been dealt with (e.g., Winford 2003: 48-51, Wohlgemuth 2009, Haspelmath & Tadmor 2009), neither the issue of lexical transfer has been tackled in light of the decisive factor of productivity, nor has productivity-determined morphological rules competition been investigated thoroughly.

Within the framework of Natural Morphology and particularly within its model of inflectional morphology (Dressler 1997, 2003), productivity is understood as grammatical productivity and envisaged as a "constitutive primitive property of inflectional patterns" (Dressler 2003: 31). As a prototypical property of rules, productivity steers the process of inflectional integration and determines its outcomes.

According to psycho- and neurolinguistic models such as the morphological race model (cf. Baayen, Dijkstra & Schreuder 1997), inflectional morphology is processed on two parallel paths, namely prelexical morphological parsing and direct access route based on stored full-form representations for morphologically complex words. Prototypically, whereas unproductive patterns are assigned the status of lexical storage and belong to static morphology, productive patterns (categories, rules and classes) are assigned the status of symbolic rule mechanism and form the core of dynamic morphology (cf. Kilani-Schoch & Dressler 2005: 118-121). As it has been observed (cf. Dressler 1997: 14-15, 2003: 51-52), when more than one productive pattern applies to the same domain, this pattern rivalry weakens the patterns' competitivity with lexical access.

In this paper, I go beyond the morphological race model and, adopting a competing motivations approach (cf. Du Bois 1985, Bates & MacWhinney 1989, Haspelmath 1999), I show not only that in the process of morphological integration different rules may compete for the same input, but also that this competition may be carried out at different *loci* of the morphological processing. Two scenarios are conceivable: either two or more productive inflectional patterns compete with each other, that is, competition occurs intra-modularly, or an inflectional pattern competes with a productive derivational one, that is, competition occurs inter-modularly. I will call the first scenario intra-modular or Inflection-Inflection (henceforth, I-I) competition and the second scenario inter-modular or Inflection-Derivation (henceforth, I-D) competition.

In Gardani (forthc.) the productivity of the nominal inflection of Latin from the beginnings of its documentation to Early Medieval Latin and of Italian from its emergence up

to 1400 has been measured on the basis of five hierarchical criteria which include the investigation of loanword integration, of indigenous conversions, and of class shift occurrences. The data on loanword integration are drawn from the contact settings of Latin with Etruscan and Ancient Greek, on the one hand, and of Old Italian with Germanic languages, Arabic, Byzantine Greek and Old French, on the other.

As far as I-I competition is concerned, the analysis of the morphological variants reveals not only different grades of morphological integration, e.g., in Latin the Graecisms *stacta -ae* (F) 'gum-resin' from στακτή -ῆς (F) and *poeta -ae* (M) 'poet' from ποιητής -οῦ (M) vs. the corresponding less integrated Graecising forms *stacte -es* (F) and *poetes -ae* (M), but also indicates competition between different inflectional classes. Moreover, there are three sub-scenarios of intra-modular competition:

Sub-scenario 1a: Competition of two or more classes which display different degrees of productivity, as shown in the examples (1) to (3):

(1) λαμπάς -άδος (F) > lampada -ae (F) (Plautus) vs. lampas -adis (F) (Plautus) 'torch'

(2) δογμα -ατος (N) > dogma -ae (F) (Laberius) vs. dogma -atis (N) (Cicero) 'a doctrin'

(3) στατήρ -ῆρος (M) > statera -ae (F) (Varro) vs. stater -eris[1] (M) (Hieronimus) 'steelyard'

Sub-scenario 1b: Competition of two or more classes with an identical degree of productivity, one of which has phonological and morphological properties that are incompatible with the original input forms, as enlightened in the examples (4) and (5):

(4) μεσπίλη -ης (F) > mespila -ae (F) (Plinius) vs. mespilus -i (F) (Plinius) 'a medlar tree'

(5) κῶλον -ου (N) > colum -i (N) (Plinius) vs. cola -ae (F) (Fronto) 'large intestine'

Sub-scenario 1c: Competition of two or more classes which display an identical degree of productivity and both have phonological and morphological properties which are incompatible with the original input forms, as in the examples (6) to (9):

(6) διαιτάρχης -ου (M) > diaetarcha –ae (M) (CIL VI, 8645) vs. diaetarchus -i (M) (CIL VI, 5187) 'a servant'

(7) κῆτος -εος, τό > cetus -i (M) (Plautus) vs. cetum -i (N) (Plinius) 'a large sea-animal'

(8) τάπης -ητος (M) > tapetum -i (N) (Livius Andronicus) vs. tapete -is (N) (Ennius) 'a woollen cloth'

(9) γαυσάπης -ου (M) > gausape -is (N) (Lucilius) vs. gausapa -ae (F) (Varro) vs. gausapum -i (N) (Ovidius) 'cloth of woollen frieze'

As far as I-D competition is concerned, derivational morphology may enter in conflict with an inflectional rule and compete with it on the attribution of an inflectional class, that is, a productive derivational suffix of the receiving language may exert pressure on the process of integration of a loanword into an inflectional class. The alluded influence may be due to

[1] In examples (1)-(3) the a-stems display full productivity, whereas the consonantal stems display mid-low productivity.

partial, superficial phonological similarities between the input form and forms of the receiving language, whereby these similarities have not an actual, once-character such as in surface analogy (e.g. Italian *marzapane* 'marzipan', from Arabic *marṭabān* on the model of *pane* 'bread') but display a serial rule-driven character, as in the following examples:

(10) Arabic *lāṭūn* > ottono (M) (Guinizzelli, a. 1276) vs. ottone (M) (Doc. fior., 1262-75) 'brass'

(11) Arabic *quṭn* > cotono vs. cotone (both M) (both in Doc. sen., 1281-82) 'cotton'

(12) Old French *destrier* > destriero (M) (Guittone, a. 1294) vs. destriere (M) (Ruggieri Apugliese, XIII m.) 'charger'

As evidenced by the examples (10)-(12), two different processes are at work here: On the one hand, the productive suffixes *-one* and *-iere* of the receiving language attract the loan-noun into their inflectional schema; on the other hand, the fully productive class *libro -i* applies. The product of this competition are allomorphic forms, whose long-term success can be determined only diachronically. The role played by the derivational suffixes involved sets instances such as *cotone* or *destriere* apart from cases in which loanword integration occurs by the explicit use of a derivational suffix, that is, cases in which the derivational suffix which is used to integrate the word morphologically and also determines its inflectional class in the receiving language, adds new semantic derivational meaning to the word borrowed. This is best exemplified by the lexeme *harpago -onis* (M) in Latin, borrowed from Ancient Greek ἁρπαγή -ῆς (F) 'hook', with the figurative augmentative-pejorative meaning of 'pilferer, rapacious person' in its earliest attestation (Plautus).

By applying the competing motivations approach to the analysis of inflectional productivity conducted on a well-documented historical corpus, this paper contributes to a better understanding of the dynamics underlying morphological integration and inflectional class attribution.

References:

André, Jacques. 1971. *Emprunts et suffixes nominaux en latin*. Genève: Droz.

Bates, Elisabeth & Brian MacWhinney. 1987. Competition, variation, and language learning. In Brian MacWhinney (ed.), *Mechanisms of language acquisition*, 157–193. London: Lawrence Erlbaum.

Baayen, R. Harald, Teun Dijkstra & Robert Schreuder. 1997. Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language* 37. 94–117.

Biville, Frédérique. 1990. *Les emprunts du latin au grec. Approche phonétique,* tome I. Louvain, Paris: Peeters.

Biville, Frédérique. 1995. *Les emprunts du latin au grec. Approche phonétique,* tome II. Louvain, Paris: Peeters.

Chang, Charles B. (in press). Phonetics vs. phonology in loanword adaptation: Revisiting the role of the bilingual. In Alex Bratkievich, Daniel Bruhn, Amy M. Campbell, Ramón Escamilla, Lindsey Newbold, and Russell Rhodes (eds.), *Proceedings of the 34th annual meeting of the Berkeley Linguistics Society: general session and parasession on information structure*. Berkeley. CA: Berkeley Linguistics Society.

Dohlus, Katrin. 2005. Phonetics or phonology: Asymmetries in loanword adaptations – French and German mid front rounded vowels in Japanese. *ZAS Papers in Linguistics* 42. 117-135.

Dressler, Wolfgang U. 1997. On productivity and potentiality in inflectional morphology. In *CLASNET Working Papers* (Université de Montréal), 7. 3-22.

Dressler, Wolfgang U. 2003. Degrees of grammatical productivity in inflectional morphology. In *Italian Journal of Linguistics* 15(1). 31-62.

Du Bois, John W. 1985. Competing motivations. In John Haiman (ed.), *Iconicity in syntax*, 343-366. Amsterdam: Benjamins.

Gardani, Francesco (forthc.). *Dynamics of morphological productivity. Noun inflection from Archaic Latin to Old Italian.* [Submitted to press]

Haspelmath, Martin. 1999. Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft* 18(2). 180-205.

Haspelmath, Martin & Uri Tadmor (eds.). 2009. *Loanwords in the world's languages. A comparative handbook*. Berlin: de Gruyter.

Kilani-Schoch, Marianne & Wolfgang U. Dressler. 2005. *Morphologie naturelle et flexion du verbe français*. Tübingen: Narr.

LaCharité, Darlene & Carole Paradis. 2005. Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry* 36(2). 223-258.

Paradis, Carole & Darlene LaCharité. 2008. Apparent phonetic approximation: English loanwords in Old Quebec French. *Journal of Linguistics* 44(1). 87-128.

Schultink, Henk. 1961. Produktiviteit als morfologisch fenomeen. *Forum der Letteren* 2. 110-125.

Winford, Donald. 2003. *An introduction to contact linguistics*. Malden, Mass.: Blackwell.

Wohlgemuth, Jan. 2009. *A typology of verbal borrowings.* Berlin: de Gruyter.

All by itself - why there are no portmanteaus in Uralic

Doreen Georgi & Larissa Kröhnert (University of Leipzig)

Claim: We argue that there are no portmanteau morphemes in the definite conjugation of the Uralic languages. There has been a long debate in the Uralic literature on this topic including Collinder 1957, Raun 1988, Salminen 1997, Künnap 1999, Körtvely 2005, Aranovich 2007. All these authors argue for the existence of portmanteaus in a a variety of Uralic languages, Abondolo 1982 being the first who denies this. Recently, the discussion on the status of portmanteaus in general has arisen anew. Caha 2008, 2010 claims that Portmanteaus must exist, whereas Trommer 2007 shows that there are alternative ways of analyzing potential portmanteaus by independently existing concepts like contextual allomorphy. We claim that a portmanteau analysis in the Uralic languages is not tenable for empirical reasons. Hence, the existence of portmanteaus is an illusion which people gained because of the complex interaction of marker composition, allomorphy and phonological interface conditions.

Background: With the term "portmanteau" we refer to a single unanalyzable morpheme which expresses features of two syntactic heads simultaneously that are realized by two separate morphemes in other contexts. In particular, we concentrate on portmanteau morphemes which express features of both arguments of a transitive verb simultaneously. Portmanteau agreement is systematically ambiguous to Ø-exponence (Trommer 2007): In a realizational morphological framework (e.g. Network Morphology (Corbett & Fraser 1993), Distributed Morphology (Halle & Marantz 1993, Halle & Marantz 1994), Paradigm Function Morphology (Stump 2001), Anderson 1992)a portmanteau M can be respresented as

i) a "real" portmanteau which realizes features of two heads x and y (x bearing subject and y bearing object features) simultaneously: M $\leftrightarrow$ [x][y]

ii) a single marker M which only expresses [x] and which is adjacent to a zero morpheme expressing [y], giving the illusion of a portmanteau on the surface. We argue that the second solution in a generalized version is the correct one for the alleged Uralic "portmanteaus" such that there are no markers M $\leftrightarrow$ [x][y]. The analysis in i) is empirically not tenable.

Uralic languages: In the Uralic languages in column 2 in (1), a transitive verb only agrees with its subject. In the languages listed in column 1 a transitive verb can agree with its subject *and* also with the object (in number and/or person) if the latter is definite (definite conjugation), but not if it is indefinite (indefinite conjugation). For our discussion of portmanteaus as defined above only the languages with subject and object agreement are important (column 1).

(1) *Conjugation patterns in the Uralic language family (cf. Körtvely 2005)*

1		**2**	
subject and object agreement		**only subject agreement**	
Khanty (Ugric)	Hungarian (Ugric)	Saami (Saami)	Mari (Mari)
Mansi (Ugric)	Nenets (Samoyedic)	Finnish (Finnic)	Udmurt (Permian)
Mordva (Ugric)	Enets (Samoyedic)	Estonian (Finnic)	Karelian (Finnic)
Selkup (Samoyedic)	Nganasan (Samoyedic)	Veps (Finnic)	
Mator (Samoyedic)	Kamas (Samoyedic)	Komi (Permian)	

Empirical evidence: We present three arguments against a portmanteau analysis of Uralic: a) There is independent evidence for the segmentation of verbal suffixes into submarkers which encode either only subject or only object features. b) Undoing phonological processes reveals that a marker has a wider distribution than we can see on the surface. c) The distribution of a marker in the indefinite conjugation suggests that it cannot be a portmanteau. We address each of them in turn on the basis of selected examples.

Segmentation: A marker M is not a portmanteau if it can be broken down into smaller submarkers M_1 and M_2 each of which expresses only the features of one argument (M_1 expresses subject and M_2 object features). It can be shown that such an analysis is advantageous and independently motivated in the Uralic languages (for the concept and motivation of this kind of marker segmentation see e.g. Pike 1965, Halle 1992, Noyer 1992, Anderson 1992, Corbett &

Fraser 1993, Halle & Marantz 1993, Wunderlich 1996, Stump 2001, Harbour 2003). The reason is that these submarkers occur as single morphemes in non-verbal paradigms. In general, verbal and nominal inflection take their inflectional exponents from the same set of markers in Uralic, e.g. person/number affixes can be attached to nouns or verbs (= the possessive declension found in all Uralic languages). Assuming that markers with an identical phonolgical form should be analysed as an instance of the same morpheme if possible (cf. the *Syncretism Principle* in Müller 2004) we segment the verbal suffixes into submarkers found in the nominal domain. We illustrate this point with Mordvin. The definite past paradigm is given in (2). The initial /i/ is the past tense marker, the second /i/ is an epenthetic vowel (cf. Abondolo 1982, Zaicz 1988). These two markers do not express ϕ-features of the arguments and thus can ignored for our purposes (they are not represented in (6)).

(2) *Definite past paradigm of Mordvin (Zaicz 1988:199)*:

S \ O	1sg	2sg	3sg	1pl	2pl	3pl
1sg	—	-itinʲ	-ija	—	-idʲizʲ	-inʲ
2sg	-imik	—	-ik	-imizʲ	—	-iʲt
3sg	-imim	-inʲzʲitʲ	-izʲe	-imizʲ	-idʲizʲ	-inʲzʲe
1pl	—	-idizʲ	inʲek	—	-idʲizʲ	-inʲek
2pl	-imizʲ	—	-iŋk	-imizʲ	—	-iŋk
3pl	-imizʲ	-idizʲ	-izʲ	-imizʲ	-idʲizʲ	-izʲ

Comparing the verbal paradigm with the personal pronouns in (3) shows that /m/ encodes 1st, /t/ 2nd person, and /iz/ (reduced to [i] in non-final position) plural. /k/ means 2nd person, too, because it is used to express the 2sg imperative, see (4). In the possessive declension of the noun *kudo* 'house' in (5) we see the same markers found in other paradigms like /m/, but also e.g. /nze, ze/ as a marker for 3rd person (e-o alternation is due to vowel harmony), /n/ for 1st, /nek/ for 1st plural, /nk/ is 2nd plural.

(3) *Personal pronouns, (Zaicz 1988:199, Raun 1988:104)*

1sg	mon
2sg	ton
3sg	son
1pl	min'
2pl	tin'
3pl	sin'

(4) *Imperative (Zaicz 1988:201)*:

a. vano-k = look-2sg = Look! (sg)
b. vano-do = look-2pl = Look! (Pl)

(5) *Possessive declension of nouns (Raun 1988:102)*:

1st sg, one possession	kudo-m
2nd sg, one poss.	kudo-t
3rd sg, one poss	kudo-zo
1st sg, more than one poss.	kudo-n
2nd sg, more than one poss.	kudo-t
3rd sg, more than one poss.	kudo-nzo
1st pl, one/more than one poss.	kudo-nok
2nd pl, one/more than one poss.	kudi-nk
3rd pl, one/more than one poss.	kudi-st

The result of segmentation of the underlying forms is shown in (6). Segmentation in Mordvin leads to a paradigm in which there is not even a potential portmanteau. What is more, the segmented paradigm allows for a simple and elegant generalization about the distribution of person and number markers which cannot be stated if we had unsegmented forms: The verb agrees in person with the object if it is local person, otherwise it references person of the subject. The verb agrees with the subject in number if it is plural, otherwise it references plural of the subject (= hierarchy effects, cf. Béjar 2003). Hence, the system is fully transparent after segmentation. Furthermore, as the smaller segments are also found in nominal paradigms, segmentation leads to syncretism between verbal and nominal paradigms and the analysis can capture the distribution in both (because markers are not specified for a category).

We applied this segmentation strategy to all other Uralic languages as well, but sometimes there still remain alleged portmanteaus after segmentation. However, even those instances turn out to be no "real" portmanteaus on the basis of the observations discussed in the next sections.

Phonological processes: In Mansi (see the partial paradigm in (7)) definite conjugation, the exponents for singular objects is /l/, for dual objects /ay/ and for plural objects /an/ (Keresztes 1988)[1]. This marker is directly attached to the stem and is followed by a subject marker. However, when a 3rd person singular subject acts upon a singular object, the object marker /l/ is missing. Hence, it could be said that /te/ expresses subject features and a singular object simultaneously: /te/ ↔ [+3 +sg]$_A$[+sg]$_P$ (abbreviations: A stands for subject and P for object).

(6) *Segmented definite past paradigm in Mordvin*:

S \ O	1sg	2sg	3sg	1pl	2pl	3pl
1sg	—	-t -n 2 1	-a 1	—	-t -iz 2 pl	-iz -n pl 1
2sg	-m -k 1 2	—	-k 2	-m-iz 1 pl	—	-iz -t pl 2
3sg	-m -m 1 1	-nze -t 3 2	-ze 3	-m -iz 1 pl	-t -iz 2 pl	-iz -nze pl 3
1pl	—	-t -iz 2 pl	-iz -nek pl 1	—	-t -iz 2 pl	-iz -nek pl 1
2pl	-m -iz 1 pl	—	-iz -ŋk pl 2	-m -iz 1 pl	—	-iz -ŋk pl 2
3pl	-m -iz 1 pl	-t -iz 2 pl	-iz -ø pl 3	-m -iz 1 pl	-t -iz 2 pl	-iz -ø pl 3

(7) *Definite non-past paradigm of the verb* uunti- *'to occupy' (Keresztes 1988:402)*:

S \ O	sg	du	pl
1sg	uunti-l-əm	uunti-jaɤ-əm	uunti-jan-əm
2sg	uunti-l-ən	uunti-jaɤ-ən	uunti-jan-ən
3sg	uunti-te	uunti-jaɤ-e	uunti-jan-e
1du	uunti-l-amen	uunti-jaɤ-men	uunti-jan-men
2du	uunti-l-en	uunti-jaɤ-en	uunti-jan-en
3du	uunti-l-en	uunti-jaɤ-en	uunti-jan-en

But there is reason to doubt this analysis: /te/ actually shows up *whenever* the subject is 3rd person singular. On the surface this cannot be seen because /te/ is reduced to /e/ after consonants, an alternation also found in the possessive declension:

(8) a. aaɤi-te 'his/her daughter'; piɤ-e 'his/her son' *(Keresztes 1988:396)*
b. waari-te '(s)he makes it'; waari-jaɤ-e 'she makes them (dual)'

(9) *Underlying forms of 3sg A acting on non-singular objects*:
a. 3sg A -> dual P: uunti-jaɤ-te b. 3sg A -> plural P: uunti-an-te

As dual and plural objects also have /te/ as an exponent, /te/ cannot be specified for the number of the object; it can only be a 3rd person non-plural subject marker and hence not a portmanteau as defined above. What is exceptional in these contexts is then the absence of the object singular marker /l/. Here we propose an account in terms of allomorphy: /l/ has a zero allomorph which is restricted to the context of [+3 +sg] subjects: Ø ↔ [+sg]$_P$ / [+3 +sg]$_A$

Distribution across conjugations: Another objection to a portmanteau analysis of a morpheme X in the definite paradigm is the fact that X also occurs in the indefinite conjugation. As long as we adhere to the Syncretism Priciple, the two occurrences of X in the definite and the indefinite conjugation should be analyzed as the same morpheme. As the indefinite conjugation comprises intransitive verbs which do not have a direct object, X cannot express features of the object. An example of such a pattern can be found in Nenets (the

[1]Insertion of the glide [j] before the object number marker applies in order to avoid a hiatus.

reasoning also applies to Enets, Nganasan, and Hungarian). The three conjugations in Nenets are shown in (10). The subject person marker is attached to the stem and is followed by the subject number marker. In Nenets, as in all Northern Samoyedic languages, the dual object receives a distinct marker which is inserted between stem and subject person marker (*-xøyu-* in Nenets). Moreover, for singular and non-singular objects, the suffixes for 1st and 2nd person subject differ (cf. Körtvély 2005, Salminen 1997). The alternating morphemes **-n** (vs. **-m**) and **-t** (vs. **-r**) occur also in the indefinite and the reflexive paradigm (expressing also only subject features), where no object is given. Consisting of single phonemes, they cannot be segmented any further either and therefore have been treated as portmanteau markers.

(10) *Indicative aorist paradigm of Tundra Nenets, segmented (Körtvély 2005:68)*:

	indefinite	definite			reflexive
O / S	—	sg	du	pl	—
1sg	-tøm	-m-ø	-xøyu-**n**-ø	-**n**-ø	-mt
2sg	-n∼t-ø	-r-ø	-xøyu-**t**-ø	-**t**-ø	-n∼t-ø
3sg	—	-t-(ʲ)a	-xøyu-t-(ʲ)a	-t-(ʲ)a	-t
1du	-**n**-ʲin	-m-ʲin	-xøyu-**n**-ʲin	-**n**-ʲin	-**n**-ʲin
2du	-**t**-ʲin	-r-ʲin	-xøyu-**t**-ʲin	-**t**-ʲin	-**t**-ʲin
3du	-xøn	-t-ʲin	-xøyu-t-ʲin	-t-ʲin	-xøn
1pl	-m-at	-m-at	-xøyu-**n**-at	-**n**-at	-**n**-at
2pl	-**t**-at	-r-at	-xøyu-**t**-at	-**t**-at	-**t**-at
3pl	-t	-t-(ʲ)ot	-xøyu-t(ʲ)ot	-t-(ʲ)on	-t-t

However, the alternation of the subject suffix in the context of a non-singular object can be captured if we postulate a discrete marker for non-singular objects which is subsegmental: [CORONAL,-continuant] ↔ [-SG]. This marker cooccurrs with the discrete subject marker found in the indefinite conjugation and with singular objects, but being subsegmental it attaches to the subject marker and changes two main place features of it, which leads to the illusion that there is a different single marker which expresses features of subject and object simultaneously: The original **-m** turns into an **-n** and the **-r** turns into a **-t** in the context of non-singular (3rd person forms are not affected by this marker because they are already coronal and [-continuant]). This analysis is superior to a portmanteau analysis because it can also explain the regular sound changes from Proto-Uralic to Proto- (Northern) Samoyedic: The original subject person markers of today's Nenets are turned into -m (1st), -r (2nd) and -t by the subsegmental number marker. These three markers resemble the Proto-Uralic personal pronouns ***mV** (1st), ***tV** (2nd) and ***sV** (3rd) from which the verbal markers are derived. Hence, the diachronic change involved the loss of the subsegmental marker in the indefinite and singular object part of the paradigm, suggesting that the marker is indeed an independent morpheme.

A similar alternation which is triggered by subsegmental object number markers which are "fused" with subject morphemes at the phonological interface and seem to produce portmanteaus can be found in the other Northern Samoyedic languages (Nenets and Enets Tundra and Forest variations), as well as in Khanty dialects (e.g. Sherkal).

Summary: The existence of portmanteaus in the definite conjugation in Uralic languages is an illusion because there are empirical reasons against this assumption (segmentation, undone phonological processes, distribution of the markers across conjugations). The relevant morphemes either express only subject or object features, or they are allomorphs of a subject morpheme or subsegmental markers which alter the surface form of a morpheme. These alternative analysis are independently motivated (explains diachronic change, segmentation leads to a uniform analysis of syncretic forms in verbal and nominal paradigms and it allows for generalizations on marker distribution).

Are prefixed units processed and represented like suffixed ones? Toward a hybrid model of morphological processing.

Hélène GIRAUDO (1) & Madeleine VOGA (2)

(1) CNRS & Université de Toulouse
(2) Université Paul-Valéry Montpellier III

The fact that in most languages affixed words are present in a very high proportion leads to the conclusion that morphology constitutes an important variable in word processing. Thirty years of investigation permitted to confirm that morphology intervenes in automatic processes that operate during the very early stages of lexical access, suggesting that morphemes were independently coded or stored somewhere in the mental lexicon. The masked priming paradigm (Forster & Davis, 1984) is the privileged technique used by the psycholinguists to examine the early processes of word recognition. The principle governing this paradigm lies indeed on the transfer of activation from a first processed stimulus (the prime) on the recognition latency of a second stimulus (the target). This activation transfer is admitted to operate on the basis of the shared representations (orthographic/phonological/morphological/semantic) by prime-target pairs. Moreover, given that the prime is presented very briefly (SOAs under 60 ms) and is generally masked (by a string of hash marks), any effect of the prime is considered to be the result of unconscious processes. In the precise case of morphology, many studies manipulated morphologically related words as well as pseudo-words and found systematically very robust positive priming effects: two morphologically related words prime each other across different languages (e.g., Boudelaa & Marslen-Wilson, 2005 in Arabic; Duntildeabeitia, Laka, Perea, & Carreiras, 2009 in Basque; Drews and Zwitserlood, 1995, in both German and Dutch; Frost, Deutsch & Forster, 1997 in Hebrew; Giraudo & Grainger, 2000 in French; Rastle, Davis, Marslen-Wilson & Tyler, 2000 in English) and in experimental settings that include multiple control priming conditions (unrelated but also orthographic/phonological and semantic controls in order to neutralize any interference effect). This general result being established, the question of the nature of morphemic units represented in long-term memory and their precise role within the lexicon remains unanswered.

Two possible hypotheses of representation have been proposed: either morphemic units stand as access units to word representations, or they organize word representations in morphological families. According to the first hypothesis, morphemic units correspond to concrete pieces of words (i.e., stems and affixes, even letter patterns resembling to morphemes but not functioning as such). Complex words are therefore processed by a decomposition mechanism that stripes off the affix in order to isolate the stem. The morphemic nature of the remaining letters is then checked out by the system in order to eliminate any procedural error. Access to word representations (i.e., word forms coded in the orthographic lexicon) can then operate via the pre-activation of the constituent morphemes. This mechanism explains why two morphologically related words prime each other, and this view is broadly shared by numerous authors interpreting their data within a sublexical approach (initially developed by Taft in 1994) that integrates morphemic representations as access units.

According to the second hypothesis, morphemic units are stored at an upper level of processing, at the interface of word and semantic representations. These intermediate units organize the lexicon in morphological families. Subsequently, each time a complex word is encountered, its recognition triggers the activation of all the word forms that can match with it. A competition is then engaged between the pre-activated forms until the right lexical unit reaches its recognition threshold (determined by its surface frequency). However, during the competition phase, competitors send positive activation to their respective base morpheme that in turn, send back positive activation to them. Two morphologically related words prime each other thanks to this

mechanism of co-activation[1]. Following this supralexical theory (Giraudo & Grainger, 2001), morphologically complex words are not "decomposed" in the proper sense (*viz*. following the same procedure described by the sublexical theory) but can trigger the activation of their constituent morphemes.

Regardless of the differences between sublexical and supralexical approaches of morphological processing, they both agree with the idea that separate morphemic units are responsible for priming effects. It is the precise location of these specific units within the architecture of the mental lexicon that specifies their role in word processing (access units vs. organizing units) as well as their own nature.

According to the sublexical view, morphemic units play the role of access units since they correspond to concrete letter clusters (i.e., bound stems, free stems and affixes) that constitute words, insensitive to any grammatical or semantic characteristic of words (i.e., transparency vs. opacity) or to their lexical environment (in terms of orthographic neighbourhood or family size). On the other hand, the supralexical view locates these units above the word forms and before the semantic units. These intermediate units are then supposed to be more abstract than those contained in the words because they have to tolerate form variations induced by the processes of derivation and inflexion (i.e., allomorphy, suppletion, phonological/morphological truncation, haplology). As a consequence, a morphemic unit does not need to exist in the real world in order to be coded in long-term memory but its existence/emergence depends on the interactions between the word and the semantic levels. Such a position also implies that all morphemes of a given language are not necessarily represented within the mental lexicon.

Recent studies explored these issues in order to test the decomposition hypothesis. Using the masked priming paradigm, it was shown that pseudo-derived words (e.g., *corner*) as well as pseudo-derived nonword (e.g., *corning*) primes composed of two existing morphemes were able to produce significant priming effects on the recognition times of their base (e.g., *corn*). Moreover it appears that the quality as well as the magnitude of these priming effects is comparable to the priming effects produced by genuine derived words (e.g., *banker-bank*). Finally, in order to separate pure morphological effects from form overlap effects, these studies used systematically orthographic control primes (i.e., morphologically simple forms whose only one part mimics a stem morpheme; such as *brothel* in which *-el* never functions as a suffix in English). Globally, the results demonstrated that the priming effects induced by derived as well as pseudo-derived primes differed significantly from these controls, suggesting that these effects resulted exclusively from the surface morphological structure of the primes. For instance, Longtin, Segui and Hallé (2003) demonstrated using French materials that a pseudo-derived word such as *baguette* (*'stick'*) (composed with the fragments *bagu-* and *–ette* that correspond to existing morphemes) facilitated the recognition of the target *bague* (*ring*) while at the same time a comparable orthographic control such as the word *abricot* (*'apricot'* in which only the fragment *abri* can be assimilated to an existing morpheme) did not facilitate the recognition of its pseudo-base *abri* (*'refuge'*). These results were replicated by Rastle, Davis and New (2004) who found a strong *corner-corn* priming effect using English materials but no priming effects with the *brothel-broth* prime-target pairs. Then, Longtin and Meunier (2005) explored the "pseudoderivation effect" using pseudo-words in order to test the resistance of early morphological decomposition face to the manipulation of the lexicality of the primes. In their masked priming study, morphologically complex pseudo-words (non existing possible words created with two existing morphemes, for instance, the base *sport-* + the suffix *-ation* produce *sportation*) were used as primes. The data revealed that pseudo-derived pseudowords (i.e., *sportation*) facilitated the recognition latencies of their base (e.g., *sport*) and did not differ from the facilitation effects obtained using transparent primes (e.g., *sportif* which is legal and

[1] It's interesting to note that under certain circumstances, a morphologically related word would not be able to facilitate or could even slow down the recognition latency of the target: when lateral inhibition is equal or stronger than excitation sent by the morphemic unit on its family members. This would be the case with prime words characterized by a high number of orthographic neighbors, a small morphological family and a weak root (in terms of its surface frequency).

semantically transparent derivation of the base *sport*). More recently, McCormick, Rastle and Davis (2008) manipulated a new category of derived stimuli, those that cannot be segmented perfectly into their morphemic components (e.g., *dropper-drop* in which there's a duplicated consonant) in order to test the flexibility of the morpho-orthographic segmentation process described by morpheme-based models. Their results demonstrated the robustness of this segmentation process in the case of various orthographic alterations in semantically related (e.g., *adorable-adore*) as well as in unrelated prime-target pairs (e.g., *fetish-fete*).

Taken together these data strongly support the robustness of a morphological decomposition effect across languages, stimuli and sensorial modalities. A complete review of the literature related to this question was made by Rastle and Davis (2008) and summarized perfectly this result in claiming: "morphological decomposition is a process that is applied to *all* morphologically structured stimuli, irrespective of their lexical, semantic, or syntactic characteristics" (p. 949). This conclusion seemed to deliver the *coup de grace* to any approach (the supralexical model in particular) that would postulate intermediate lexematic units situated above word units.

Nevertheless, the very recent study conducted by Crepaldi, Rastle, Coltheart, & Nickels (2010) opened a breach in this wall of certainty. A series of masked priming experiments were carried out on English irregularly inflected forms (viz. allomorphs). Interestingly enough and in total contradiction to their starting hypothesis, the authors found that allomorphs (e.g., *fell*) whose construction disables decomposition, primed their verbal base (e.g., *fall*) more than did orthographically-matched (e.g., *fill*) and unrelated control words (e.g., *hope*). This result had been already found by Pastizzo & Feldman (2002), and discussed enough by morphologists, but it had not been attributed the right importance by the tenants of the sublexical approach because of minor pitfalls in the control conditions (that do not have any incidence on the results, as the comparison between the data of Pastizzo & Feldman and those of Crepaldi et al. demonstrates). The authors conceded the "existence of a second higher-level source of masked morphological priming" and proposed a lemma-level composed of inflected words acting "at an interface between the orthographic lexicon and the semantic system".

However, this double source of morphological priming leads us to differentiate the nature of the coded morphemes. If we turn back to the locus issue that, according to us, determines the content of the units reflecting (and explaining) morphological effects, it is important to highlight that more than 90% of the experimental studies manipulated suffixed words or pseudowords. However, prefixed and suffixed words show many differences in terms of (1) their position relative to the stem, (2) relative number of suffixes and prefixes, (3) their grammatical properties (Montermini, 2008; Stump, 2001). To our knowledge, only few experimental studies were consecrated to affix processing representation. Two experimental papers (Colé, Beauvillain, & Segui (1989); Meunier & Segui, 1999) presented data obtained through naming and lexical decision tasks suggesting that the processing of prefixes and suffixes might differ. But masked priming studies conducted on one hand in French (Giraudo & Grainger, 2003) and on the other hand in Spanish (Duñabeitia, Perea, & Carreiras, 2008) presented contradictory results. While Giraudo and Grainger found that only prefixed primes – but not suffixed ones - produced morphological facilitation on target recognition latencies (e.g., *prénom-préface*), Duñabeitia and coll. get suffix priming (using a different experimental design[2]). We have nevertheless to observe that when it comes to the test of the decomposition hypothesis ALL the studies were conducted using suffixed words.

[2] Giraudo & Grainger examined affix priming effects using two types of affixed words, prefixed and suffixed, and three priming conditions: (1) an affix condition (e.g., *prénom-préface*), a pseudo-affix condition (e.g., *préfet-préface*) and a unrelated baseline condition (e.g., *guitare-préface*). Only prefixed prime-target pairs produced facilitation that differed significantly relative to pseudo-affixed and unrelated primes, suggesting a genuine morphological effect. Duñabeitia and coll. compared suffix priming using two kinds of words: polymorphemic (e.g., *igualdad*) vs monomorphemic (e.g., *certamen*) and two priming conditions: related vs unrelated. While they found facilitation effects for polymorphemic words (e.g., *brevedad-igualdad* faster than *plumaje-igualdad*), these effects did not occur for monomorphemic words (e.g., *volumen-certamen* equivalent to *topacio-certamen*).

The present paper attempts to bring new elements relatively to two related but unanswered questions: are pseudo-derivation effects observed using prefixed primes and are prefixes represented in long term memory? Four masked priming experiments were conducted using French materials. In these experiments we selected either morphologically simple targets (e.g., *nom*) or morphologically complex targets (e.g., *surnom*) and we systematically manipulated three priming conditions: a morphologically related condition (M+), an orthographic condition (O+) and an unrelated condition (M-O-). While the M+ condition used prefixed word primes (e.g., *prénom*), the O+ condition used either pseudo-prefixed nonwords in Exp.1 and 2 (e.g., *dénom*), nonwords containing a related stem in Exp.3 (e.g., *danom*) or nonwords containing a prefix but no stem in Exp. 4 (e.g., *prénam*). Globally, the results seem to indicate that prefixed and pseudoprefixed primes produce equivalent facilitation effects on both simple and complex target recognition. Moreover, they highlight the strong dependency of these effects on the presence of two existing morphemes within the prime, since nonword primes containing a related stem were not able to produce priming (and, in fact, this condition did not differ from the unrelated baseline condition). Taken together, these results are in line with those found using suffixed and pseudo-suffixed words (Longtin and coll., 2003, 2005; Rastle and coll., 2004; 2008). In order to integrate pseudo-derivation effects as well as affix effects (restricted to prefixes) within the same lexical architecture, while keeping in mind that for certain morphologically complex words and in particular those that cannot be decomposed into morphemes, there is a need to represent morphology at a higher level of processing (as suggested by Crepaldi and coll. 2010), we present a new architecture composed of four levels (Figure 1):

(1) Submorphemic units that only correspond to surface morphemes.
(2) Word units defining a separate level of processing which constitutes the orthographic lexicon.
(3) Base lexemes connected at the upper level with their family members.
(4) Concept units connected to both word and base lexeme units.

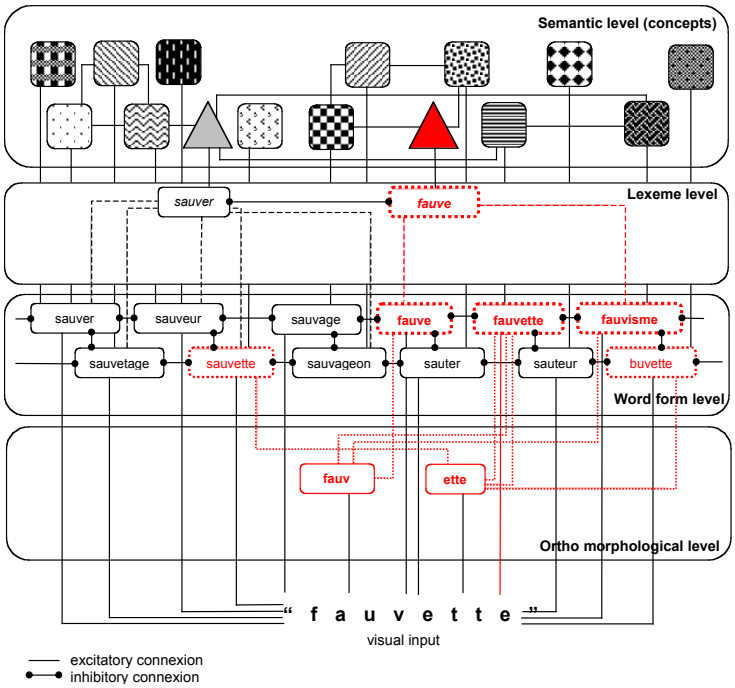


Figure 1: Hybrid model of morphological processing. The visual input *fauvette* triggers at the same time the activation of the ortho-morphological level (i.e., the morphemes *fauv-* and *–ette* are positively activated and send excitation to the related word forms ***fauve***, ***fauvette***, ***fauv***isme, sau***vette***, buv***ette***, *etc.*) and the word level (i.e., orthographic neighbours compete with each other via inhibitory connexions). Then, the lexeme *fauve* is activated but this activation is indirect because it's not triggered by the input *fauvette* but by its related word forms *fauve* and *fauvisme*.

Morphological information contained within words is then coded according to two dimensions, their surface form and their internal structure. The first level captures the perceptive regularity and the saliency of morphemes within the language. It contains the stems and affixes that can be extracted from words according to a simple segmentation process. At this level of coding, morphologically complex words, pseudo-derived words and nonwords whose surface structure can be divided into (at least two) distinct morphemes are similarly processed. As a consequence, this level cannot be considered as a morphological level in a proper sense, but rather as an ortho-morphological level. Contrarily, the second level deals with the internal structure of words, how they're formed according to morphological rules. This level contains lexemes (nouns, verbs and adjectives) abstract enough to tolerate orthographic and phonological variations produced by derivation and inflection. Lexeme representations are connected to morphologically related word representations and the connections are determined by the degree of semantic transparency between the word forms and the lexeme. Morphologically complex words that are semantically transparent, are connected with both their constituent lexemes and morphemes. However, words whose the morphological structure is semantically opaque (e.g., *fauvette 'warbler'* that is not related anymore to its free-standing stem *fauve 'tawny'*) or illusory (e.g., *baguette 'stick'* in which *bagu-* is not a stem and has nothing to do with *bague 'ring'*) are not connected with their lexeme. Both types of items are nevertheless connected with their constituent morphemes situated at the ortho-morphological level.

References

Colé, P., Beauvillain, C., & Segui, J. (1989). On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory and Language, 28*, 1-13.

Crepaldi, D., Rastle, K, Coltheart, M., & Nickels, L. (2010). Early morpho-orthographic segmentation and masked irregular priming. *Journal of Memory and Language, to appear*.

Diependaele, K., Sandra, D. & Grainger, J. (2005). Masked cross-modal morphological priming: Unravelling morpho-orthographic and morpho-semantic influences in early word recognition. *Language and Cognitive Processes, 20*, 75-114

Duñabeitia, J.A., Laka, I., Perea, M., & Carreiras, M. (2009). Is Milkman a superhero like Batman? Constituent morphological priming in compound words. *European Journal of Cognitive Psychology, 21(4),* 615-640.

Duñabeitia, J.A., Perea, M., & Carreiras, M. (2008). Does darkness lead to happiness? Masked suffix priming effects. *Language and Cognitive Processes, 23*, 1002-1020.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 10*, 680-698.

Giraudo, H., & Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin and Review, 8*, 96-101.

Giraudo, H., & Grainger, J. (2003). On the role of derivational affixes in recognising complex words: Evidence from masked priming. In H. Baayen (Ed.), *Aspects of morphological processing*. New York: Mouton de Gruyter.

Longtin, C. M., Segui, J., & Hallé, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes, 18(3)*, 313-334.

Longtin, M.C., & Meunier, F (2005). Morphological decomposition in early visual word processing. *Journal of Memory and Language, 53(1)*, 26-41

Meunier, F., & Longtin, C. M. (2007). Morphological decomposition and semantic integration in word processing. *Journal and Memory and Language, 56*, 457-471.

Meunier, F., & Segui, J. (1999). Frequency effects in auditory word recognition: The case of suffixed words. *Journal of Memory and Language, 41*, 327-344.

McCormick, SF., Rastle, K., & Davis, MH (2008), "Is there a 'fete' in 'fetish'?: Effects of orthographic opacity on morpho-orthographic segmentation in visual word recognition. *Journal of Memory and Language, 58*, 307-326.

Montermini, F. (2008). *Il lato sinistro della morfologia. La prefissazione in italiano e nelle lingue del mondo*. Milano: FrancoAngeli.

Pastizzo, M. J., & Feldman, L. B. (2002). Discrepancies between orthographic and unrelated baselines in masked priming undermine a decompositional account of morphological facilitation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 28*, 244-249.

Rastle, K, Davis, M.H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes, 23*, 942-971.
Rastle K, Davis M H, New B (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin and Review, 11,* 1090-1098.
Stump, G. (2001). Affix position, in Haspelmath M., König H., Oesterreicher W., Raible W. (a c. di) 2001, *Language Typology and Language Universals*, Berlin – New York, de Gruyter.708-714.
Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes, 9*, 271-294.

The representation and storage of lexical units.
An analysis of Catalan nouns and adjectives

Aurélie Guerrero
Fabio Montermini
(CLLE-ERSS, CNRS & Université de Toulouse)

The goal of this paper is to discuss the format under which lexical units are stored in memory, and the way in which inflectional data are organized within morphological competence. Its main focus is Catalan adjectival and nominal inflection. Catalan presents some peculiar characteristics. In particular, some word forms are apparently overtly marked for gender and / or number inflection, while others are unmarked (cf. *prim*$_{\text{MASCSG}}$ vs. *primes*$_{\text{FEMPL}}$ 'thin'). Catalan data crucially raise the question of the status to be attributed to final segments: as we will see, the same segment may sometimes explicitly mark a feminine gender inflection (as in *prima* / *primes*, where *e* and *a* both correspond to [ə]), or do not mark a gender distinction (as in *còmode*$_{\text{MASCSG}}$ / *còmoda*$_{\text{FEMSG}}$ 'comfortable' or *belga*$_{\text{MASCSG}}$ / *belga*$_{\text{FEMSG}}$ 'Belgian'). Whether these segments should be considered as inflectional morphemes (in some cases or in all cases), theme vowels, or whether they should be considered as part of the stem, will also be discussed.
The analysis proposed is realized in a Word-and-Paradigm realizational model of morphology (Stump 2001). We will claim that a realizational view is particularly compatible with an emergent view of morphology (cf. Blevins 2006), in which the relations between the forms are not expressed as traditional rules, but rather as generalizations on the existing lexicon. In this respect, frequency plays a major role. In the final part of this paper a quantitative analysis of some subclasses of nouns and adjectives will be presented. Another issue that will be discussed in the paper is the division of labour between phonology and morphology. We consider that the effect of phonology is limited to phonotactic rules operating automatically, while other relations between inflected forms (whether they are phonologically motivated or not) are purely morphological or lexical. An advantage of this view is that it allows to distinguish between phonological rules which are part of the speakers' linguistic competence, and relations between forms which have a clear historical explanation, but for which it is not clear whether they still play a role in synchrony.
The paper will be organized into three parts: the first part will be devoted to a general discussion of models of lexical storage and of the kinds of units that are included in a speaker's lexicon; in the second part we will present a global analysis of Catalan nominal and adjectival inflection within a Word-and-Paradigm model; finally, in the third part we will present a case study, namely the case of nouns and adjectives presenting a masculine singular form ending in a stressed [a].

Present-day theories of the lexicon may be divided into two main groups: those considering that the lexicon only contains idiosyncratic information (so-called impoverished theories, mainly defended within formal models of morphology, such as generative ones), and those considering that the lexicon does not necessarily store information non redundantly, and that regularly formed forms may be stored along with irregular ones (cf. e.g. Bybee 1985; Blevins 2006). Clearly, a realizational and emergent model of morphology, as the one we defend, is more compatible with the second than with the first view. In a full-entry model units does not need to be decomposed into smaller parts, and the identification of such sublexical units as morphemes, roots, stems, etc. may be viewed more as an empirical problem than as a really theoretical issue. Note that some recent developments of such symbolic models as OT recog-

nize the fact that lexemes should be memorized as a whole in the lexicon (cf. Bermúdez Otero 2009).

Catalan nouns and adjectives apparently display overt affixes for gender and number inflection. The singular number is unmarked, while the plural number is systematically marked by [s]. The masculine gender may be unmarked, but there are at least some masculine forms ending in [u] (graphically *-o*); the feminine is generally marked by [ə] (graphically *-a*). According to traditional morphemic analyses (cf. Clua, 2002 for an overview), [s] is the plural suffix, [u] is the masculine suffix for some nouns and adjectives, and [ə] is the feminine suffix. Nevertheless, there are also some masculine nouns ending in [ə] (*cotxe* 'car') and some feminine nouns that are unmarked (*sal* 'salt'); the same is true for adjectives (cf. *pobre* / *pobra* 'poor' vs. *gran* / *gran* 'big'). For adjectives like POBRE a phonological explanation is often invoked: in these cases, the [ə] is an epenthetic vowel that resolves a sequence otherwise impossible word-finally. However, while it is the case for POBRE, other adjectives displaying a final sequence, which would be acceptable in final position also end in [ə]: *còmode* ('comfortable'), *belga* ('Belgian'). There are four types of adjectives in Catalan:

(1)	a.	prim	prima	'thin'
	b.	flonjo	flonja	'soft'
	c.	gran	gran	'big'
	d.	pobre	pobre	'poor'

We consider that the paradigm of Catalan adjectives includes (at least) two stems, one for the masculine forms and one for the feminine forms. For some adjectives, in fact, these forms display an alternation that cannot be accounted for simply by phonological rules (cf. the data in (2)):

(2)	Masc Sg	Masc Pl	Fem Sg	Fem Pl	
	clar ['kla]	clars ['klas]	clara ['klaɾə]	clares ['klaɾəs]	'clear'
	car ['kar]	cars ['kars]	cara ['kaɾə]	cares ['kaɾəs]	'dear'

Under our analysis, the phoneme [ə] is part of the feminine stem and, in default cases, at least for adjectives like PRIM and CAR, the masculine (A) and the feminine (B) stems are linked by a function of the type X – Xə. For adjectives like CLAR, on the other hand, both stems have to be stored in the lexical representation of the lexeme. Of course, this function only holds for one class of adjectives. For the other types (represented in (1) by FLONJO and by GRAN and POBRE, respectively), other default functions are active, namely Xu – Xə for the first type, and X – X (identity) for the second. Thus, the final vowels are disconnected from the expression of gender, a desirable effect, since, as we have seen, they may or may not appear in forms realizing a specific gender.

To sum up, what we have in Catalan is an inflectional scheme common to all adjectives (3a), three inflectional classes, each specifying a function linking the two stems (3b), and a series of rules for the construction of actual inflected forms (3c):

(3)

a.

	Masc	Fem
Sg	A	B
Pl	A	B

b.
Class I: X – Xə (Stem B = Stem A+ə)
Class II: Xu – Xə (Stem B = Stem A–u+ə)
Class III: X – X (Stem B = Stem A)

c.
Masc Sg = Stem A
Masc Pl = Stem A+s
Fem Sg = Stem B
Fem Pl = Stem B+s

What is stored in the lexical representation of Catalan adjectives and nouns is then a fully specified word form and a pattern of relations between the stems. In the default case, one form is sufficient, and the relations in (3) allow to reconstruct the whole paradigm. The final vowels have no inflectional status (they do not directly mark gender), and they are not theme vowels, since they do not indicate inflectional class membership, which is ambiguous (a masculine in consonant may belong either to Class I or to Class III, and a feminine in [ə] to any of the three classes). Class membership is in fact indicated by the relation pattern attached to the lexical representation.
In the final part of this paper we present a case study, namely the inflection of adjectives (and nouns variable in gender) whose masculine ends in a stressed [a]. The data in (4) illustrate the different types of alternation that can be found in Catalan:

(4)	clar ['kla]	clara ['klaɾə]	'clear'	car ['kar]	cara ['kaɾə]	'dear'
	pla ['pla]	plana ['planə]	'plane'	nan ['nan]	nana ['nanə]	'dwarf'
	lilà [li'la]	lilà [li'la]	'lilac'			

Contrary to what is sometimes claimed (Mascaró 1986; Bibiloni 2002), as illustrated in (4), the relation is not phonological, at least in synchrony. We have realized an extensive research of adjectives whose masculine form ends in a stressed [a] in the *Diccionari de l'Institut d'Estudis Catalans 2* (henceforth *DIEC 2*, http://dlc.iec.cat/). ['a]/ ['anə] is indeed the most frequent pattern: in our corpus, 424 lexemes display this alternation. 381 of them (about 90%) are constructed words including a relational suffix *-à*:

(5) from places names: Catalunya < català / catalana 'Catalan'
proper nouns: Shakespeare < shakespearià / shakespeariana 'Shakespearian'
common nouns: crani < cranià / craniana 'cranial'

However, the other patterns illustrated in (4), although they are less frequent, are far from marginal. Note that the sequence ['an] in word final position is not phonologically excluded in Catalan, since several other noun and verb forms display it:

(6) blan ['blan] / blana ['blanə] 'soft'
nan ['nan] / nana ['nanə] 'dwarf'
comprèn [kum'pɾɛn] 'he understands'

Thus, when a speaker of Catalan encounters a novel and / or unknown word ending in one of the sequences above, he / she must rely on the existing lexicon, rather on an abstract rule, to inflect it. The statistic data we present suggest that frequency plays a role in the strategies chosen by speakers to reconstruct the paradigm of unknown lexemes, a conclusion that mili-

tates in favour of a probabilistic approach to morphological organisation (cf. Baayen 2003; Albright 2009).

References

Albright A. (2009), "Modelling analogy as probabilistic grammar" in Blevins J.P., Blevins J. (eds) *Analogy in Grammar*: p. 185-213. Oxford: Oxford University Press.

Baayen R.H. (2003), "Probabilistic approaches to morphology". In Bod R., Hay J., Jannedy S. (eds), *Probabilistic Linguistics*: p. 229-87. Cambridge, Ma: MIT Press.

Bermúdez Otero R. (2009), "The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features". Manuscript. University of Manchester.

Bibiloni G., (2002), "Elisió de -N i -R, distribució de les ròtiques i altres fenòmens consonàntics en el mot". In Solà J., Lloret M.R., Mascaró J., Pérez Saldanya M. (eds.), *Gramàtica del català contemporani*. Barcelona, Editorial Empúries: 273-285.

Blevins J.P. (2006), "Word-based morphology", *Journal of Linguistics* 42: p. 531-573.

Bybee J.L. (1985), *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.

Clua E. (2002), "Gènere i nombre en els noms i en els adjectius". In Solà J., Lloret M.R., Mascaró J., Pérez Saldanya M. (eds.), *Gramàtica del català contemporani*. Barcelona, Editorial Empúries: 485-534.

Mascaró J., (1986), *Morfologia*. Barcelona, Enciclopèdia catalana.

Stump G.T. (2001), *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.

Nominalisation en *–erie* à partir d'adjectifs en français et construction du sens : de l'occurrence à la propriété.

Aurore Koehl
Nancy-Université & CNRS-ATILF
Aurore.Koehl@atilf.fr

0. Introduction

Cet article consacré au groupe des noms en *-erie* s'inscrit dans le cadre de l'étude globale des nominalisations désadjectivales en français (*i.e.* les suffixations *-(i)té, -esse, -eur, -itude, -ise, -erie*), (cf. Dubois (1962, 1999), Bescherelle (1976), Lüdtke (1978)). Cette analyse se situe dans le cadre théorique, descriptif et prescriptif de la morphologie lexématique (cf. entre autres Anderson (1992), Aronoff (1994), Fradin (2003)). Des noms comme FOURBERIE ou BEAUTÉ dénotent respectivement soit la propriété[1] d'être FOURBE ou BEAU, soit des paroles ou événements manifestant cette propriété (une/des niaiserie(s), une/des beauté(s)). Les noms en *-erie* semblent pourtant se démarquer sémantiquement des autres suffixations (*e.g. -ité*) en raison du type sémantique des bases (§3.1) et du sens construit par cette suffixation (§3.2). Cette spécificité des noms en *-erie* les écarte du patron sémantique des autres noms désadjectivaux, c'est-à-dire dénoter prioritairement la propriété « être Adjectif ».

1. Problématique

En français, la suffixation en *–erie* sélectionne des bases nominales (1), verbales (2) ou adjectivales (3) (cf. Drapeau & Boulanger (1982), Temple (1991), (1996)).

(1) ORFÈVRE$_N$ > ORFÈVRERIE$_N$ « activité de l'orfèvre ; lieu de cette activité »
(2) IMPRIMER$_V$ > IMPRIMERIE$_N$ « activité d'imprimer ; lieu de cette activité »
(3) NIAIS$_A$ > NIAISERIE$_N$ « propriété de ce qui est niais »

Selon M. Temple, ces noms résultent de l'application de deux règles différentes. L'une dérive des noms d'activité sur la base de noms et de verbes (notés N/V*erie*$_N$) (4). L'autre dérive des noms de propriété sur des adjectifs (notés A*erie*$_N$), cf. (5).

(4) Règle 1 : N/V >$_{-erie}$ N/V*erie*$_{N\text{-ACTIVITÉ}}$ cf. (1-2)
(5) Règle 2 : A >$_{-erie}$ A*erie*$_{N\text{-PROPRIÉTÉ}}$ cf. (3)

Cette description engendre plusieurs questions. La première concerne le nombre de règles. La frontière entre ces deux règles est-elle véritablement aussi nette ? En d'autres termes, faut-il vraiment distinguer deux patrons de suffixation en *–erie* ? Un premier obstacle réside dans la difficulté à déterminer la catégorie grammaticale de la base (V, N ou A ?) d'un nom en *-erie*. Quelle est, par exemple, la base de BAVARDERIE (6) ?

(6) a. BAVARD$_N$ >$_{-erie}$ BAVARDERIE$_N$ « activité du bavard »
b. BAVARD$_A$ >$_{-erie}$ BAVARDERIE$_N$ « propriété de ce(lui) qui est bavard »
c. BAVARDER$_V$ >$_{-erie}$ BAVARDERIE$_N$ « activité de bavarder »

Cette hésitation conduit à se demander si tous les A*erie*$_N$ sont des noms de propriété. De manière plus générale, y a-t-il réellement superposition entre la catégorie de la base (V/N et A) et le sens construit (N-ACTIVITÉ et N-PROPRIÉTÉ) ? Si tel est le cas, tous les adjectifs bases de la règle 2 devraient dénoter des propriétés, ce qui n'est pas évident. En (7b) par exemple, *badineries* ne renvoie pas à des propriétés mais à des événements.

[1] Le terme *propriété* renvoie aux états et aux qualités, tels qu'ils sont définis par D. Van de Velde (1997).

(7) a. Marie est d'une **niaiserie** confondante.
= Marie a la *propriété* d'être niaise.
b. Il est, certes, loisible de se permettre quelques récréations et **badineries**.
= se permettre quelques *actes ou paroles* badines.

Pour éclaircir ce problème, notre étude se focalise sur les noms en *–erie* désadjectivaux, qui devraient tous renvoyer à de la propriété. Les bases sélectionnées par la suffixation en *–erie* seront comparées à celles sélectionnées par les autres suffixations désadjectivales. Si les A*erie*$_N$ sont des noms de propriété, ils doivent présenter le même fonctionnement que les autres noms désadjectivaux (eg. BANALITÉ, TENDRESSE, GRANDEUR, AMPLITUDE, BÊTISE), à savoir permettre et exclure les mêmes contextes syntaxiques que les autres noms de propriété.

2. Données empiriques

Ce travail prend en compte trois types de corpus écrits différents, dans le but de ne pas se limiter au seul registre de langue soutenu du français. Notre corpus comprend des données lexicographiques, extraites du TLF*i*, des données extraites du corpus Le Monde[2] (années 1987, 1991, 1995, 1999) et des données provenant de la Toile[3] (ici, résultats du 03/12/2009). La colonne 1 du Tableau 1 répertorie respectivement le nombre de A*erie*$_N$ traités pour chaque source. Nous analyserons plusieurs types de noms selon les propriétés sémantiques de leur base adjectivale, dont les effectifs sont enregistrés dans les colonnes 2 à 5. Parmi les bases se mêlent adjectifs ethniques (notés Aethno), idéologiques (Aidéo), évaluatifs de comportement (Acomp) et adjectifs de propriété (Aprop). La colonne 6 indique combien de noms en *-ie* ont été reconstruits avec la forme *-erie* (eg. BOURGEOISERIE). Notre analyse résulte de l'examen de quelque 846 A*erie*$_N$.

	1	2	3	4	5	6
	Nb total A*erie*$_N$	Aethno	Aidéo	Acomp	Aprop	Reconstruction
TLF*i*	204	16 **7,85%**	0 **0%**	162 **79,40%**	26 **12,75%**	-
Le Monde	21	5 **23,81%**	0 **0%**	11 **52,38%**	5 **23,81%**	0 **0%**
La Toile	621	101 **16,26%**	18 **2,90%**	295 **47,50%**	173 **27,86%**	34 **5,48%**
TOTAL	**846**	122 **14,42%**	18 **2,13%**	468 **55,32%**	204 **24,11%**	34 **4,02%**

Tableau 1. Données chiffrées de la construction A>A*erie*$_N$ selon source

3. Analyse

3.1. Quelles bases adjectivales ?

Les bases A des A*erie*$_N$ sont définies en fonction des différentes classes d'adjectifs fournies par la littérature. Un premier constat est que A n'est jamais un adjectif relationnel (cf. Mélis-Puchulu (1991), Fradin (2007)). Généralement, les adjectifs qualificatifs sont décrits comme désignant soit des états, soit des qualités (cf. Van de Velde, 1997). Les adjectifs d'état correspondent aux *stage level predicates* (cf. Carlson (1977), Kratzer (1995)) et les noms morphologiquement reliés aux adjectifs d'états psychologiques ont la particularité d'apparaître dans des structures locatives (8). Les noms correspondant aux adjectifs de qualité apparaissent dans des génitifs de qualité (9) (cf. Van de Velde, 1997).

[2] Ces données ne prennent pas en compte les A*erie*$_N$ déjà répertoriés dans le TLF*i*.
[3] Par abus de langage, nous nommons « corpus la Toile » l'ensemble fini constitué des données que nous avons récupérées en ligne à la date du 03/12/2009, via le moteur Yahoo™. Pour une discussion sur l'utilisation de ces données, cf. par exemple Kilgarriff & Grefenstette (2003) et Hathout & al. (2009).

(8) Marie est *dans* une fureur noire.
(9) Marie est *d'*une grande beauté.

Il existe un sous-type d'adjectifs de qualité, qu'on retrouve chez divers auteurs. Ces adjectifs évaluent une qualité comportementale. Quelques uns de ces adjectifs sont proposés sous (10).

(10) AIMABLE, BABILLARD, BADIN, BALOURD, BAVARD, BÉGUEULE, BIGOT

Il s'agit des *comportementaux* ('attitudinals') de P.M. Bertinetto (1994), des adjectifs d'attitude J.C. Anscombre (1995, 1996), des adjectifs orientés-agent de P. Bouillon (1997), des adjectifs de « qualité » de F. Meunier (1999), ou plus récemment des adjectifs dynamiques de P. Haas & F. Tayalati (2008).
P.M. Bertinetto (1994 : 391) définit les comportementaux comme référant à des situations dans lesquelles l'occurrence répétée d'un événement donné est pris pour être la propriété définitoire d'un individu ou d'un objet. L'auteur montre que les comportementaux se composent d'activités (au sens donné par Vendler (1967)) qui sont contextuellement transformées en états permanents.
P. Bouillon (1997) montre qu'il existe des adjectifs orientés-agent « qui évaluent un individu relativement à ce qu'il fait (cf. Ernst (1984)) ». P. Bouillon (1997 : 149) montre que les adjectifs orientés-agent font référence à la fois à un état (ou la manifestation de cet état, cf. (11)) et à un événement. C'est le cas de nombreux adjectifs bases des noms en *-erie* (12).

(11) a. Marie est **niaise** de lire ces magazines. [statique]
b. Lire ces magazines est **niais** de sa part. [manifestation]
c. Lire ces magazines **manifeste sa niaiserie**. [manifestation]
(12) Cette **niaiserie** a duré deux heures. [événement]

Enfin, P. Haas & F. Tayalati (2008) rassemblent des tests qui mettent en valeur (i) la dynamicité d'un adjectif et (ii) son agentivité. Pour ces auteurs, « le passage de la dénotation d'une occurrence d'un comportement à celle d'une propriété s'effectue par la réitération, source de la dénotation d'une habitude ». Pour ces auteurs, la spécificité des adjectifs dynamiques est « qu'ils sont prioritairement rattachés à un sujet par l'intermédiaire d'une action ». Les auteurs proposent un continuum dans la construction du sens de l'adjectif de l'interprétation d'occurrence (mode perfectif) à l'interprétation stative en passant par une interprétation habituelle (mode imperfectif).

En testant les A*erie*$_N$, à savoir en les plaçant dans les structures présentées en (11) et (12), il apparaît qu'A est dans de fortes proportions un adjectif d'évaluation de comportement (cf. Tableau 1) et correspond aux descriptions citées dans ce paragraphe.

Par ailleurs, une part non-négligeable des A*erie*$_N$ a une base ethnique ou de gentilé (Tableau 1). Nous verrons que ces A*erie*$_N$ ne renvoient pas à la propriété mais à des objets ou événements (*e.g.* une/des chinoiserie(s)). La « propriété de ce qui est chinois » est dénotée par les noms SINITÉ ou SINITUDE (cf. Dal & Namer, 2008).

3.2. A*erie*$_N$ dans le système des A*suf*$_N$.

Outre la connotation négative apportée par la suffixation en *-erie* et la préférence de cette règle pour les bases adjectivales péjoratives (suffixées en *-âtre*, *-asse* et *-ard*), nous montrerons que les A*erie*$_N$ se démarquent des autres A*suf*$_N$ de propriété en présentant une construction du sens différente.

(13) a. BANAL$_A$ >$_{-ité}$ BANALITÉ$_A$
b. la banalité de la situation = La situation est banale. [propriété]
c. Nous échangeâmes quelques banalités. [paroles]

Les noms désadjectivaux en *-eur*, *-esse*, *-ise*, *-itude* ou *-(i)té* (cf. 13) dénotent la propriété. L'interprétation parole ou événement est possible mais non systématique (*e.g.* ACCESSIBILITÉ ne renvoie pas à une occurrence). Dans le cas des A*erie*$_{N}$, l'interprétation objet, parole ou événement est générale. On ne peut attribuer la propriété A*erie*$_{N}$ à un individu qui ne s'est pas comporté d'une manière jugée A. Ce phénomène peut être élargi à la nominalisation des adjectifs d'évaluation de comportement (*e.g.* MÉCHANCETÉ). Cela est remarquable dans le cas de la suffixation en *-erie*, qui sélectionne préférentiellement ces adjectifs, alors que la suffixation en *-(i)té* par exemple ne présente pas cette contrainte.

4. Références bibliographiques

Anderson, S. R. (1992). *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Anscombre, J.-C. (1995). Morphologie et représentation événementielle : le cas des noms de sentiments et d'attitude. *Langue Française, 105*, 40-54.

Anscombre, J.-C. (1996). Noms de sentiments, noms d'attitude et noms abstraits. In N. Flaux, M. Glatigny & D. Samain (Eds.), *Les noms abstraits, histoire et théories* (pp. 257-273). Lille : Presses Universitaires du Septentrion.

Aronoff, M. (1994). *Morphology by itself*. Cambridge: MIT Press.

Bertinetto, P. M. (1994). Statives, progressives, and habituals: analogies and differences. *Linguistics, 32*, 391-423.

Bescherelle, D. (1976). *La dérivation des noms abstraits en français : concurrence des suffixes.* Université Nancy 2, Nancy.

Bouillon, P. (1997). *Polymorphie et sémantique lexicale : le cas des adjectifs*. Villeneuve d'Ascq: Presses Universitaires du Septentrion.

Carlson, G. N. (1977). *Reference to Kinds in English.* Unpublished PhD dissertation, University of Massachusetts.

Dal, G., & Namer, F. (2008). *French property nouns from toponyms or ethnic adjectives: A case of base variation*. Paper presented at the 13th International Morphology Meeting.

Drapeau, L., & Boulanger, A. (1982). Les suffixes *-erie* en français populaire : implications pour la théorie morphologique. *Revue québécoise de linguistique, 11*(2), 73-91.

Dubois, J. (1962). *Étude sur la dérivation suffixale en français moderne et contemporain*. Paris: Larousse.

Dubois, J., & Dubois-Charlier, F. (1999). *La dérivation suffixale en français*. Paris: Nathan.

Fradin, B. (2003). *Nouvelles approches en morphologie*. Paris: Presses Universitaires de France.

Fradin, B. (2007). *On the semantics of denominal adjectives*. Paper presented at the 6th Mediterranean Morphology Meeting.

Haas, P., & Tayalati, F. (2008). Les adjectifs français et l'opposition aspectuelle *statif* vs *dynamique*. *Travaux de Linguistique, 56*(2008-1).

Hathout, N., Namer, F., Plénat, M., & Tanguy, L. (2009). La collecte et l'utilisation des données en morphologie. In B. Fradin, F. Kerleroux & M. Plénat (Eds.), *Aperçus de morphologie du français* (pp. 267-287). Saint-Denis: Presses Universitaires de Vincennes.

Kilgarriff, A., & Grefenstette, G. (2003). Introduction to the special issue on the Web as a corpus. *Computational Linguistics, 29*(3), 333-347.

Kratzer, A. (1995). Stage-Level and Individual-Level Predicates. In G. N. Carlson & F. J. Pelletier (Eds.), *The Generic Book*. Chicago: University of Chicago Press.

Lüdtke, J. (1978). *Prädikative Nominalisierung mit Suffixen im französichen, katalanischen und spanischen*. Tübingen: Niemeyer.

Mélis-Puchulu, A. (1991). Les adjectifs dénominaux : des adjectifs de 'relation'. *Lexique, 10*, 33-60.

Meunier, A. (1999). Une construction complexe N°hum être Adj de V°-inf W caractéristique de certains adjectifs à sujet humain. *Langages, 33*(133), 12-44.

Temple, M. (1991). Définir le sens d'un mot construit : application aux noms d'activité en *-erie*. *Lexique, 10*, 163-209.

Temple, M. (1996). *Pour une sémantique des mots construits*. Villeneuve d'Ascq : Presses Universitaires du Septentrion.

Van de Velde, D. (1997). Adjectifs d'états, adjectifs de qualités. In D. Amiot, W. De Mulder, N. Flaux & M. Tenchéa (Eds.), *Fonctions syntaxiques et rôles sémantiques*. Arras: Artois Presses Université, 151-160.

Vendler, Z. (1967). Verbs and times *Linguistics and Philosophy* (pp. 97-121): Cornell University Press.

Morpho-phonological Blocking of Valence Changing:
Evidence from Hebrew and Palestinian Arabic

Lior Laks, Tel-Aviv University

This talk examines morpho-phonological constraints that restrict the application of valence changing operations in Modern Hebrew (hereafter MH) and Palestinian Arabic (hereafter PA). Different thematic realizations of the same verbal concept (e.g. passive, decausative and reflexive) are assumed to be derived via operations that manipulate the syntactic valence of verbs. Such operations in MH and PA usually manifested by relations among prosodically distinct configurations called binyanim (e.g. *CiCeC*, *hitCaCeC*). Following Reinhart & Siloni (2005), I assume these operations apply in the lexicon in these two languages, apart from MH passivization that applies in the syntax. Lexical operations are considered relatively less productive than syntactic ones, that is, they demonstrate gaps in the derivation of predicates. The MH verb *raxac* 'wash' has a reflexive counterpart *hitraxec* 'washed oneself', while the verb *cavat* 'pinch' has no such counterpart **hictabet* 'pinch oneself'.
What is it that restricts the application of such operations and prevents the formation of theoretically possible verbs? I contend that while some of the lexical gaps seem arbitrary, the blocking of verb formation in lexical operations is affected by morpho-phonological criteria. I present three cases of lexical gaps within the two languages, where the lack of application can only be accounted for by morpho-phonology. In addition, I show that a word-based derivation captures such restrictions and therefore should be favored.

1. Gaps of the MH CiCeC- hitCaCeC paradigm

The *CiCeC- hitCaCeC* paradigm is considered very productive in verb formation and valence changing operations. *CiCeC* usually hosts transitive verbs that are basic entries in the lexicon, while *hitCaCeC* is mainly used for the formation of derived verbs via a reduction of the syntactic valence of transitive ones (e.g. *ximem* 'make warm' – *hitxamem* 'become warm'). Verbs whose initial stem consonant is *t* or *d* usually escape *hitCaCeC*, since such derivation creates the homorganic /*tt*/ or /*dt*/ clusters, which are prohibited in Hebrew. It is not surprising that most of *CiCeC* transitive verbs with no intransitive alternate in *hitCaCeC* are ones whose initial stem consonant is *t* or *d*. The verb *dike* 'make depressed', for example, has no decausative alternate that denotes 'become depressed', as such a formation would result either in an undesired homogenous cluster (**hitdake*) or in deletion of a consonant (**hidake*). The morphological component escapes these two options and hence the formation of a possible predicate is blocked. Furthermore, a few *CiCeC* verbs that begin with *d* have intransitive alternates in the *niCCaC* template. The *CiCeC- niCCaC* paradigm is highly rare and such formations are attested only in case where the *hitCaCeC* formation is blocked due to morpho-phonological reasons. The verb *diber* 'talk', for instance, has a derived reciprocal alternate in *niCCaC* (*nidbar* 'talk to one another') rather than in *hitCaCeC* (**hitdaber*/ **hidaber*) for this reason.

2. Gaps of the PA *CaCaC- inCaCaC* paradigm

PA *CaCaC* template is used for the formation of both transitive and intransitive verbs. The intransitive derived counterparts of *CaCaC* are formed in *inCaCaC* in case of passivization and decausativization (e.g. *bana* 'build – *inbana* 'be built'). Some *CaCaC* transitive verbs have no intransitive counterpart at all. The verb *daras* 'study', for example, has no derived passive form (**indaras* 'be studied') for no apparent reason. Again, some of the gaps are idiosyncratic but others can be predicted. A dictionary search reveals that apart from two forms, verbs whose initial stem consonant is nasal have no derived counterparts in *inCaCaC* (e.g. *najad* 'rescue'- **innajad* 'be rescued', *manaħ* 'award' - *inmanaħ* 'be awarded'). This is

not surprising because such formation would result in an undesired cluster of nasal consonants. As shown for MH, theoretically possible verbs are not formed due to morpho-phonological constraints. In other cases, the morphological component finds a way of deriving such predicates by forming them in a less typical template. The verb *našar* 'spread', for instance, has a derived counterpart in the *iCtaCaC* template (*intašar* 'be spread') rather than in *inCaCaC* in order to avoid a homorganic cluster (**innašar*), although *iCtaCaC* is not used for such verbs.

3. Blocking of PA passive formation

PA passive verbs are formed in two main templates: *inCaCaC* and *tCaCCaC*. Passive formation is possible only when the input transitive verb is formed in certain template, *CaCaC* and *CaCCaC*. The former is used as a base for *inCaCaC* passive verbs (e.g. *baʿa* 'sell' - *inbaʿa* 'be sold'), while the latter is used for the formation of *tCaCCaC* passive verbs (e.g. *s'allaħ* 'fix' - *ts'allaħ* 'be fixed').

There are verbs in other templates such as *iCtaCaC* (e.g. *iqtaraħ* 'suggest') and *istaCCaC* (e.g. *istaqrad'* 'borrow'), which have no passive alternates. Which factors prevent the formation of such passive verbs? There seem to be no thematic, syntactic or pragmatic reason for this blockage of valence changing. Furthermore, passive counterparts of such verbs exist in other languages cross-linguistically (e.g. MH and English). I contend that the reason is morpho-phonological. Forming such passive verbs in one of the passive templates would involve a rather complex morpho-phonology. Non-existing (but theoretically possible) forms such as **inqaraħ* ('be suggested') or **tqarra*d' (be borrowed') cannot be derived directly from transitive alternates by adding a prefix (*iqtaraħ* 'suggest' and *istaqrad'* 'borrow' respectively). The morphological component cannot handle such formations and therefore they are entirely blocked.

The analysis reveals the effect of morpho-phonological constraints on thematic operations. The voice gaps within the three cases discussed above can only be explained via morpho-phonological restrictions. The above constraints demonstrate that the morphological component operates directly on words rather than roots and stems (Bat-El 1994, Ussishkin 1999 among others). It has to examine both the input and the output forms and keep them as faithful as possible to one another by making only the minimal changes. Such restrictions are mostly typical to operations that apply in the lexicon, in contrast to syntactic operations (e.g. MH passivization) that are much more productive with no morphological limitations. The analysis therefore supports the claim that morphology is an independent component of the grammar that interacts with the lexicon (Aronoff 1976, Anderson 1977, Scalise 1984 among others), as it can also be responsible for blocking effects on valence changing.

References

Anderson. S.R. 1977. *On the formal description of inflection*. CLS 13, 15-44.

Aronoff, M. 1976. *Word Formation in Generative Grammar.* MIT Press, Cambridge, Mass.

Bat-El, O. 1994. Stem modification and cluster transfer in Modern Hebrew. *NLLT* 12, 572-596.

Reinhart, T. & T. Siloni. 2005. The Lexicon-Syntax Parameter: Reflexivization and other Arity Operations. *Linguistic Inquiry* 36, 389-436.

Scalise, S. 1984. *Generative Morphology.* Foris, Dordrecht.

Ussishkin, A. 1999. The inadequacy of the consonantal root: Modern Hebrew Denominal Verbs and Output-output Correspondence. *Phonology* 16: 401-442.

Internally structured morphemes at the Phonology-Syntax interface: evidence from the Bosnian declensional system

Nicola Lampitelli (Université Paris-Diderot Paris 7)

The realizational process that associates a form to morphemes, i.e. Vocabulary Insertion, is a central issue in piece-based morphological theories (Halle & Marantz 1993, Marantz 2001, Embick & Halle 2005). In these models, a morpheme is a feature-matrix associated to a given terminal node. Hence, Vocabulary Items (VI) compete for insertion at node level (Embick & Marantz 2008).

In this paper, I argue for a specific interpretation of such a device. Namely I propose that only one VI corresponds to each feature-matrix, and therefore the surface exceptions are the result of either phonological processes (Lowenstamm 2008) or contextual local allomorphy rules (Embick 2010). Both processes apply post-syntactically.

The Bosnian[1] declensional system (1) provides interesting evidence to this proposal. As the correlation between gender and declension seems predictable (Corbett & Browne 2008:337-343), I consider that gender is overtly marked on nouns and that it coincides with declension.[2] Note that: (i) the suffixes on nouns are formed by one vowel;[3] (ii) only M sg. NOM displays a phonological *zero*-morpheme.

The underlying form of a Bosnian noun is shown in (2) (cf. Halle & Vaux 1998 and Halle & Nevins 2009 for similar structures).

Given (2), then M. sg. NOM must be marked by three *zero*-morphemes, i.e. gender/declension, number (#) and case (K), respectively. This is shown in (3).

A question arises now: what are the role and the structure of such a final vowel on Bosnian nouns? Standard responses within piece-based theories normally postulate processes like Fusion or Impoverishment in order to account for the mismatch between the structure (cf. 2) -where three terminal functional nodes appear- and the surface, where only one vowel is observed.

What if we look at these vowels by hypothesizing that they *do* have an internal structure (cf. Bendjaballah 2003 for a similar approach)?

For this purpose, consider the theory of Elements (Kaye et al. 1985). As far as a five-vowel language like Bosnian is concerned, this gives the results in (4) (cf. Ségéral 1995 for the use of Elements as morphemes).

As a consequence, the application of this theory to each final vowel on nouns in (1), gives the situation depicted in table (5) where case endings are shown as decomposed phonological objects.

By hypothesis in (3), NOM is marked by *zero*, therefore it is the simplest case on a phonological ground. Thus, what we observe in (5.a) is the combination of gender and # morphemes only. In (6), I recast the data including both NEU(ter) endings: *-o* and *-e* (cf. fn 7).

The analysis proceeds as follows: I_{pl} marks pl., as it appears in M and F pl. only, whereas sg. is marked by *zero*, as by hypothesis in (3). The Elements ***A*** in one side and ***I*** and ***U*** in the other, mark F and NEU, respectively. Again, (3) tells us that M is marked by *zero*. The situation is clear for M and F, but NEU needs more explications. First, notice that the surface difference between [o] and [e] is given by ***I*** vs. ***U***. Secondly, note that NEU is characterized by NOM-ACC syncretism (cf. 1), which -I claim- is formally explained by the presence of the Element A in both NOM and ACC.

---

[1] My informant comes from Bosnia-Herzegovina and for this reason I refer to Bosnian-Serbo-Croatian as to simply Bosnian.

[2] A fourth small group exists, cf. *stvar* 'thing', where all the nouns are F, but the declension is different from group 1. Note that group 1 only contains M nouns whereas group 3 only NEU ones. Group 2 has some M nouns: cf. *jedan sudji**a*** 'one.M judge' vs. *jedna kuć**a*** 'one.F house'.

[3] Whenever a consonant is present (pl. DAT and INSTR), the suffix is: /Vm(a)/, where V stands for the alternating vowel. I am aware that some F nouns can have a INSTR -ju *stvarju* 'with a thing', but these belong to the marginal -i declension (cf. fn 2) which is nevertheless possibly accounted for by my approach by positing a contextual allomorphy phenomenon.

This brings us to show all the underlying morphemes, as in (7). Each terminal has only one Vocabulary Item associated to, e.g. pl.↔I_{pl}, ACC↔A or DAT/LOC↔U, etc.[4]

I propose the structure in (8.a) and the corresponding complex head at PF (8.b) in order to account for each noun in (1) (Th is inserted as by a general requirement of Bosnian): *kuće* 'houses' F, pl, NOM is taken as an example.[5] At PF, each terminal can be associated to a CV syllable (in the sense of Lowenstamm 1996, 2008): in this model, only the phonological material associated to the CVCV.. cluster surfaces.

Crucially, each node/morpheme in the structure corresponds to a phonological primitive contained in each final vowel, allowing for a one-to-one parallelism between a given node and a VI.

Mismatches between phonology and syntax can however occur, cf. double-framed cases in (8). White-framed cases are explained by the theory of Elements itself (no possible /U+I/ combinations in five-vowel languages), whereas grey ones are instances of contextual allomorphy rules which apply locally (cf. Embick 2010). Neither Fusion nor Fission must be postulated.[6]

Final vowels on Bosnian nouns are complex phonological objects that are internally structured in such a way that they reflect syntactic structure. The consequence of such an approach is a more abstract phonological representation (cf. Lowenstamm 2008) and the vacuity of the notion of "paradigm", as well-formedness is locally determined (cf. Bobalijk 2008); finally, class is not a feature of stems (cf. Halle & Marantz 2008).

(1)	group 1 (M)		group 2 (F)		group 3 (NEU)	
	sg.	pl.	sg.	pl.	sg.	pl.
a. NOM	*okvir*	*okvir-**i***	*kuć-**a***	*kuć-**e***	*sel-**o***[7]	*sel-**a***
b. GEN	*okvir-**a***	*okvir-**a***	*kuć-**e***	*kuć-**a***	*sel-**a***	*sel-**a***
c. DAT-LOC	*okvir-**u***	*okvir-**ima***	*kuć-**i***	*kuć-**ama***	*sel-**u***	*sel-**ima***
d. ACC	*okvir-**(a)***[8]	*okvir-**(e)***	*kuć-**u***	*kuć-**e***	*sel-**o***	*sel-**a***
f. INSTR	*okvir-**om***	*okvir-**ima***	*kuć-**om***	*kuć-**ama***	*sel-**om***	*sel-**ima***
	'frame'	'frames'	'house'	'houses'	'village'	'villages'

(2) *Underlying structure of a noun*: Root + gender/declension + # + K

(3) *Null Morphemes*: a. M = *zero*; b. sg. = *zero*; c. NOM = *zero*.

(4) *Decomposed vowels*: a. [a] = /A/; b. [i] = /I/; c. [u] = /U/; d. [e] = /A.I/; e. [o] = /A.U/

(5) *Decomposed vocalic case endings*

	M sg.	M pl.	F sg.	F pl.	NEU sg.	NEU pl.
a. NOM	*zero*	I	A	A.I	A.U	A
b. GEN	A	A	A.I	A	A	A
c. DAT/LOC	U	I(ma)	I	A(ma)	U	I(ma)
d. ACC	(A)	A.I	U	A.I	A.U	A
f. INSTR	A.U(m)	I(ma)	A.U	A(ma)	A.U(m)	I(ma)

[4] Note that NEU examples in 8 have I_{pl}, as by hypothesis. 6.c-d are incomplete representations of the structures.
[5] Cf. Halle 1997 and Lowenstamm 2008 respectively for case and gender featural decomposition.
[6] Note that GEN forms seem to be "impostors" like in Russian, cf. Bailyn & Nevins (2008).
[7] NEU nouns can be marked in sg. direct cases (NOM & ACC) by [e], too i.e. *srce* 'heart'. The theory proposed here accounts for both endings in NEU, as 6 below shows. The crucial issue is that a unique underlying Element A marks both ACC and NOM in the NEU paradigm as it is the spell-out for the ACC case. This accounts for the NOM-ACC syncretism.
[8] M nouns are marked by -a in sg. ACC when the referent is animate. On the other hand, Pl. ACC is always -e.

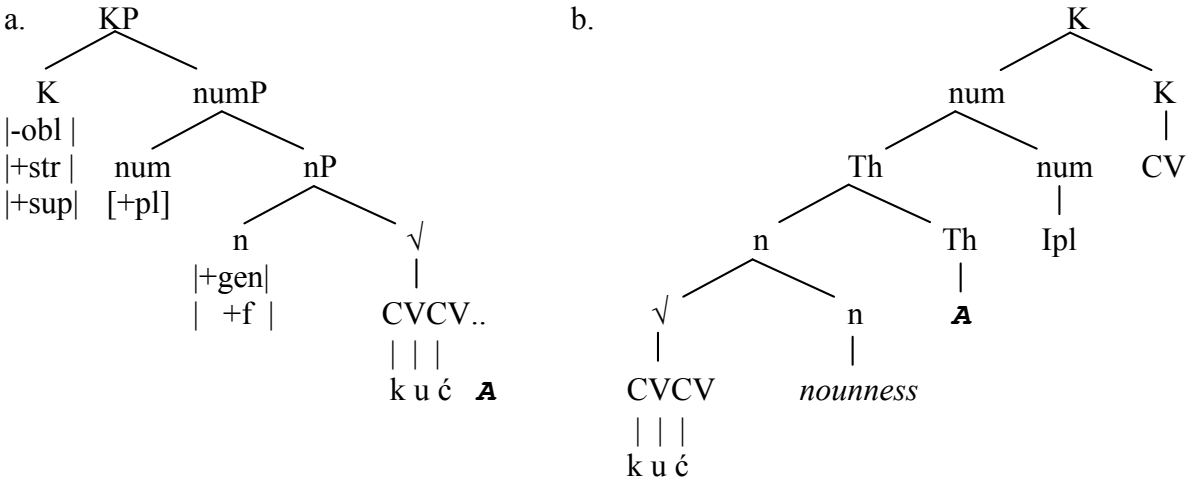
(6) *Decomposed NOM endings*

a. Masculine		b. Feminine		c. Neuter I		d. Neuter II	
sg.	pl.	sg.	pl.	sg.	pl.	sg.	pl.
zero	I_{pl}	***A***	***A***.I_{pl}	A.***I***	A	A.***U***	A

(7) *Underlying morphological structures for final vowels: gender/declension.#.K*

	M sg.	M pl.	F sg.	F pl.	NEU sg.	NEU pl.
a. NOM	ø.ø.ø	ø.Ipl.ø	***A***.ø.ø	***A***.Ipl.ø	***U***.ø.A	***U***.Ipl.A
b. GEN	ø.ø.A	ø.Ipl.A	***A***.ø.A	***A***.Ipl.A	***U***.ø.A	***U***.Ipl.A
c. DAT/LOC	ø.ø.U	ø.Ipl.U	***A***.ø.U	***A***.Ipl.U	***U***.ø.U	***U***.Ipl.U
d. ACC	ø.ø.A	ø.Ipl.A	***A***.ø.A	***A***.Ipl.A	***U***.ø.A	***U***.Ipl.A
f. INSTR	ø.ø.[A.U]	ø.Ipl.[A.U]	***A***.ø.[A.U]	***A***.Ipl.[A.U]	***U***.ø.[A.U]	***U***.Ipl.[A.U]

(8) *Structure for a noun*



output: [kuće] ‘houses’

References

Bailyn, J. & A. Nevins. 2008. “Russian genitive plurals are impostors”. In *Inflectional Identities*, ed. by A. Bachrach & A. Nevins, 237-270. Oxford University Press.

Bobaljik, J. 2008. “Paradigms (Optimal and Otherwise): A Case for Scepticism”. *Inflectional Identity*, ed. by A. Bachrach and A. Nevins, 29-54. Oxford University Press.

Bendjaballah, S. 2003 “The internal structure of the determiner in Bedja”. *Research in Afroasiatic Grammar*, ed. by J. Lecarme, 35-52. Amsterdam: John Benjamins.

Corbett, G. & W. Browne. 2008. “Serbo-Croat. Bosnian, Croatian, Montenegrin, Serbian. In *The World’s Major Languages*, ed. by B. Comrie, 330-346. New York: Routledge.

Embick, D. 2010. *Localism versus Globalism in Morphology and Phonology*. MIT Press Monograph.

Embick, D. & M. Halle. 2005. “On the status of *stems* in Morphological Theory”. *Proceedings of Going Romance 2003*, ed. by T. Geerts, I. van Ginneken & H. Jacobs, 37-62. Amsterdam: John Benjamins.

Embick, D. & R. Noyer. 2001. “Movement Operations after Syntax”. *Linguistic Inquiry* 32.4:555-595.

Halle, M. 1997. “Distributed Morphology: Impoverishment and Fission”. *Papers at the Interface*, ed. by B Bruening, Y. Kang, M. McGinnis, 425-449. MIT Press, vol. 30 of *MIT Working Papers in Linguistics.*

Halle, M. & A. Marantz. 1993. “Distributed Morphology and the pieces of inflection” In *The view from the Building 20*, ed. by K. Hale & S. J. Keyser, 111-176. MIT Press, Cambridge.

Halle, M. & A. Marantz. 2008. “Clarifying ‘Blur’: Paradigms, Defaults and Inflectional Classes”. In *Inflectional Identities*, ed. by A. Bachrach & A. Nevins, 55-72. Oxford University Press.

Halle, M. & A. Nevins. 2009. “Rule Application in Phonology”. *Contemporary Views on Architecture and Representations in Phonology*, ed. by E. Raimy and C. Cairns, 355-382. Cambridge, Mass.: MIT Press.

Halle, M. & B. Vaux. 1998. “Theoretical aspects of Indo-European nominal morphology: The nominal declensions of Latin and Armenian”. *Mir Curad: Studies in Honor of Clavert Watkins*, edited by J. Jasanoff, H. C. Melchert, and L. Olivier, Innsbrucker Beitrage zur Sprachwissenschaft, pp. 223–240. Institut f¨ur Sprachwissenschaft der Universit¨at Innsbruck, Innsbruck.

Kaye, J., J. Lowenstamm & J. R. Vergnaud. 1985. “The internal structure of phonological elements: a theory of charm and government” *Phonology Yearbook* 2:305-328.

Kiparsky, P. 1996. “Allomorphy or Morphophonology?” *Trubetzkoy’s Orphan: Proceedings of The Montréal Roundtable ‘Morphophonology: Contemporary Responses’*. Ed. by R. Singh, 12-31. Amsterdam: John Benjamins.

Lowenstamm, J. 1996. “CV as the only syllable type”. *Current Trends in Phonology: Models and Methods*, ed. by J. Durand & B. Laks, 419-441. Manchester: Salford.

Lowenstamm, J. 2008. “On *n*, nP and √”. In *The Sounds of Silence: Empty Elements in Syntax and Phonology*, ed. by J. Hartmann, V. Hegedus & H. van Riemsdjik, 105-144. Amsterdam: Elsevier.

Marantz, A. 2001. “Words and Things”. Ms. MIT.

Ségéral, Ph. 1995. *Une théorie généralisée de l’apophonie*. PhD diss. Université Paris VII.

Les suffixations en *–iser* et en *–ifier* : vérifier les données pour vériser les hypothèses ?

Stéphanie LIGNON
ATILF, MSH & Université Nancy 2
stephanie.lignon@univ-nancy2.fr

Le français dispose de différents moyens afin de construire des verbes dits de changement d'état à base nominale et/ou adjectivale : les préfixations en *a-*, *en-*, *dé-* e.g. ALLONGER, ENLAIDIR ou DÉSHERBER ; la conversion e.g. BALISER, BLANCHIR ; les suffixations en *–iser* et *–ifier* e.g. ATOMISER, ALCOOLISER et IMMENSIFIER, MOMIFIER. Nous nous proposons de nous concentrer ici sur les règles de construction en *–iser* et en *–ifier*. Notre objectif est d'identifier les paramètres régissant l'adjonction de l'un ou l'autre de ces exposants de règles de construction de lexèmes (désormais RCL). Les descriptions qui leur sont consacrées (Nyrop (1936 : 206), Dubois (1962 : 19), TLFi) ne permettent pas de les opposer. Nous nous intéresserons tout d'abord aux formes construites présentes dans le TLFi, pour vérifier ensuite les hypothèses que nous aurons établies à partir d'un corpus de néologismes. Nous nous proposons de montrer que la suffixation en *–ifier* n'est actuellement qu'une suffixation supplétive, dont l'unique rôle est de pallier les carences morphophonologiques de la suffixation en *-iser*.

1. Les formes du TLF

L'étude systématique des formes construites en *–iser* et en *–ifier* du TLFi (désormais **Xiser** et **Xifier**) permet de recueillir 789 **Xiser** et 151 **Xifier**, soit 5 fois moins. Bien que toujours considérée par le TLF comme disponible, la suffixation en *–ifier* est donc très minoritaire par rapport à celle en *–iser*.

1.1. La sémantique des Xiser et Xifier

D'autres auteurs se sont intéressés avant nous à la concurrence entre les **Xiser** et les **Xifier** ou à leurs équivalents dans d'autres langues (F. Namer (2009) pour le français, I. Plag (1999) pour l'anglais, F. Rainer (1993) pour l'allemand). Pour Plag (1999), les formes en *–ize*, l'équivalent anglais du suffixe *–iser*, se rencontrent dans 7 emplois distincts en fonction des relations entre base et dérivé. Ces emplois peuvent également être utilisés pour le français (cf. Tableau 1 ci-dessous). Les emplois de *–ize* peuvent être subsumés par une structure lexicale conceptuelle unique sous-spécifiée (en opposition à Lieber (1996) qui en propose 5) et l'interprétation du dérivé sera fonction de la nature sémantique et catégorielle de la base, ainsi que de ses propriétés pragmatiques. I. Plag s'intéresse également à la concurrence entre la suffixation en *–ize* et en *–ify*. On retrouve *–ify* dans de nombreux emplois identiques à ceux de *–ize*. Cette concurrence sémantique se retrouve en français :

locatif	mettre dans X	HOSPITALISER	
décoratif	munir de X	ARBORISER	PLASTIFIER
causatif	rendre (plus) X	BANALISER	ACIDIFIER
résultatif	transformer en X	AGATISER	MOMIFIER
inchoatif	devenir un X	CICATRISER	BARONIFIER
performatif	faire X	COURTISER	NIDIFIER
similatif	agir comme un X	CHAMANISER	

Tableau 1 : Différents emplois de *–iser* et de *-ifier*, à partir de I. Plag (1999 : 196)

F. Namer (2009) a également étudié la concurrence entre *–iser* et *–ifier* en vue du traitement automatique et propose une analyse fondée sur la catégorie grammaticale de la base (N ou Adj), sur la syntactique du verbe construit ainsi que sur les rapports sémantiques qu'entretient

le Nbase avec les différents actants du verbe construit. Elle propose notamment 5 catégories concurrentes pour les verbes en *–iser* et *–ifier* construits sur base nominale (Namer : 183).
De notre côté, nous avons constaté que les **Xifier** du TLFi sont majoritairement dénominaux. Seul un tiers des **Xifier** a pour base un adjectif, généralement gradable. De plus, l'interprétation des **Xifier** est rarement ambiguë, contrairement à ce que l'on peut observer sur les construits en *–iser*, comme par exemple avec ARISTOCRATISER, dont l'interprétation peut tout à la fois renvoyer à ARISTOCRATE$_N$ ou ARISTOCRATIQUE$_A$ (1). Les interprétations des construits en *-ifier* sont donc beaucoup plus restreintes que celles des construits en *–iser*, ce qui explique également la plus grande disponibilité sémantique de la RCL en *–iser*.

(1) « […] animé par le même projet d'égalité abstraite, reste néanmoins emblématique d'une idéologie héroïque visant à ***aristocratiser*** les masses. »

Le consensus sur la concurrence sémantique entre les RCL en *–iser* et en *–ifier* nous permet de nous intéresser à la distribution morphophonologique de ces affixes.

1.2. Morphophonologie

Les travaux de I. Plag ont montré qu'en anglais, la répartition entre *–ify* et *–ize* était pour partie conditionnée par la morphophonologie. Un certain nombre de paramètres, notamment accentuels, expliquent le choix de la forme affixale. En français, la répartition des deux affixes est également dépendante de paramètres morphophonologiques.

1.2.1. Dissimilation

L'analyse des **Xiser** et **Xifier** du TLF montre que les phonèmes précédant immédiatement le suffixe ont une répartition différente devant chacun des suffixes. La dernière colonne du tableau comporte l'écart de pourcentage entre les **Xiser** et **Xifier**.

		Nombre de **Xifier**	%	Nombre de **Xiser**	%	écart
Consonnes	**s**	**21**	**14,09%**	**17**	**2,15%**	**-11,94%**
	d	13	8,72%	13	1,65%	-7,07%
	t	31	20,81%	114	14,47%	-6,34%
	OL	11	7,38%	13	1,65%	-5,73%
	z	**5**	**3,36%**	**0**	**0,00%**	**-3,36%**
	ɲ	4	2,68%	1	0,13%	-2,55%
	b	3	2,01%	3	0,38%	-1,63%
	g	1	0,67%	2	0,25%	-0,42%
	k	1	0,67%	2	0,25%	-0,42%
	p	1	0,67%	4	0,51%	-0,16%
	v	1	0,67%	7	0,89%	+0,22%
	ʒ	1	0,67%	10	1,27%	+0,60%
	f	0	0%	5	0,63%	+0,63%
	ʃ	0	0%	8	1,02%	+1,02%
	m	5	3,36%	37	4,70%	+1,34%
	r	20	13,42%	137	17,39%	+3,97%
	n	16	10,74%	133	16,88%	+6,14%
	l	**10**	**6,71%**	**274**	**34,77%**	**+28,06%**
Voyelles	E	5	3,36%	6	0,76%	-2,60%
	O	0	0%	1	0,13%	+0,13%
	U	0	0%	1	0,13%	+0,13%
		149		788		

Tableau 2 : pourcentage d'apparition des phonèmes devant *–iser* ou *-ifier*

Si la majorité des phonèmes se retrouvent dans des proportions similaires devant *–iser* ou *-ifier*, on peut toutefois constater un certain nombre de faits remarquables :

- les sifflantes ont une fréquence d'apparition plus de 8 fois plus importante devant *–ifier* que devant *-iser*. Cette répartition s'explique aisément par le respect de contraintes

dissimilatives, dont le poids a été démontré pour le français entre autres dans Plénat (2000), Roché (1997) et Lignon et Plénat (2008). Que le suffixe *–iser* répugne à s'adjoindre à des bases présentant en finale, ou en dernière attaque dans le cas de bases à finale vocalique, un phonème identique à l'un de ceux qu'il comporte lui-même est donc parfaitement attendu. Il a déjà été établi qu'un des moyens pour pallier une incompatibilité phonologique de cette nature pouvait être le recours à un suffixe concurrent. Le suffixe *–ifier* ne présentant pas cette incompatibilité, et la RCL qui l'utilise étant largement similaire à celle utilisant *–iser*, tant sur le plan sémantique que syntactique, il peut donc parfaitement prendre le relais. Ce qui est plus surprenant, c'est que le suffixe *–iser* s'adjoigne malgré tout à des formes *a priori* inconciliables phonologiquement. Or, dans de nombreux cas (12), il existe dans le paradigme dérivationnel du Nbase un N en *–isme* et on observe des triplets N/Nisme/Niser : MARX/MARXISME/MARXISER, CATHOLIQUE/CATHOLICISME/CATHOLICISER, etc. Parfois le triplet se réduit à un doublon Xisme/Xiser : FASCISME/FASCISER, EXORCISME/EXORCISER, OSTRACISME/OSTRACISER, etc.

- le /l/ : Le /l/ apparaît de façon tout à faire remarquable devant *–iser*. Cette fréquence exceptionnelle est due à la sélection massive par *–iser* de base adjectivale ou tout du moins de forme adjectivale en *–al* : NORMALISER, BESTIALISER, CENTRALISER, etc.

1.2.2. Taille

Plus de la moitié des **Xifier** du TLF est construite à partir de bases monosyllabiques, alors que le pourcentage de bases monosyllabiques pour les **Xiser** est inférieur à 4%. L'attirance remarquable de *–ifier* pour les bases monosyllabiques est sans doute à ramener à la structure segmentale de l'exposant de RCL. En effet, l'adjonction de *-ifier* (sans marque flexionnelle) augmente de deux syllabes les bases à finale consonantique (majorité des cas), une syllabe pour les bases à finale vocalique, par substitution. En revanche, le suffixe *–iser* construit des dérivés dont la taille est supérieure d'une syllabe pour les bases à finale consonantique ou isosyllabiques pour les bases à finale vocalique.

La répartition entre les suffixes *–iser* et *–ifier* semble donc partiellement conditionnée par des paramètres morphophonologiques : la suffixation en *–iser* est la suffixation par défaut. La suffixation en *–ifier* est cantonnée aux bases monosyllabiques et aux bases présentant en finale ou en dernière attaque une sifflante.

2. Xiser et Xifier sur la Toile

2.1. Méthodologie

Nous nous proposons de vérifier les hypothèses établies sous 1 à l'aide de néologismes attestés sur la Toile (extraits à l'aide du moteur de recherche Google courant avril 2010). Pour cela, nous avons extrait les noms et les adjectifs de la base de données BRULEX, à partir desquels nous avons construit une liste de lexèmes candidats en *–iser* et en *–ifier*. Nous avons ensuite effectué des requêtes sur la Toile à l'aide de la forme infinitive de ces lexèmes à l'aide du robot Walim (Namer, 2003).

2.2. Disponibilité

Même si *–iser* et *–ifier* sont présentés comme toujours disponibles (cf. TLF), les néologismes en *–ifier* sont nettement moins fréquents que ceux en *–iser*. Nous nous proposons de vérifier, auprès d'un corpus de formes néologiques, le poids des contraintes morphophonologiques identifiées ci-dessus :

- La suffixation en *–iser* est la suffixation par défaut, attendue pour toutes les bases, sauf présence d'une sifflante en finale ou dernière attaque : sur SARKOZY on trouve *sarkozifier* et non *sarkoziser*, sur VORACE, *voracifier* et non *voraciser*, etc.
- La suffixation en *–ifier* est réservée aux bases monosyllabiques sauf si elles présentent en finale un /f/ (ou sa correspondante sonore /v/). Dans ce cas, c'est la forme par défaut qui

l'emporte : SAGE>SAGIFIER, ROND>RONDIFIER, POULE>POULIFIER, mais SUAVE>SUAVISER et non SUAVIFIER ; POUF>POUFISER et non POUFIFIER.

Dans tous les autres cas, *–iser* sera l'exposant par défaut.

(1) « Bref, s'il veut commencer à remonter la pente, le président aurait tout intérêt à se "***filloniser***". Allez, un petit effort. Auteur : Franz-Olivier Giesbert **»**

(2) « A condition aussi qu'il continue d'infléchir, de ***villepiniser*** son discours devenu, il est vrai, beaucoup plus laïc et républicain, […] »

2.3. Concurrence des Xiser/Xifier

On trouve un certain nombre de doublons sur la Toile avec des bases pour lesquelles on attendrait la suffixation en *–ifier* pour des raisons de taille, comme par exemple *jeuniser* (92) et *jeunifier* (169), *bulliser* (47) et *bullifier* (9), *blondiser* (64) et *blondifier* (261). C'est aussi le cas, dans une moindre mesure, avec les bases à finale (ou dernière attaque) en sifflante. La forme *pétassiser*, construite sur PÉTASSE, apparaît deux fois plus que sa concurrente *pétassifier*, pourtant mieux adaptés morphophonologiquement :

(3) « Ils sont obligés de la ***pétassiser*** au maximum pour détourner l'attention de sa voix horripilante. Et ça marche. **»**

Et *poufiassiser*, *radassiser* ou *blondassiser* sont les seules formes attestées sur la Toile :

(4) « je pense que sa m'aurais detruit de te voir te ***poufiassiser*** autant. **»**

(5) « Bien placé dans la capitale je pourrai enfin me ***radassiser***! »

(6) « réfléchissé bien qui on va ***blondassiser*** a la premiere acti !! je vous aime mes blondasses !!!!!!!!!!!!! **»**

D'une part, la pression exercée par le poids des formes en *–iser* est susceptible de prendre le pas sur les contraintes morphophonologiques et les formes en *–iser* sont susceptibles d'apparaître alors même que l'on serait en droit d'attendre *–ifier*.

D'autre part, la nature morphologique des bases sélectionnées ainsi que leur appartenance à des paradigmes lexicaux peuvent également être des facteurs de choix de la forme affixale. Nous avons déjà observé la construction de formes en *–iser* sur des bases en sifflantes dans le TLFi (cf. 1.2.1), bases ayant dans leur paradigme lexical des formes en *–isme* et/ou en *–ique*. L'attirance remarquable de la suffixation en *–iser* pour les bases en *–ique* permet la construction de formes telles que LYRICISER alors même que LYRIFIER est attesté dans le TLFi et que la forme construite contrevient à la contrainte de dissimilation.

(7) « ça vogue de nuée en nuée…enfin je voudrais pas *lyriciser* hein ! à côté de ça, son article "L'effet de réel " est vraiment habité »

Conclusion

De nos jours, la RCL en *–ifier* n'est plus productive, contrairement à la suffixation en *-iser*, comme l'ont observé Dal *et alii* (2008) dans leur étude sur la productivité affixale. Cette forme affixale n'intervient que dans des contextes morphophonologiques défavorables à la suffixation en *–iser*. La description de la concurrence de ces deux formes affixales doit nécessairement prendre en considération les liens sémantiques et syntactiques entre la base et le construit, les contraintes morphophonologiques de bonne formation mais également la nature morphologique de la base et son appartenance à un paradigme lexical.

Bibliographie

Dal, G., Fradin, B., Grabar, N., Lignon, S., Namer, F., Plancq, C., Yvon, F. et Zweigenbaum, P. (2008). « Quelques préalables au calcul de la productivité des règles constructionnelles et premiers résultats ». *CMLF 2008*, Paris.

Plag, I. (1999), *Morphological Productivity. Structural Constraints in English derivation*. Berlin, New-York : Mouton de Gruyter.

Lieber, R. (1998), « The Suffix *–ize* in English. Implications for Morphology » in Steven G. Lapointe, Diane K. Brentari, and Patrick M. Farrell (eds.) *Morphology and its Relation to Phonology and Syntax*. Stanford : CLSI : 12-33.

Lignon S. et M. Plénat. (2008). « Echangisme suffixal et contraintes phonologiques (Cas des dérivés en *-ien* et en *-icien*)". *In* B. Fradin, F. Kerleroux & M. Plénat (éds). *Aperçus de morphologie du français.* Saint-Denis, Presses Universitaires de Vincennes, pp. 65-81.

Namer, F. (2003). « WaliM : valider les unités morphologiquement complexes par le Web » *in* D. Corbin, B. Fradin, B. Habert, F. Kerleroux, & M. Plenat (eds.), Mots possibles et mots existants, Actes du colloque de Villeneuve d'Ascq (3ème Forum de Morphologie, 20-21 septembre 2002, Lille), [= Silexicales 3], p.142-150. Villeneuve d'Ascq: Presses Universitaires du Septentrion.

Nyrop Kristoffer (1936), *Grammaire Historique de la langue française*, Tome troisième, *Copenhague*, Gyldenskale Boghandel Nordisk Forlag.

Namer F., 2009, *Morphologie, lexique et traitement automatique des langues : l'analyseur Dérif*, Paris : Hermès.

Plénat**,** M.**.** (2000). « Quelques thèmes de recherche actuels en morphophonologie française ». *Cahiers de lexicologie* 77, pp. 27-62.

Rainer, F. (1993), *Spanische Wortbildungslehre*. Tübingen : Niemeyer.

Roché, M. (1997), « *Briard, bougeoir* et *camionneur*. Dérivés aberrants, dérivés possibles », *in* D. Corbin, B. Fradin, B. Habert, F. Kerleroux, & M. Plenat (eds.), Mots possibles et mots existants, Actes du colloque de Villeneuve d'Ascq (Forum de morphologie, 1res rencontres, 28-29 avril 1997), [= Silexicales 1], pp. 241-250. Villeneuve d'Ascq: Presses Universitaires du Septentrion

Bases de données

BRULEX, http://www.lexique.org/

Natural selection in self-organizing morphological systems

Mark Lindsay
Stony Brook University

Mark Aronoff
Stony Brook University

In recent years, it has been shown that certain subsystems of human languages can be profitably investigated in terms of self-organizing or emergent systems (Baayen (1993), Anshen and Aronoff (1999), Albright (2002), Plag and Baayen (2009)). Using new methods made available by the Internet and modern databases, we will show that the birth of productive affixes from borrowed vocabulary can be treated as an emergent system, where affixes survive or perish due to circumstances of environment and competition. We present two types of evidence for this position. The first is historical and addresses the question of how a language borrows affixes at all. We examine the history of two sample suffixes in English, *-ment* and *-ity*, both of which came into English from French, using the OED as a database. The second type of evidence is a synchronic investigation of the rival suffix pair *-ic* and *-ical*, using Google as a measurement of relative productivity.

Borrowed Suffixes *-ment* and *-ity*

In Figure 1, we see the rate of borrowings of words containing *-ment* and *-ity* from 1250 to the present day (adjusted for the total number of words entering the OED during that time period). Both of these suffixes had a large number of borrowings from French early on, followed, unsurprisingly, by a gradual decline; this decline reached nearly zero by the end of the 20th century.

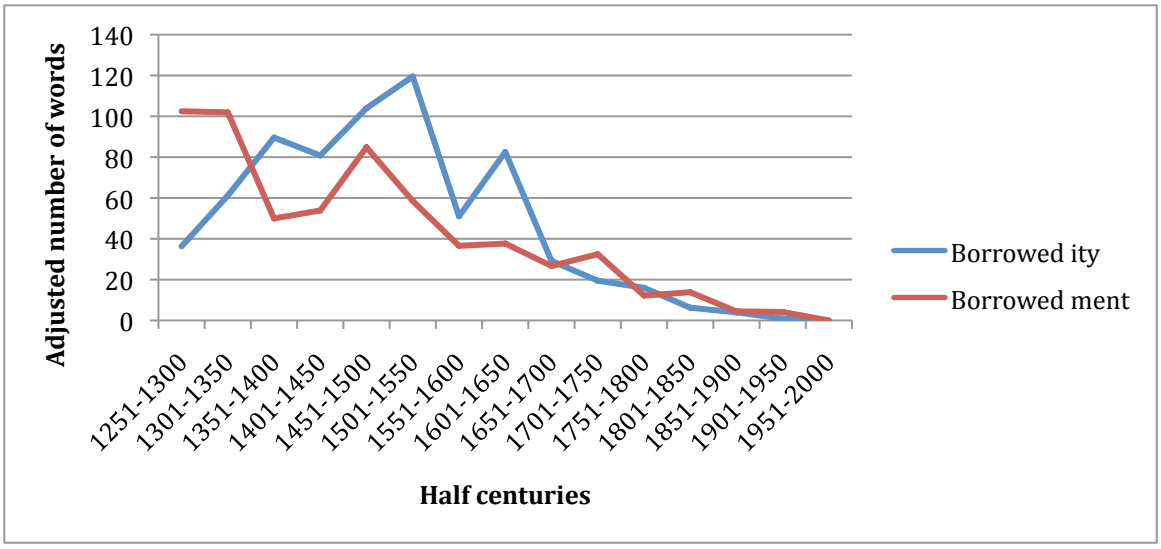


Figure 1 – Borrowed -ity vs borrowed -ment (adjusted) (modified from Anshen & Aronoff 1999)

In Figure 2, we see that the fate of these two suffixes was quite different. Early on, few words of English were derived using *-ment* or *-ity* as a productive suffix. The number of derivations increased, presumably as the number of exemplars increased as a result of continued French borrowings. However, in the early 17th century, the productivity of *-ity* and *-ment* began to change drastically. While *-ity* flourished, creating hundreds of new derived forms, *-ment* began a decline that has resulted in zero derived forms by the present day.

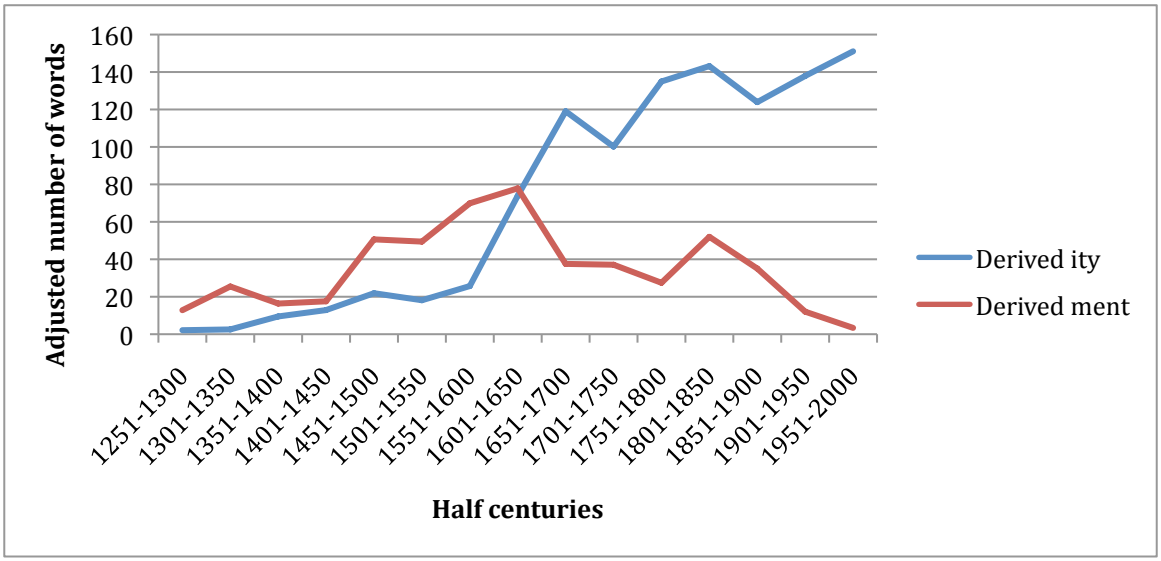


Figure 2 – Derived -ity vs. derived -ment (adjusted) (modified from Anshen & Aronoff 1999)

Why did *-ity* sustain itself as a productive affix while *-ment* failed? By their very nature, productive affixes exist in morphological ecosystems, where they depend on new words as sources for continued productivity. These two suffixes had different productive "niches": *-ity* attached to adjectives (e.g. *equal* → *equality*) and *-ment* attached to verbs (e.g. *punish* → *punishment*). We see in Figure 3 that the number of new verbs declined over the centuries, with its most significant dip taking place in the 17th century, the same time that *-ment* began its decline in productivity. On the other hand, the number of adjectives increased and remained far above the number of verbs. Thus, *-ment* died out because it could no longer be sustained in its ecosystem.

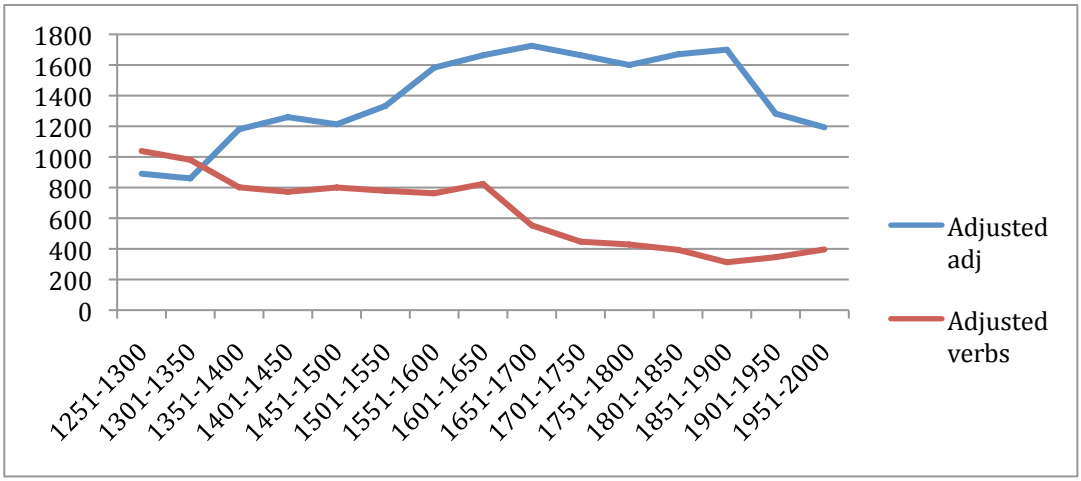


Figure 3 – Total new adjectives vs. new verbs per half-century (adjusted)

Rival suffixes *-ic* and *-ical*

Another type of evidence comes from the study of the rival suffix pair *-ic* and *-ical*. We call the pair 'rivals' because they are synonymous, and thus are in direct competition with each other. Why would a language tolerate synonymous affixes? The question is even more mysterious in this case because, although both have their origins in borrowed suffixes (Greek *–ik* and Latin *–al*), *-ical* is more or less an English creation.

Although the language exhibits doublets, e.g. *historic* and *historical*, the distinctions in meaning or usage exist only in the full word pairs; the difference between *historic* and *historical* cannot be generalized to predict that of, e.g., *electric* and *electrical*.

Furthermore, in most pairs, one form is strongly preferred over the other: *electronic* is much more common than *electronical*, while *surgical* is much more common than *surgic*. The first question we ask is whether there are grounds for saying that one suffix is more productive than the other. What is novel about this research is the way in which we measure productivity: the number of Google hits for each word ending in each suffix. For every word, we do a Google search on the exact term and list the number of hits; we then look for numerical patterns in these numbers to determine productivity.

Using basic regular expression matching, we identified a total of 11966 stems of English words ending in either *-ic* or *-ical* in Webster's Second International dictionary. For each stem, either one or both derivatives was listed in the dictionary. We then ran automated Google queries for words ending in both *-ic* and *-ical* in each of the 11966 stems, storing the results in a database. We then determined for each pair whether *-ic* or *-ical* had more hits; this word was considered the 'winner'. For most stems, a Google search resulted in hits for both *-ic* and *-ical* words (e.g. *historic* and *historical*), while for a relatively small number, only one word had any hits (e.g. *radiosurgical/*radiosurgic* but *overenthusiastic/*overenthusiastical*). Ninety percent of all comparisons yielded a winner by a margin of at least one order of magnitude. Overall, we identified 10613 *-ic* winners vs. 1353 *-ical* winners with an overall ratio of 7.84 in favor of *-ic*. This demonstrates conclusively that, overall, *-ic* is more productive than *-ical*.

Finer-grained analysis, however, reveals a subtler story — namely that *-ical* is potentiating (Williams 1981) within a certain domain. To investigate this, we first sorted all the stems in our database into left-to-right alphabetical neighborhoods of one to five letters, e.g. all stems ending in *-t-* (there are 4166 of these), or all stems ending in *–graph-* (there are 294 of these). We find that, when we sort the words in this way, the only set of words ending in *-ical* with a neighborhood over 100 in size is *–olog(ical)*; for this subset only, *-ical* is the winner over *-ic* (e.g. *psychological* over *psychologic*) by a ratio of 8.30, almost the exact reverse of the ratio of the full set (7.84 in favor of *-ic*). In other words, using this Google search technique, we find that, although overall *-ic* is more productive than *-ical*, the reverse is true for words ending in *-olog(ical)*.

Although *-olog(ical)* forms a large subset (475 members), no other large sets in the system have resisted the overall trend favoring *-ic*. However, the *-olog(ical)* subset is unique in another way: the string 'olog' has a strikingly low number of competitors. For example, there are 79 stems ending in *-rist-*, but 660 stems ending in the substring *–ist-*, and this number jumps to 4166 stem ending in *–t-*. Thus, *-rist-* makes up only 1.9% of stems ending in *-t-*, leaving many similar competitors. We found that, on average, a neighborhood of length 2 (e.g. *-st-*) only accounts for 27.84% of the words in the corresponding neighborhood of length 1 (e.g. *-t-*), while a neighborhood of length 4 makes up just 10.47% of its size-2 set. However, even at length 4, the *-olog-* subset still makes up 66.62% of its length 1 (*-g-*) set. This means that 66% of words ending in *-gic(al)* also end in *-ologic(al)*, which is the strongest of all sets by a wide margin (the next strongest is *-graph-* at 34%).

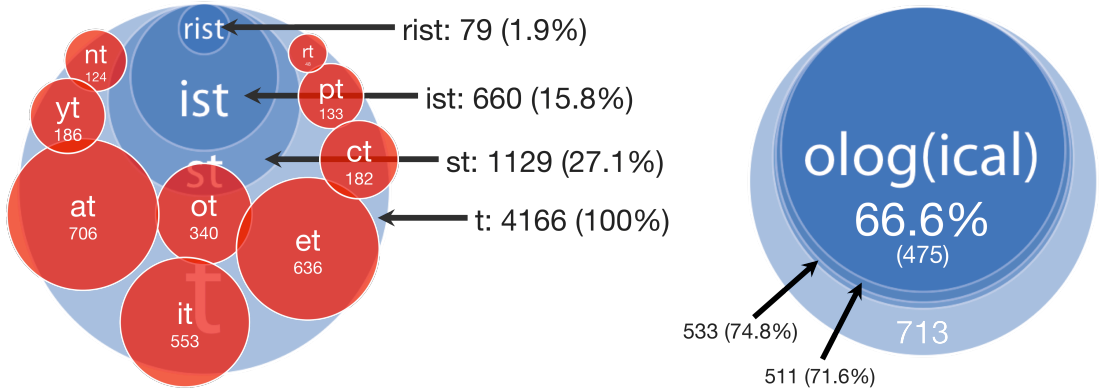


Figure 4 – Competing similar forms in the -rist- neighhorhood vs. the relative uniformity of -olog-

Thus, *-olog(ical)* is a subsystem that is not only sufficiently large, but also has distinctly few competitors, leaving it uniquely suited to sustain itself in spite of patterning inversely to all other *-ic/ical* pairs.

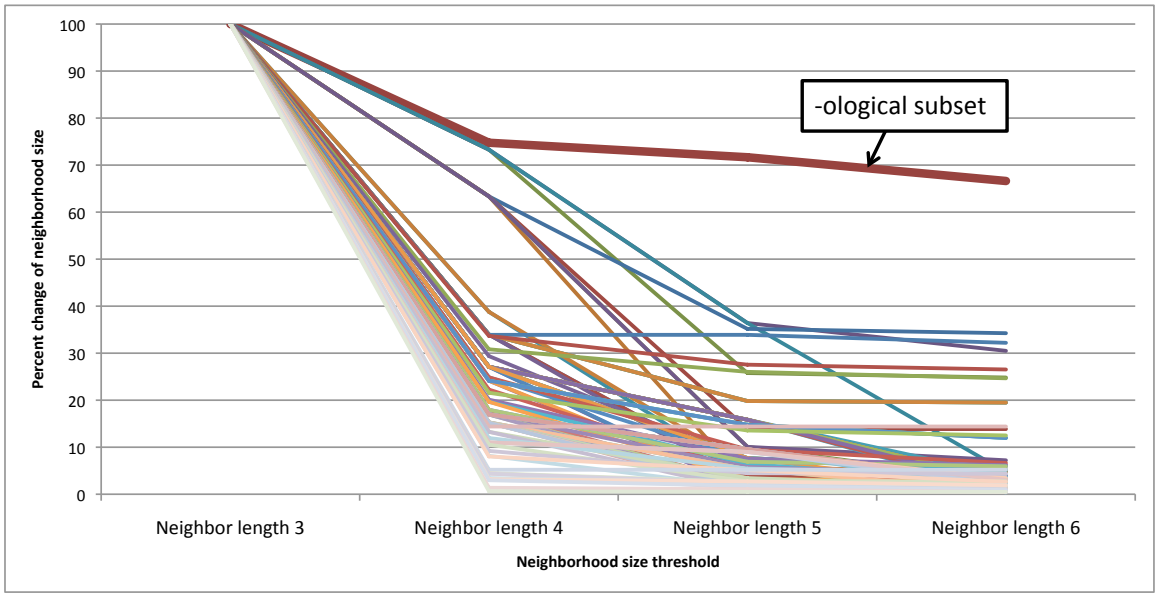


Figure 5 – Change in neighborhood size as a function of threshold length

We know from our historical study of *-ity* that borrowed suffixes in English form emergent self- organizing systems. If words ending in *-ic* or *-ical* formed a simple emergent system, we might predict, based on the overall preponderance of *-ic* words, that this rival would eventually win out and that *-ical* would lose. Instead, we see that a strong regularity, even one that is the reverse of the normal pattern, can develop in a subset if the subset stands out. We expect to explore other such subsets in further studies. For example, we know that *-ity* and *-ness* are rivals and we have a variety of types of evidence showing that *-ness* is more productive, with some evidence that *-ity* is potentiating in a subset of words (words ending in *–ible* and *-able*). We will use the same Google search techniques to explore the set of words ending in either *-ness* or *-ity*, to see whether there are structural commonalities with the set of words that we have explored here. Other areas for future investigation include *-ize* and *-ify*, the triplet *-dom/-hood/-ship* (German *-tum/-heit/-schaft*), and the broader question of suffix ordering in English.

Overall, and somewhat surprisingly, English derivational morphology, especially when it involves the emergence of productive affixes from sets of borrowed words (in which English is especially rich), is a fertile proving ground for the study of self-organizing systems in languages, in part because of the databases that electronic resources provide.

Selected References

Albright, Adam. 2002. The Identification of Bases in Morphological Paradigms. Ph. D. thesis, UCLA.

Anshen, Frank & Mark Aronoff. 1999. Using dictionaries to study the mental lexicon. Brain and Language 68.10.

Baayen, R. Harald. 1993. On frequency, transparency, and productivity. In Booij, G. and J. van Marle (eds.) Yearbook of morphology 1992. Dordrecht: Kluwer. 181-208.

Plag, Ingo. 1999. Morphological productivity : structural constraints in English derivation. Berlin ; New York: Mouton de Gruyter.

Plag, Ingo and Harald Baayen. 2009. Suffix ordering and morphological processing. Language, 85, 106-149.

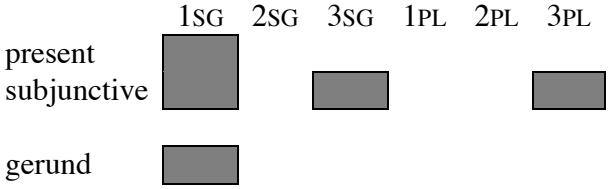
Williams, Edwin. 1981. Argument structure and Morphology. Linguistic Review, 1:81-114.

A 'phonologically conditioned morphome'?
An apparent paradox in the history of Romanian

Martin Maiden. Research Centre for Romance Linguistics, Oxford University

I explore the morphological history of two very similar patterns of root allomorphy in the verb system of Daco-Romance (that branch of Romance principally represented by Romanian), and its implications for notions of 'autonomous morphology'. Both patterns are the result of ancient (and extinct) contextually conditioned sound changes, and share a nearly identical, and highly idiosyncratic, pattern of distribution within the inflectional paradigm of the verb, defined over a heterogeneous set of paradigm cells. The first sound change operated on various consonants, in the environment of a yod, producing the pattern of alternation in (1), which I label, for reasons here unimportant, the 'L-pattern'. The resulting sets of consonantal alternants for modern Romanian are given in (2):

(1) The L-pattern



(2) L-pattern alternant-sets

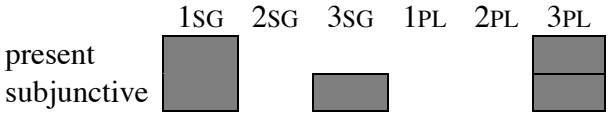
Alternants occurring in L-pattern cells	alternating with
/ts/	/t/
/dz/ (> /z/)	/d/
/j/	/n/
/j/	/l/

(3) Early modern Romanian example of L-pattern verb (N.B. the consonantal alternant of the 2SG in this verb has a different origin):

	1SG	2SG	3SG	1PL	2PL	3PL
present	*vădzu* 'I see'	*vedzi*	*vede*	*vedem*	*vedeţi*	*vădu*
subjunctive	*vădzu*	*vedzi*	*vadză*	*vedem*	*vedeţi*	*vadză*
gerund	*vădzându*					

The second change operated on velars before front vowels, but in almost exactly the complementary set of cells to the first change. It thus resulted in a nearly identical distributional pattern of alternation (which I label, again for reasons here unimportant, the 'U-pattern'), as in (4), and the sets of consonantal alternants in modern Romanian in (5):

(4) The U-pattern



gerund

(5) U-pattern alternant-sets

Alternants occurring in U-pattern cells	alternating with
/k/	/tʃ/
/g/	/dʒ/
/sk/	/ʃt/

(6) Early modern Romanian example of U-pattern verb

	1SG	2SG	3SG	1PL	2PL	3PL
present	*pla*[k]*u* 'I please'	*pla*[tʃ]*i*	*pla*[tʃ]*e*	*plă*[tʃ]*em*	*plă*[tʃ]*eţi*	*pla*[k]*u*
subjunctive	*pla*[k]*u*	*pla*[tʃ]*i*	*pla*[k]*ă*	*plă*[tʃ]*em*	*plă*[tʃ]*eţi*	*pla*[k]*ă*
gerund	*plă*[k]*ându*					

In terms of their paradigmatic distribution, which I argue to be irreducible to any semantic or functional common denominator, both the L-pattern and the U-pattern qualify as 'morphomic' in the sense of Aronoff (1994). They also show the diachronic evidence for the 'psychological reality' of morphomes discussed by Maiden (e.g., 2001a, 2005, 2009), in that it can be shown that their distribution remains robustly intact over most of the observable history of Daco-Romance, and that they are 'coherent': that is, if any morphological change affects the root allomorph in one of the cells implicated, it equally and always affects all of the others in exactly the same way.

I will demonstrate that the original phonological processes have been extinct for at least a millennium but, as examples (3) and (6) show, it remains the case that in U-pattern verbs the alternation is unfailingly correlated with the presence of a following 'non-front' vowel (although final /u/ has largely been deleted in modern Romanian and many dialects), and that the L-pattern alternation also finds itself strongly (although less perfectly) correlated with that same environment. I do not accept that this fact, of itself, means that the patterns may be seen as 'phonologically conditioned', and Maiden (2001b, 2009) argues against claims of this kind (e.g., albeit from very different perspectives, Fanciullo 1998 and Burzio 2004), for cognate and closely similar phenomena in the history of Italo-Romance and Ibero-Romance varieties. But if the substantive evidence from the history of these latter languages argues (I claim) against 'phonologizing' accounts, that from Daco-Romance turns out to be problematically different. Through much of their history in Daco-Romance the two patterns are subject to extensive analogical *extension*, being introduced into verbs where they are not historically justified, and sometimes giving rise to novel alternant-sets, e.g., (7):

(7) 'Coherent' analogical creation of novel U-pattern alternant set in early modern Romanian (the root *ucig-* replaces earlier *ucid-*)

	1SG	2SG	3SG	1PL	2PL	3PL
present	*ucigu* 'I kill'	*ucizi*	*ucide*	*ucidem*	*ucideţi*	*ucigu*
subjunctive	*ucigu*	*ucizi*	*ucigă*	*ucidem*	*ucideţi*	*ucigă*
gerund	*ucigându*					

In such developments, both patterns are very robustly ‘coherent’. In more recent centuries there has been considerable (if always lexically sporadic) analogical ‘levelling out’ of the L/U-pattern alternants. This levelling is often also 'coherent', but there are quite frequent examples of 'incoherence', where not all cells are affected equally. A rather obvious explanation of this could dispense altogether with any notions of ‘morphomic’ structure: extension occurred coherently simply because it operated in a *phonologically specifiable* environment (non-front vowels); levelling was incoherent because it involved introduction of alternant which originated in the phonologically complementary environment, and therefore could not, by its very nature, be distributionally sensitive to ‘non-front vowels’. In short, the whole story could be a matter of the presence or absence of a phonological conditioning environment: no ‘morphome’ at all.

A closer examination of the comparative-historical facts leads, however, to the following, more nuanced and perhaps paradoxical, observation: the L-pattern and U-pattern really are morphomic, but they are much more ‘coherent’ when they are also ‘supported’ phonologically. The ‘purely phonological’ hypothesis will be shown not to work for L-pattern verbs because the alternants systematically fail to appear before non-front vowels in certain morphologically defined circumstances, although in its support it has the fact that the L-pattern alternant is selected before derivational affixes containing non-front vowels. In the case of the U-pattern, I shall adduce some evidence that the alternants may be analogically extended independently of the nature of the ending. The purely 'morphomic' hypothesis looks doubtful to the extent that the history of the L- and U-patterns is often incoherent when it comes to analogical ‘levelling out’ of the alternants. In a significant minority of cases, levelling affects the first person singular present and subjunctive, but not the third person subjunctive, sometimes leaving the latter as the sole remnant of L-pattern alternation. However, at least in western Romanian dialects, the pattern of breakdown between 1SG.PRS/SBJV and 3SBJV turns out to be *structured*, in that loss of the alternant in the latter regularly implies its loss in the former. The evidence is that speakers really do make a generalization across a synchronically arbitrary set of paradigm cells but that, unlike 'canonical' morphomes of the type discussed by Aronoff (1994), and involving patterns of stem identity, what the Daco-Romance data reveal is a probabilistic implicational hierarchy between cells (cf. Blevins 2006:570), along the following lines:

i. If there is a distinctive allomorph in the first person singular present and first person singular subjunctive, then there is a very strong likelihood that the same distinctive allomorph will appear in the third person singular and plural of the subjunctive.
ii. If there is a distinctive allomorph in the third person singular and plural of the subjunctive, there is a strong likelihood that the same distinctive allomorph will appear in the first person singular present and first person singular of the subjunctive.
iii. If there is a distinctive allomorph in the gerund, then there is a likelihood that it will also appear in at least one of: the third persons singular and plural of the subjunctive and first person singular present and first person singular of the subjunctive.
iv. If there is a distinctive allomorph in at least one of the third persons singular and plural of the subjunctive or first person singular present and first person singular of the subjunctive, then there is a likelihood that it will also appear in the gerund.

The foregoing is a set of purely morphological generalizations implicationally linking an arbitrary set of cells. That they form a distinctive set at all is an accident of long-extinct sound changes. What is also clear from the diachronic evidence is that all these implicational bonds are greatly reinforced, in all directions, if there is also a phonologically specifiable cue in the form of a following non-front vowel.

To the extent that a phonologically statable context is implicated in many phases of the history of the Daco-Romance U-pattern and L-pattern, these patterns are not 'classical' morphomes. But then what are they? On the phonological side, I shall show that what is involved is far from 'natural' in universal terms, and crucially has no generality even within the phonological system of Daco-Romance, outside the confines of verb morphology. But on the semantic side the paradigmatic distribution of the alternants does not correspond to any class of features which 'coheres' in a universal sense, or is relevant to any other phenomenon in the synchronic structure of the language. The Daco-Romance L-pattern and U-patterns actually suggest that the 'morphological autonomy' of some alternation pattern can be a matter of *degree*, not an absolute — a possibility that is, as I shall stress, actually strongly implicit in Aronoff's own discussion of the nature of morphomes (1994:25;166;167). In particular, I shall argue for a perspective in which some morphological phenomena are essentially and irreducibly 'autonomous', yet may be 'reinforced' by a recurrently associated phonological environment. The relationship between such facts and the notion of the 'emergence of the unmarked' (Rubach and Booij) will also be considered. I shall briefly review the relevance of my findings to certain other recent claims for phonological, rather than morphomic, interpretations of the distribution of allomorphy (e.g., Anderson 2008).

References

Anderson, Stephen 2008. 'Phonologically conditioned allomorphy in the morphology of Surmiran (Rumantsch)', *Word Structure*, 1, p. 109-134.

Aronoff, Mark 1994. *Morphology By Itself*. Cambridge, MA.: MIT Press.

Blevins, James, 2006. 'Word-based morphology', *Journal of Linguistics*, 42, 531-573.

Burzio, Luigi 2004. 'Paradigmatic and syntagmatic relations in Italian verbal inflection', in Auger, Julie, Clancy Clements, J. and Vance, B. (eds.), *Contemporary Approaches to Romance Linguistics*. Amsterdam: Benjamins, pp. 17-44.

Fanciullo, Franco 1998. 'Per una interpretazione dei verbi italiani a "inserto" velare', *Archivio Glottologico Italiano* 83:188-239.

Maiden, Martin 2001a. 'A strange affinity: perfecto y tiempos afines', *Bulletin of Hispanic Studies* 58:441-64.

Maiden, Martin 2001b. 'Di nuovo sulle alternanze velari nel verbo italiano e spagnolo', *Cuadernos de filología italiana* 8.39-61.

Maiden, Martin 2005. 'Morphological autonomy and diachrony', *Yearbook of Morphology 2004*, 137-75.

Maiden, Martin 2009. 'From pure phonology to pure morphology. The reshaping of the Romance verb'. *Recherches de linguistique de Vincennes* 38:45-82.

Rubach, J. and Booij, G. 2001 'Allomorphy in Optimality Theory: Polish iotation'. *Language* 77:26-60.

Stage level and Individual level Readings of Quality Nouns
Deadjectival Suffixes as Aspectual Disambiguators

Fabienne Martin, Stuttgart University

1. Introduction

-Ité, *-isme*, *-esse*, *ence/ance*, *-(er)ie*, *-ise*, *-itude* and *-ion* are among the suffixes used to derive Quality Nouns (QNs) from evaluative adjectives (*stupid, intelligent*, etc.) in French. Although several in-depth studies have been recently devoted to some of them (cf. e.g. Kerleroux (2008) on *-ion*, Koehl (2009) on *-ité* or Roché (2010) on *-isme*), their morpho-semantic differences have not been investigated since Bécherel (1981, 1976) in French (but see Daude (2002) for Latin and Rainer (1988, 1989) for Italian). Our goal is to compare the aspectual polysemy of the adjectival root and the derived noun. Our main hypothesis is that some of the deadjectival suffixes at least are endowed with some aspectual value, by which they contribute to disambiguate the root on the aspectual side. Let us immediately note that this hypothesis contributes to explain the existence of several doublets of QNs derived from the same root, cf. (1). This would have been *a priori* problematic for the Blocking Effect (Aronoff (1976)) if both forms were functionally identical.

(1) a. coquett**erie**/coquett**isme** (>*coquet* 'coquettish')
b. drôl**erie**/drôl**isme** (>*drôle* 'funny, weird')
c. fanfaronn**erie**/fanfaronn**isme** (>*fanfaron* 'swanky')
d. crapul**erie**/crapul**isme** (>*crapule* 'scoundrel')
e. crétin**erie**/crétin**isme** (>*crétin*'moron')
e. importun**ité**/importun**isme** (>*importun* 'out of place')

The aspectual value of >150 deadjectival nouns are analysed for the study. The adjectives chosen are the first 130 ones in a list of evaluative adjectives ordered alphabetically. In relevant cases, neologisms were identified on the Internet. The created list of 150 QNs includes 45 Ns in *-ité*, 33 Ns in *-ence/ance*, 18 in *-ion*, 17 in *-erie*, 15 in *-isme*, 8 in *-esse*, 7 in *-ie*, 5 in *-ise* and 3 in *-itude*. As expected (cf. Koehl (2009)), *-ité* is the most represented deadjectival suffix.

2. Aspectual polysemy of evaluative adjectives

It is now well-known that *stupid* predicates are aspectually polysemous; besides their use as *individual level predicates* (ILPs), they can be coerced in stage-level ones (SLPs) in some contexts or constructions (cf. Fernald (1999), Barker (2002), Geuder (2000)), cf. e.g. (2).

(2) a. Pierre a joué avec la sourdine. Il a été intelligent./*Peter played with the soft pedal. He was clever*.
b. Feynman était stupide de danser sur la table./*Feynman was stupid to dance on the table.*

Besides, it has been several times independently argued that in some of their SLP uses, *stupid* predicates tend to presuppose the occurrence of an action through which the quality is concretely manifested (cf. Barker (2002), Geuder (2000)). For instance, (2b) presupposes that Feynman danced on the table.

3. Aspectual polysemy of derived QNs

The aspectual polysemy of derived *stupidity* nouns have been comparatively less investigated. Daude (2002) for Latin and Meunier (1981), Beauseroy and Knittel (2007) or Beauseroy (2009) for French distinguish for QNs (i) a (stative) IL reading under which the QN denotes a quality [1] , (ii) an event reading (or e-reading), (iii) a metonymic object-reading (o-reading).

1. Some authors have argued that NPs like *Peter's intelligence* or *Mary's happiness* denote *tropes*, cf. e.g. Moltmann (2004) and Villalba (2009). A problem with this view is that according to the definition of tropes endorsed by these authors, tropes can be placed on temporal and spatial axes. Villalba's example (3a) supports this view for *happiness*. However, at least for French, this type of data can only be extended to a subset of QNs, cf. (3b) vs (3c). I see several solutions to this problem. Firstly, one can distinguish, among QNs, those who denote tropes (compatible with temporal predicates), and those who do not. Secondly, one could give up the idea that tropes are concrete entities (cf. e.g. Campbell (1990) for such a view), and admit that all QNs uniformly denote tropes. In the latter case, one should still explain why some QNs only are compatible with the type of predicates illustrated in (3).

(3) a. John's happiness lasted only one year.
b. *L'intelligence de John a duré toute sa vie./* John's intelligence lasted his whole life.
c. La maladresse de John a duré toute sa vie/John's clumsiness lasted his whole life.

According to our data, the acceptability of QNs with spatio-temporal predicates seems related to the availability of the eventive reading (as described *infra*). We will not adopt a particular view for now and simply assume that QNs denote (particular instances of) qualities/dispositions.

	-erie	*-ie*	*-esse*	*-ise*	*-ité*	*-ance*	*-isme*
[+EV] QNs	100%	71%	62%	40%	25 %	25 %	0%
[+SL] QNs	100%	100%	88%	100 %	68 %	66%	26%

Table 1: Aspectual readings of QNs classified according to the deadjectival suffix used

(4) a. *Pierre est d'une grande gentillesse.* (IL-reading)
Pierre is of-a great kindness
'Pierre is very kind'
b. *Pierre a fait des gentillesses.* (e-reading)
Pierre did sm kindness-PL.
'Pierre did some acts of kindness'
c. *Pierre a apporté des petites douceurs.* (o-reading)
Pierre brought some little delicacies
'Pierre brought some little delicacies'

This typology reflects the description of dictionaries like the *Trésor de la langue française*. The indicators for the e-readings are among others the optionality of argument structure, cf. (8a)-(8b), the verbs *faire* or *commettre* 'commit', cf. (4b) and (8a), and the verb *avoir lieu* 'take place', cf. (8b). [2]

(8) a. *La chatte a commis **une méchanceté**.*
The cat committed a malicious-TÉ (=malicious act).
b. ***Ces bêtises** ont eu lieu dans la cuisine.*
These stupid-ISE-s (=stupid acts) took place in the kitchen.

3.1. A temporary stative reading

Differentiating three readings is not enough though, since among the [−EV] QNs like *discrétion* or *intelligence* incompatible with the e-reading indicators *faire/commettre* or *avoir lieu* (cf. (9)), some like *discrétion* are still acceptable in environments which force them to denote a temporary eventuality (i.e. are [+SL]). Among the environments imposing the SL reading to the QN are the demonstrative following an event description in (10) and the temporary adverbial in (11).

(9) **Pierre a fait une discrétion/intelligence.*
Pierre did a discrete-ION/intellig-ENCE. (intended: discrete/intelligent act)

(10) *Pierre n'a rien dit au comité. Cette discrétion/*cette intelligence m'a étonné.*
Pierre didn't say anything to the committee. This discrete-ION/this intellig-ENCE surprised me.

(11) *Sa discrétion/# son intelligence de ce matin m'a beaucoup étonné.*
His discrete-ION/intellig-ENCE of this morning surprised me a lot.

Given these facts, I will assume the existence of two temporary (stage-level) readings for QNs, an eventive one (the e-reading), cf. (9), and a stative one (the SL s-reading), cf. (10)-(11). Under the SL s-reading, QNs describe the quality as temporarily delimited by the act which manifests it. Note that while the adjective *intelligent* has a SL s-reading (cf. (2a)), the derived QN *intelligence* does not (cf. (9)-(11), but see the comments about (14) below). This supports the hypothesis that the nominalisation process contributes to aspectually disambiguate the base.

3.2. The polysemy of quality nouns and their suffixes

A first indication that deadjectival suffixes play a role in the aspectual interpretation of QNs is that the range of their possible readings significantly varies with the suffix at hand. The first line of Table 1 gives the percentage of [+EV] QNs (QNs which accept the e-reading) for each type of suffixes. The second one gives the percentage of [+SL] QNs, i.e. QNs which accept a (stative or eventive) SL reading.
Although these results should be confirmed on a larger set of QNs and by a larger set of speakers [3] to be fully

2. The interpretation of *beaucoup* 'a lot' also differs when the e-reading is available (cf. e.g. Obenauer (1983) and Doetjes (2002) on the polysemy of *beaucoup*). With [+EV] QNs like *tendresse*, *beaucoup* can have a quantitative reading paraphrasable by *souvent*, cf. (5), hence the possibility to have (6). With [−EV] QNs like *intelligence*, the intensive reading is the only one available, cf. (7). With these nouns, it is impossible to build consistent discourses of the type (6).

(5) *Pierre a fait montre de beaucoup de tendresse.*
Pierre showed much tenderness.
i. Pierre showed tenderness many times.
ii. Pierre showed a great tenderness

(6) *Pierre fait montre de beaucoup de tendresse, mais n'en donne qu'un petit peu à la fois.*
Pierre shows much tenderness, but gives only a bit at once.

(7) *Pierre a fait montre de beaucoup d'intelligence.*
Pierre showed much intelligence.
i. #Pierre showed intelligence many times.
ii. Pierre showed a great intelligence.

Given the difference illustrated through these examples, we disagree with Flaux and Van de Velde (2000) or Beauseroy (2009) according to which the quantity reading of determiners like *beaucoup* is never available with QNs.

3. For most nouns, we used the *TLF*, Frantext and the Internet to check our intuitions of native speaker.

convincing, they still strongly suggest that (i) it is useful to differentiate two types of SL readings as proposed above (stative or eventive), and (ii) it is neither true that permanency is equally exhibited by any type of QNs as suggested by Daude 2002:229 for Latin, nor that all QNs are aspectually underspecified on the lexical level as claimed by Beauseroy 2009:129 for French. The data presented in Table 1 suggest on the contrary that deadjectival suffixes used to form QNs differ from each other by their range of aspectual values. More particularly, *-isme* seems the only suffix to almost univocally convey permanency once attached to QNs (and the related suffix *-iste* also appears to univocally select the IL reading of its root, cf. *infra*). Two additional facts support this claim. First of all, (12a) and (12b) below do not have the same truth conditions: the speaker of (12a) might find that Pierre was an idiot at one occasion only (thanks to the present perfect on *être*), while the one of (12b) seems committed to the belief that Pierre is permanently stupid. [4] The same way, there is no entailment from (13a) to (13b): being one time in favour of X does not entail that I am a X-ist, because 'X-ist' implies that I am permanently in favour of X, hence the oddness of the present perfect in (13b) (which can be accommodated once admitted that Pierre is a versatile person). [5]

(12) a. *Pierre a été crétin. Ca m'a étonné.*
Pierre was stupid. It surprised me.
b. *Le crétinisme de Pierre m'a étonné.*
The stupid-ISME of Pierre surprised me.

(13) a. *Ce matin, Pierre a été favorable à la suspension.*
This morning, Pierre was in favour of the suspension.
b. *Ce matin, Pierre a été suspensionniste.*
This morning, Pierre was suspension-ISTE.

(14) *# Son infantilisme de ce matin m'a étonné.*
His childish-ISME of this morning surprised me.

Secondly, when *isme*-Ns are combined with temporal punctual adverbials like *ce matin*, instead of denoting a 'temporary disposition' like in (11), they tend to refer to a (possibly permanent) *subkind* of disposition. For instance, (14) does not imply that *x* was temporarily childish, but rather that *x* manifested a subkind of childishness this morning.

3.3. Factors at play for the emergence of the e-reading

The availability of the e-reading depends on several interacting factors of morphological, semantic and pragmatic nature. Firstly, although QNs are derived from an adjectival **root**, the e-reading is almost automatically obtained when the derivational family of the root includes an agentive verb, which could then possibly be the **base** of the derived noun (see Roché (2010) on the distinction between the root and the base of a derived word). For instance, while the adjectival root *extravagant* has the agentive verb *extravaguer* in its derivational family, *exubérant* is only related to the rare and largely unknown verb *exubérer*, which is, besides, unaccusative; as (15) shows, *extravagance* only has an e-reading, although both QNs are built with the same deadjectival suffix.

(15) a. *Pierre a commis une extravagance*.
Pierre committed an extravag-ANCE (=an extravagant act)
b. **Pierre a commis une exubérance*.
Pierre committed an exuber-ANCE.(intended: an exuberant act)

Secondly, when the derivational family of the adjectival root contains no verb, the e-reading is *ceteris paribus* easier to get when the QN denote an 'abnormal', 'salient', 'marked' disposition, cf. the contrasts in (16) below. This seems due to the fact that a disposition is more easily conceived as temporally delimited to an event and gets more easily the cognitive salience of an event when it is *abnormal*.

(16) a. *Pierre a fait/commis une indiscrétion/*une discrétion*.
Pierre did an indiscrete-ION/discrete-ION (=indiscrete/discrete act)
b. *Pierre a fait/commis une indélicatesse/#une délicatesse*.
Pierre did an insensitive-ESSE/sensitive-ESSE (=insensitive/sensitive act)
b. *Pierre a fait/commis une bêtise/*une intelligence*.
Pierre did a stupid-ISE/intellig-ENCE (=stupid/intelligent act)

Thirdly, on the relevant e-reading, several QNs behave like *pluralia tantum* (even when they denote 'abnormal' dispositions). For instance, contrary to *bêtise* 'stupidity', *courtoisie* 'courtesy/kindness' only gets the e-reading once pluralised, cf. the contrast in (17).

(17) a. *Pierre lui a fait ?une courtoisie/OK des courtoisies.*
Pierre committed a courteous-ITÉ/some courteous-ITÉ-s (=a/some courteous act(s))
b. *Pierre a fait ?une frivolité/OK des frivolités.*
Pierre committed a frivolous-ITÉ/some frivolous-ITÉ-s (=a/some frivolous act(s))
c. *Ils ont commis ?une cruauté/OK des cruautés*.
Pierre committed a crual-ITÉ/some crual-ITÉs (=a/some crual act(s))

4. Note that the entailment would go through with a [+SL] noun like *amabilité*, which confirms that no other factor intervenes in the construction we have in (12).

5. Among the three models of *-isme* derivations distinguished by Roché (2010), QNs systematically instantiate Model 3 (the 'be' model), and nouns like *suspensionnisme* 'prosuspension' instantiate Model 1, in which N-*isme* is based on the predicative axiological relation *to be favorable to N* (for instance, *esclavagisme* 'proslavery' is the property of being in favour of slavery).

The fact that plural introduces countability certainly contributes to explain why the plural is sometimes compulsory for the e-reading; events are countable, states are not. But this does not explain in which case exactly it has to be used for the e-reading to be accessible. Again, this varies with the suffix. None of the Ns in *-ance* or *-erie* require the plural to get the e-reading, contra 60% of the Ns in *-esse* or 66% of the Ns in *-ité*. At least for *-erie*, data can be explained by the fact that the suffix is already frequentative/collective by itself (cf. e.g. Dubois (1962), Lieber (2004)).

Interestingly, QNs which do not have an e-reading, like *isme*-QNs in general or other [-EV] QNs like *intelligence*, cannot pluralise under their non-generic readings. For instance, (18a) cannot be paraphrased as in (18b). This is expected if these QNs are inherently stative, since states are massive and thus non-countable. The only available reading is the (often pragmatically weird) subkind reading illustrated in (18c).

(18) a. ?*Les infantilismes/intelligences de Pierre*.
Peter's childishnesses/intelligences.
b. ≠ *les manifestations d'infantilisme de Pierre*
The manifestations of Pierre's childishness
c. = ?*Les différents types d'infantilisme/d'intelligence de Pierre.*
The different types of Pierre's childishness.

4. Conclusions

French deadjectival suffixes differ from each other by their aspectual value. In the doublets presented in the introduction (1) for instance, QNs in *isme* have a strong bias towards the IL reading while the corresponding QNs in *erie* can easily have a (stative or eventive) SL one. It is interesting to note that most of the doublets we found until now consist of *erie*-QNs vs *isme* ones. Given the fact that the *erie/isme* suffixes present the biggest differences in terms of aspectual values (cf. Table 1), the prevalence of this kind of pairs can be explained by the fact that they exhibit the greatest degree of semantic dissimilarity.

References

Aronoff, M. (1976). *Word Formation in Generative Grammar*. MIT Press, Cambridge, Mass.

Barker, C. (2002). The Dynamics of Vagueness. *Linguistics and Philosophy*, 25:1–36.

Beauseroy, D. (2009). *Syntaxe et sémantique des noms abstraits statifs : des propriétés verbales ou adjectivales aux propriétés nominales*. PhD thesis, Université de Nancy.

Beauseroy, D. and Knittel, M.-L. (2007). Nombre et détermination. le cas des noms de qualité. *Revista di linguistica*, 2:231–262.

Bécherel, D. (1976). *La dérivation des noms abstraits en français*. PhD thesis, Université Paris-Sorbonne.

Bécherel, D. (1981). Différenciation morpho-sémantique des suffixes nominalisateurs de l'adjectif. *Cahiers de Lexicologie*, 38:45–59.

Campbell, K. (1990). *Abstract Particulars*. Blackwell, Oxford.

Daude, J. (2002). Les substantifs abstraits de qualité. In Kircher-Durand, C., editor, *Grammaire fondamentale du latin, t. IX. La création lexicale : la formation des noms par dérivation suffixale*, pages 225–305. Peeters, Leuven/Paris.

Doetjes, J. (2002). Comparing Adverbs of Quantity. Manuscript, Utrecht University.

Dubois, J. (1962). *Étude sur la Dérivation Suffixale en Français Moderne*. Nathan, Paris.

Fernald, T. (1999). Evidential Coercion: Using Individual-level predicates in Stage-level Environments. *Studies in the Linguistic Sciences*, 29/1:43–63.

Flaux, N. and Van de Velde, D. (2000). *Les noms en français. Esquisse de classement*. Ophrys, Paris.

Geuder, W. (2000). *Oriented Adverbs*. PhD thesis, Universität Tuebingen.

Kerleroux, F. (2008). Des Noms indistincts. In Fradin, B., editor, *La Raison Morphologique. Hommage à la Mémoire de Danielle Corbin*, pages 113–132. John Benjamins, Amsterdam.

Koehl, A. (2009). Are French *-ité* Suffixed Nouns Property Nouns? In Montermini, F., Boyé, G., and Tseng, J., editors, *Selected Proceedings of the 6th Décembrettes*.

Lieber, R. (2004). *Morphology and Lexical Semantics*. Cambridge University Press, Cambridge.

Meunier, A. (1981). *Nominalisations d'adjectifs par verbes supports*. PhD thesis, Université Paris 7.

Moltmann, F. (2004). Properties and Kind of Tropes: New Linguistic Facts and Old Philosophical Insights. *Mind*, 123/1:1–41.

Obenauer, H.-G. (1983). Une quantification non. canonique: la quantification à distance. *Langue française*, 58:66–88.

Rainer, F. (1988). Towards a Theory of Blocking: the Case of Italian and German Quality Nouns. In van Maarle, J. and Booij, G., editors, *Yearbook of Morphology*. Foris Publications, Dordrecht.

Rainer, F. (1989). *I nomi di* qualità *nell'italiano contemporaneo*. Braumüller, Vienna.

Roché, M. (2010). Quel traitement unifié pour les dérivations en *-isme* et en *iste*? In Roché, M., Boyé, G., Giraudo, H., ad S. Lignon, N. H., Montermini, F., and Plénat, M., editors, *Des unités morphologiques au lexique*. Presses Universitaires de Vincennes. to appear.

Villalba, X. (2009). Definite adjective nominalizations in spanish. In Espinal, M. T., Leonetti, M., and McNally, L., editors, *Proceedings of the IV Nereus International Workshop 'Definiteness and DP Structure in Romance Languages*. Universität Konstanz. Arbeitspapier 12x Fachbereich Sprachwissenschaft.

Irregularity *and* decomposability of inflected verbal forms

Fanny Meunier[1], Delphine Fabre[1], François-Xavier Alario[2]
contact : fanny.meunier@univ-lyon2.fr

[1] *Laboratoire Dynamique du Langage (CNRS& Université Lyon 2, Lyon, France)*

[2] *Laboratoire de Psychologie Cognitive (CNRS & Aix-Marseille, Marseille, France)*

In this article we investigated the issue of stem representations of French inflected verbs in the lexicon by the means of a lexical decision task and the manipulation of the cumulative and surface frequencies. Briefly stated, there are two major psycholinguistic hypotheses about the representation of polymorphemic words in the mental lexicon. Some models postulate that morphology is not explicitly represented in the lexicon. This is for example the case of the whole-word representation hypothesis (e.g. Butterworth, 1983; Manelis & Tharp, 1977), which posits that all words have a unitary representation that does not take into account their internal morphological structure. This class of models also includes connectionist models, in which morphology is not represented *per se*, but rather in terms of weighted connections between distributed whole-word orthographic, phonological and semantic representations (e.g. Rumelhart & McClelland, 1986; Seidenberg & McClelland, 1989; Sereno & Jongman, 1997). Words belonging to the same morphological family tend to share some or much of their orthographic and phonological forms, as well as their semantics. Models with whole word representations capture this regularity by postulating (or implementing learning algorithms that lead to) links that are stronger between the representations of these related words than between words that are solely orthographically or phonologically related. The alternative models postulate morpheme-decomposed representations. In this type of models, polymorphemic words are stored in the lexicon in terms of their morphemic constituents (Taft & Forster, 1975). Morphology is explicitly represented in the sense that access to a word like *painter* is mediated by the activation of the stem representation for *paint-* and the suffix representation for *-er*. Consequently, polymorphemic words do not have their own lexical entry; rather, they share a stem representation with all the words from the same morphological family.

One profitable experimental method to test these contrastive hypotheses has been to measure the influence of the frequency of the representations postulated by the different models on polymorphemic word recognition, as measured for instance in the classic lexical decision task. The frequency of whole-word representations, referred to as *surface frequency*, is

defined as the word's frequency of occurrence as a free-standing lexical item. The frequency of a morphological component such as the stem can be defined as the sum of the frequencies of all the affixed words that carry that stem (e.g. for the stem *paint*: the surface frequency of *paint* + the surface frequency of *painter* + the surface frequency of *unpainted* + etc). This measure, referred to as *cumulative frequency*, captures the hypothesis that each time a morphologically complex word comprising the stem is used, the stem representation is activated. The manipulation of these variables in a lexical decision paradigm allows testing the role played by whole-word and stem representations during the recognition of polymorphemic words.

Using priming paradigm it has been shown that English *regular* inflected verb form primes the recognition of its stem (e.g., *walked* - *walk*). By contrast, no priming effect (or a priming significantly smaller than that obtained in the case of regular verbs) is observed between *irregular* inflected forms and their stem (e.g., *drove* - *drive*; see for example Kempley & Morton, 1982; Napps, 1989; Stanners, Neiser, Hernon, & Hall, 1979). This result has also been interpreted as evidence for a decomposition procedure of the prime, a procedure that would only operate with regular forms. In this interpretation, processing the prime *walked* leads to its decomposition and to the activation of the stem *walk.* In the case of English irregular verb forms, no priming effect is observed, thus no decompositional procedure is hypothesized. This interpretation suggests that lexical access for irregular verbs can only be achieved through their whole-word representation. In French, Meunier and Marslen-Wilson (2004) reported two verbal priming experiments in French, one involving masked priming and the other cross-modal priming. In masked priming, the prime is presented visually for a very brief duration followed by a pattern mask (e.g. random alphanumeric characters). In this situation, the prime word is not consciously detected by the participants (Forster & Davis, 1984; Forster, 1998). In cross-modal priming, the prime is presented auditorily while the target is presented visually (see for example Marslen-Wilson, Tyler, Waksler & Older, 1994). Meunier and Marslen-Wilson (2004) reported results that were rather different from the English results described above (see also Orsolini & Marslen-Wilson, 1997, for similar results in Italian, a Romance language that has an inflectional system close to French). They found similar priming effects across regular, irregular, and morpho-phonological verb targets. In other words, the priming observed between irregular pairs such as *buvais* – *boire* ('you drunk' – 'to drink') and morpho-phonological pairs such as *sèment* – *semer* ('they sow' – 'to sow') did

not differ from the priming observed between regular pairs such as *mangeons* – *manger* ('we eat' – 'to eat').

The priming effect observed in French with the regular class of verbs can be explained with the same rationale that was used to account for the results observed with English regular verbs. Following Marslen-Wilson and Zhou (1999), the priming effect observed with the morpho-phonological class of verbs can be explained with the hypothesis that the stem alternations of the morpho-phonological variants of a verb share an abstract representation. By contrast, the priming effect observed for irregular verbs is more difficult to account for. This is because the stem included in a prime like *buvais* overlaps only minimally with the target *boire*. Thus if *buvais* is decomposed, the remaining stem *buv-* do not overlap with the target stem *boir-* as in the case of regular verbs (e.g. *mangeons* – *manger*). Consequently, the representation of irregular verbs in French remains unresolved.

In the experiments reported, we attempted to clarify the representation of French verb stems. To address this issue, we used a simple lexical decision task where we manipulated the surface and the cumulative frequencies of the verbs.

We began by testing regular verbs which have only one stem. We manipulated the cumulative frequency effect by selecting pairs of verb forms that have similar surface frequencies but whose stems have different cumulative stem frequencies (for example *mangeas* 'you have eaten' and *grimperas* 'you will climb' have the same surface frequency but different cumulative frequencies). We also manipulated the surface frequency. We expected to observe a cumulative frequency effect. In the logic developed earlier, such effect would be suggestive of stem activation during the processing the inflected verb forms.

In the second experiment, we investigated the morpho-phonological class of verbs, to test whether allomorphic variants activate an underspecified stem (following Marslen-Wilson and Zhou, 1999). For this purpose, we selected verbs whose two surface forms were contrasted in cumulative stem frequency. For example, the two stems of the verb *semer* have respectively a high cumulative frequency (*sem-*) and a low cumulative frequency (*sèm-*). We contrasted verb forms that had low cumulative frequency and were contrasted on surface frequency (e.g. *sème* vs. *sèmerai*), as well verb forms that had high cumulative frequency and were contrasted on surface frequency (e.g. *semais* vs. *semée*). If representations are shared and phonologically underspecified, no cumulative frequency effect should be observed. By contrast, if the two variations of the stem have distinct representations, we predict a cumulative frequency effect similar to that predicted for regular verbs.

Finally, we tested the irregular class of verbs. For this purpose, we selected verbs whose two idiosyncratic stems were contrasted in cumulative frequency. For example the two stems of the verb *boire* have respectively a high cumulative frequency (*boir-*) and a low cumulative frequency (*buv-*). If irregular inflected forms have a single shared representation for irregular and canonical stems, then we should not observed a cumulative frequency effect. By contrast, if there are distinct representations for the two stems, we should observe a pattern of results similar to that expected for regular verbs, namely a cumulative frequency effect.

Our results revealed that the surface and cumulative frequencies influence the recognition of inflected verbal forms for the three classes of verbs. This suggested that, in French, all verbal forms are decomposed during visual recognition, be they regular or not. This leads to the idea that morphological decomposition is trigged by the decomposability of verbal forms more than by their regularity *per se*.

Références

Butterworth, B. (1983). Lexical representation. In B.Butterworth (Ed.), *Language production*, vol.2, (pp. 257-294). San Diego, CA: Academic Press.

Manelis, L., & Tharp, D. (1977). The processing of affixed words. *Memory and Cognition, 5,* 690-695.

Rumelhart, D.E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In D.E. Rumelhart & J. L. McClelland & P.R. Group (Eds.). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume2. Psychological and biological models* (pp. 216-271). Cambridge, MA: MIT Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

Sereno, M. S., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory and Cognition, 25*, 425-437.

Taft, M., & Forster, K.I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behaviour, 15*, 638-647.

Kempley, M., & Morton, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology, 73*, 441-454.

Napps, S. (1989). Morphemic relationships in the lexicon: Are they distinct from semantic and formal relationships? *Memory and Cognition, 17*, 729-739.

Stanners, R.F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Behavior, 18*, 339-413.

Meunier, F., & Marslen-Wilson, W. (2004). Regularity and irregularity in French verbal inflection. *Language and Cognitive Processes, 19(4),* 561-580.

Forster, K.I. (1998). The pros and cons of masked priming. *Journal of Psycholinguistic Research, 27*, 203-233.

Forster, K., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 680-698.

Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological review, 101*, 3-33.

Orsolini, M., & Marslen-Wilson, W. (1997). Universals in morphological representation: Evidence from Italian. *Language and Cognitive Processes 12 (1)*, 1-47.

Marslen-Wilson, W., & Zhou, X. (1999). Abstractness, allomorphy and lexical architecture. *Language and Cognitive Processes, 14 (4)*, 321-352.

Adjectival bases of French *-aliser* and *-ariser* verbs: syncretism or under-specification?

Fiammetta Namer – UMR 7118 ATILF – Nancy-Université – MSH Lorraine
fiammetta.namer@univ-nancy2.fr

Issue: This paper follows a previous study, proposed in (Namer, 2009), and is about the constructed meaning of 227 deadjectival verbs ending with *-aliser* or *-ariser* (hereafter *-aRiser*). Although the verb base is always an adjective, the verb itself seems to instantiate three constructional patterns: some are genuinely deadjectival (1), others have a denominal interpretation (2), and in other cases the adjectival base bears an anaphoric function, namely it refers to a « Noun Adj » phrase, in which Adj matches the verb base (3). According to the context, a single verb form may have more than one reading (4). The examined 227 verbs come either from the TLF or from online documents[1]. The investigation aims at determine the conditions in which these various verb meanings are observed. It leads us to questioning the nature of the morphological processes that form *-aRiser* verbs. Our assumption is that the verb is formed on a semantically underspecified base; the verb's interpretation pattern, involving a noun, an adjective or a noun phrase, is revealed only in context.

(1) UNIVERSALISER : rendre UNIVERSEL - *Que faire pour **universaliser** l'accès à l'eau*
[UNIVERSAL-IZE: make [sth] become UNIVERSAL: What to do to **universalize** access to water]

(2) LOCALISER : déterminer le LIEU - *On a **localisé** le naufragé*
[LOCAL-IZE: determine [sth's] PLACE: we have **localized** the shipwreck survivor]

(3) MURALISER : faire une [fresque] MURALe - *Lorsque l'on évoque le sujet des fresques murales, Serge Berkowicz se met tout de suite à en parler avec passion. Ce sentiment de "créer" de laisser une "trace" visible, sont autant d'éléments qui le motive à **muraliser***
[MURAL-IZE: make a MURAL [fresco] - When mural fresco matter is advocated, SB starts immadiately to talk about it with passion. This impression of « creating » of letting a visible « trace », are as many elements that motivate him to **muralize**]

(4) AMICALISER : rendre AMICAL - *Les cadeaux ne servent qu'à "**amicaliser**" les tribus non-amicale, dans le but de faire des alliances*
AMICALISER : ajouter à sa liste d'AMIS - ***Amicalise**-moi sur Facebook*
[FRIENDLY-IZE : make [sb] become FRIENDLY: Gifts just help friendly-ize unfriendly tribes, with the intention of forming alliances
FRIENDLY-IZE : add [sb] to a FRIENDS list (=to friend): **Friendly-ize** me on Facebook]

Preliminaries : French *-iser* suffix is generally known as a verbalizer, i.e. verb forming suffix selecting a noun or an adjective as a base. When they are noun based, verbs derivatives express a whole range of concepts, since the verbal predicate can be causative, resultative, locative, inchoative, performative, etc.: many authors agree to assign several interpretative patterns to the *-iser* suffixed verbs (as well as to the corresponding suffixed verbs in other languages, e.g. the English *-ize*). Following Lieber and Plag terminology, respectively in (Lieber, 2004 :71-89) and (Plag, 2003 :118), *-ize* is polysemous ; In Plag opinion, cf. (Plag, 1999 :121-194), *-ize* (or *-iser*) derivational pattern verbal outputs are semantically underspecified. Moreover, still according to (Plag, 1999), the expected meaning of a new coined *-ize* suffixed verb results from many interactions, among which the base semantic features, pragmatic factors, and the weight of the attested lexicon. The French *-iser*

[1] Google and Yahoo searches have been performed through the WaliM robot (Namer 2003), between February and April 2010. Native spelling and spelling errors have been left untouched in utterances given as examples.

suffixation rule alternates with other derivation rules, such as *-ifier*, *en-*, *a-*. Mechanisms that rule this competition are examined from a semantic viewpoint in (Roger, 2003). The morphophonological perspective is accounted for English in (Plag, 1999), and for French in (Lignon, 2010).

A fact seems to emerge from the various studies: when the base is an adjective, the *-iser* suffixed verb is either causative or unaccusative, and it expresses change-of-state affecting the patient, whose final state is characterized by the property described by the adjective base. In other words, this adjective, which refers to a property, has to be a predicate (cf. (Wisniewski and Love, 1998) for more details).

Syncretism: In this talk we are disputing this claim. Actually, TLF records 904 *-iser* suffixed verbs. The base of a hundred of them is a denominal adjective suffixed with *-al, -el* or *-aire*. More than 50 of these adjectives, corresponding to 5,5% of the 904 *-iser* verbs set, bear a strictly relational meaning (Fradin, 2008) (5). Correspondingly, the meaning of the derived verb directly depends from that of the noun the base adjective is derived from (6):

(5) SCOLARISER < SCOLAIRE = relatif à l'école
[SCHOOL-AR-IZE < SCHOOL-AR (academic, educational) = relating to school]
INSTRUMENTALISER < INSTRUMENTAL = relatif à un instrument
[INSTRUMENTAL-IZE < INSTRUMENTAL = relating to an instrument]

(6) SCOLARISER = mettre dans une école (*rendre scolaire)
[SCHOOL-AR-IZE = Send [sb] to SCHOOL (*make sth/sb become academic, educational)]
INSTRUMENTALISER = transformer en instrument (*rendre instrumental)
[INSTRUMENTAL-IZE = turn [sth] into an INSTRUMENT (*make sth/sb become instrumental)

The so-called authentic corpora (newspapers, articles, forums, blogs, etc.) gather *-aRiser* verbs absent from the TLF and having similar characteristics:

(7) CATÉGORIALISER = faire des CATÉGORIES - *Il faut une faculté capable de procéder à un tri, de sélectionner, de* ***catégorialiser***.
[CATEGORICAL-IZE = make CATEGORIES: A faculty is needed, capable of sorting, selecting, **categorical-ize**] (see also (Sablayrolles, 2007))

The first, spontaneous explanation to the discrepancy between form and meaning observed with verbs such as in (5)-(7) is that they define cases of syncretism, following the same line of argument as the one (Koehl, 2009) or (Dal and Namer, 2010) hold to account for the *-ité* nouns formation. A. Koehl (2009) remarks that, although *-ité* suffixed nouns are usually property adjectives, some of these nouns are denominal, and the *-ité* derived noun refers to a measure: the base noun surfaces as the adjective forms it is morphologically related to. For instance, the semantic base of MORTALITÉ (MORTALITY) is the noun MORT (DEATH) under the form /mɔʁtal/ borrowed from the adjective MORTEL (MORTAL). In (Dal and Namer, 2010) the reasoning is quite similar: for ethnic property nouns, the default base is an adjective, e.g. ITALIEN > ITALIANITÉ (ITALIAN > ITALIAN-ITY). However, in given circumstances, the stem of the morphologically related toponym replaces the adjective. So, PORTUGUAIS > PORTUGALITÉ (PORTUGUESE > PORTUGAL-ITY) instead of the unattested *PORTUGUAISITÉ (PORTUGUESE-ITY). For (Koehl, 2009) and (Dal and Namer, 2010), substitution is purely formal: the base word diverts to its own profit the default stem of another word belonging to its morphological family (a similar approach is used in (Booij, 1997) for the analysis of Dutch toponym/ethnonym/adjective triplets).

All in all, adopting the syncretism solution turns to assume that, for instance, INSTRUMENTALISER is based on the noun INSTRUMENT, surfacing as the suppletive stem /ɛ̃stʁymɑ̃tal/ borrowed from the related adjective INSTRUMENTAL.

At first sight this solution is seducing, because she solves the semantic mismatch issue between the verb and its apparent base. So, it accounts for the fact that SCOLARISER does not mean *makes sth become academic/educational*, and that the online attested DÉMONIALISER (DEAMON-AL-IZE) can only be defined from the noun DÉMON (DEMON):

(8) « *Comment ne pas **démonialiser** des gens qui prescrivent des amphétamines aux petits enfants* »
[How not to **demon-al-ize** people who prescribe amphetamine to little children]

At the same time, this hypothesis offers additional evidence about the syncretic nature of *-al/-el* and *-aire*. The near-complete lack of X*iser* forms competing with X*aRiser* in TLF also argue in favour of the syncretism solution. The form taken by the nominal base X of a verb X*aRiser* systematically matches the stem of the corresponding X*al*, X*el* or X*aire* denominal adjective, from the moment that this adjective exists.

Polysemy : However, this solution is insufficient. In fact, it does not explain why, for a large number of TLF X*aRiser* verbs, the same verb form may have either a denominal meaning or a deadjectival one, depending on the context in which it occurs:

	VERB	Verb meaning according to its BASE	EXAMPLE
a	PROVINCIALISER (PROVINCIAL-IZE)	Assign sth to the PROVINCE_N	*Ne faudrait-il pas plutôt **provincialiser** ces compétences ?* Should not we rather **provincial-ize** these juridictions?
b		(Make sth) Become PROVINCIAL_A	*La langue des écrivains s'est créolisée, **provincialisée**, barbarisée* Writers language creolized, **provincial-ized**, barbarized
c	ARTÉRIALISER (ARTERIAL-IZE)	turn sth (= a blood vessel) into ARTÈRE_N (ARTERY)	*Il est préférable **d'artérialiser** la veine radiale* It is better to **arterial-ize** the radial vein
d		(Make sth) Become ARTÉRIEL_A(ARTERIAL)	***Artérialiser** le sang veineux, c'est en modifier l'oxygénation* To **arterial-ize** venous blood means to modify its oxygenation

Table1: X*aRiser* denominal and deadjectival readings

We could put forth the hypothesis that there are two lexemes PROVINCIALISER (Table 1, a,b). The former would be defined with respect to the noun PROVINCE, which refers to the final destination of the verb patient referent. The latter would be adjective-based and describe the transition leading the patient to the property of 'being PROVINCIAL'. The same line of argument would also apply to ARTÉRIALISER (Table1, c,d). One of the verb readings describes

a process consisting in "turn a blood vessel into an artery" (c). The other one refers to the fact of "providing sth with arterial features" (d).
However, assuming homomorphic X*aRiser* verb pairs is not a satisfying solution: in fact, for all X*aRiser*, it presumes two identical verb forms being systematically constructed from two different bases, this device being implemented despite the fact that other better suited morphological means are available to avoid polysemy. For instance, instead of ARTÉRIALISER in (c), the non-existing ARTÉRIFIER (ARTERIFY) could have been coined, following the model of the attested MOMIFIER (MUMMIFY): "turn sth into a MOMIE (MUMMY)".

Our analysis: Our assumption is that the meaning of these particular verb forms is decided within speech utterances, and that the adjective category, which is always the selected base, is semantically underspecified. This hypothesis, close to that what is proposed in (Roché, 2008), and going against the conclusions stated in (Fradin and Kerleroux, 2003), has been tested with a set of deadjectival *-aRiser* verbs with a denominal base: 95 TLF-attested verbs, and 132 other ones searched from the Internet have been examined in detail. The results we came across are the following:

- In most cases, the verb unique interpretation is either denominal (9), or deadjectival (10). The verb meaning reflects the adjective nature, which is either strictly predicative (HORIZONTAL) or relational (FLUVIAL).

(9) FLUVIALISER < FLUVIAL=FLEUVE - *Entrepôts estimés intéressants car susceptibles de* ***fluvialiser*** *leur flux*.
[fluvial-ize < fluvial=river: Warehouses considered interesting because likely to **fluvial-ize** their flow]

(10) HORIZONTALISER < HORIZONTAL ≠ HORIZON - *En augmentant sa vitesse de nage le nageur va s'****horizontaliser*** *grâce à l'accroissement de la force de portance*
[HORIZONTAL-IZE < HORIZONTAL ≠ HORIZON: By raising its swimming speed, the swimmer is going to **horizontal-ize** itself by increasing lift strength]

- Just like what Table1 shows, two readings are sometimes possible for a given verb; the detection of each reading is context depending. The verb polysemy reveals the twofold nature (predicative/relational) of its base adjective:

(11) COLONIALISER < COLONIAL = qui rappelle les colonies - *Besoin d'idées pour* ***"colonialiser"*** *mon vieux buffet*
[COLONIAL-IZE < COLONIAL = that reminds colonies: Need ideas to '**colonial-ize**' my old sideboard]
COLONIALISER < COLONIAL = COLONIE - *Ba tu les defendrais moins si ils venaient* ***colonialiser*** *la France*
[COLONIAL-IZE < COLONIAL = COLONY: Well you would'nt defend them this way if they came to **colonial-ize** France]

- Moreover, the inspection of the Web data reveals another trend: the verb adjectival base corresponds to the adjectival modifier occurring in a Noun-Adjective collocation, which is sometimes found in the speech occurring before the *-aRiser* verb. For instance, in (12), the adjectival base TERRITORIAL of TERRITORIALISER (TERRITORIAL-IZE) does not refer to TERRITOIRE (TERRITORY) but to *collectivité territoriale* (*territorial collectivity* = region with a measure of autonomy). In (13), the BANCARISER (BANK-AIRE-IZE) base is the adjective BANCAIRE (BANK/BANKING), referring to the phrase *compte bancaire* (*bank account*): indeed

BANCARISER's meaning is neither "put sth in a bank" nor "make sth become bank-related", but "provide sb with a bank account". In example (12) (resp. (13)), the *-aRiser* verb adjectival base plays in a way an anaphoric role, since it is co-referential with the nominal phrase *collectivité territoriale* (resp. *compte bancaire*), occurring in the previous sentence.

(12) *méconnue et peu utilisée surtout en collectivité territoriale. **....** elle **territorialise** la plupart de ses services, crée des équipes*
[neglected and not much used, especially in a territorial collectivity … it **territorial-izes** most of its departments, creates teams]

(13) *Le produit est un compte bancaire « light » et accessible à tous et qui contribuera à « **bancariser** » les tunisiens*
[The product is a "light" bank account, affordable to anyone and that will help to "**bank-aire-ize**" Tunisian citizens]

- Finally, the anaphoric interpretation is for some verbs one of the two possible readings: for instance the meaning of RELATIONNALISER (RELATIONAL-IZE) in (14) is "having relationships", whereas in (15) the adjectival base RELATIONNEL (RELATIONAL) refers to the phrase *trouble relationnel* (*relational trouble*). Consequently, the verb paraphrase in (15) is "become a relational trouble":

(14) *Sur Second Life c'est pareil. Et même pire. Je **relationnalise** aisément, mais ça s'arrête souvent là.*
[On Second Life it is the same. Even worse. I **relational-ize** easily, but it is often the end of the matter]

(15) *… ni celle du graphologue : un trouble instrumental peut se « **relationnaliser** », une difficulté relationnelle peut s'» instrumentaliser »*
[… nor that of the graphologist: an instrumental trouble may "**relational-ize**" itself, a relational issue may "instrumental-ize" itself]

Questions: The following questions arise from the data analysis, and (part of) them will be given a tentative answer during the talk.
- Is it possible to predict the possible X*aRiser* interpretation(s)? At a first sight, the denominal or deadjectival verb reading depends on the predicative/relational adjective nature
- Are the verb denominal readings correlated to a restricted number of interpretative schemata, or do these nouns play any of the semantic roles generally observed with genuine denominal *-iser* verbs (e.g. locatum, instrument, result, process, etc.)?
- What is the exact nature of the base unit in *-aRiser* verbs? First, this unit is probably not a semantically specified lexeme, because of the nominal/adjectival reading ambiguity we have illustrated in (9-11). Moreover, it may be the case that it is even not a lexeme, as shown in (12-15). Thus, could it be that the base under-specification is not only semantic, but even categorial, for *-aRiser* verbs?

Another interesting issue requires further investigation to be answered: is the verb behaviour consistent with that of its morphological family (X*aRiser*, *dé*X*aRiser*, *re*X*aRiser*), e.g. (INSTRUMENTALISER [INTRUMENTAL-IZE], DÉSINSTRUMENTALISER [UN-INSTRUMENTAL-IZE], RÉINSTRUMENTALISER [RE-INSTRUMENTAL-IZE])? In other words, does the same interpretation variation occur for X*aRiser* and for its related prefixed verbs? If we were able to give a positive answer to this question, this would provide a further confirmation to the impact of paradigm in morphology, as claimed in (Hathout, à paraître).

Bibliography

Booij, Geert. 1997. Autonomous morphology and paradigmatic relations. *Yearbook of Morphology* 1996:35-53.

Dal, Georgette, and Namer, Fiammetta. 2010. French property nouns based on toponyms or ethnic adjectives: A case of base variation. In *Variation and Change in Morphology*, eds. Franz Rainer, Wolfgang U. Dressler and Hans Christian Luschützky. Amsterdam: John Benjamins:53-74.

Fradin, Bernard, and Kerleroux, Françoise. 2003. Quelles bases pour les procédés de la morphologie constructionnelle ? In *Silexicales 3 : les unités morphologiques*, eds. Bernard Fradin et al. Villeneuve d'Ascq: Presses Universitaires du Septentrion:76-84.

Fradin, Bernard. 2008. On the semantics of Denominal Adjectives. Online Proceedings of the *6th Mediterranean Morphology Meeting (MMM6)*, Ithaca, Greece.

Hathout, Nabil. à paraître. Une analyse unifiée de la préfixation en anti-. In *Des Unités Morphologiques au Lexique*, ed. Michel Roché. Paris: Hermès.

Koehl, Aurore. 2009. Are French /ite/ suffixed nouns property nouns? In *Morphology in Bordeaux - Selected Proceedings of Decembrettes 6*, eds. Fabio Montermini and Gilles Boyé. Somerville, MA: Cascadilla Proceedings Project: 95-110.

Lieber, Rochelle. 2004. *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.

Lignon, Stéphanie. 2010. Les suffixations en -iser et en -ifier : vérifier les données pour vériser les hypothèses ? In: *Décembrettes 7*. Toulouse.

Namer, Fiammetta. 2003. WaliM : valider les unités morphologiquement complexes par le Web (3ème Forum de Morphologie, 20-21 septembre 2002, Lille). In *Silexicales 3 : les unités morphologiques*, eds. Bernard Fradin et al. Villeneuve d'Ascq: Presses Universitaires du Septentrion:142-150.

Namer, Fiammetta. 2009. *Morphologie, Lexique et TAL : l'analyseur DériF*: TIC et Sciences cognitives. London: Hermes Sciences Publishing.

Plag, Ingo. 1999. *Morphological productivity. Structural constraints in English derivation*. Berlin/New York: Mouton de Gruyter.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

Roché, Michel. 2008. Structuration du lexique et principe d'économie: le cas des ethniques. *1er Congrès Mondial de Linguistique Française*, Paris:1559-1573.

Roger, Coralie. 2003. De la pertinence de la notion de paradigme pour les procédés de construction des verbes de changement d'état (3ème Forum de Morphologie, 20-21 septembre 2002, Lille) In *Silexicales 3 : les unités morphologiques*, eds. Bernard Fradin et al. Villeneuve d'Ascq: Presses Universitaires du Septentrion:179-187.

Sablayrolles, Jean-François. 2007. A propos de quelques métanéonymes. *Cahier du CIEL* 2007-2008:105-115.

Wisniewski, Edward J., and Love, Bradley C. 1998. Relation vs properties. *Journal of Memory and Language* 38:177-202.

Bare N(ominal)N(ominal) Concatenations in Turkish: Compounds or Syntactic Fallacies?

Ralli, A[*]., Bagriacik, M., & Andreou, M.

University of Patras
ralli@upatras.gr, mbagria@upatras.gr, andreoum@upatras.gr

Bare NN concatenations in Turkish, such as TAŞ DUVAR 'stone wall', TAHTA KAFA lit. *wood head* 'idiot' or DEMİR EL lit. *iron hand* 'squeezer', have long been cited as compounds (Kornfilt 1997; Göksel & Kerslake 2005; Gökdayı 2007) even though there is a disagreement on whether these concatenations are AN or NN compounds (see Kornfilt (1997) for a NN view, and Gökdayı (2007) among others for arguments in favor of a AN structure).

In this paper, we argue that: a) many of these NN concatenations are not true compounds, b) the real reason for the disagreement about the grammatical category of their internal constituents results from accepting the existence of two different nominal categories, nouns and adjectives, and c) NN concatenations are distinguished into three types, compounds, constructs, and mere noun phrases, all of them members of a continuum.

Following Braun & Haig (2000), we argue that in Turkish there are no distinct categories of nouns and adjectives, but rather a set of nominals, the members of which belong to a categorial continuum: '(more) nouny', 'no-pref(erence)', '(more) adjective-like'. In order to determine where a nominal stands on this continuum, we use five tests, three of them (3-5) have been proposed by Braun & Haig (2000): 1) modifiability, 2) *bir* insertion, 3) suffixation with *-lI* and *–sIz*, 4) gradability, and 5) intensifying reduplication. Assuming that the lexical items *kara* 'black', *taş* 'stone', and *duvar* 'wall' are adjective-like, no-pref. and nouny, respectively, they respond to these tests as follows:

1. Modifiability: As can be seen by the examples (1a-c) modification is leftwards, i.e. *adjective-like + no-pref*; *no-pref + nouny*; *adjective-like + nouny*:

(1)a. KARA TAŞ	b. TAŞ DUVAR	c. KARA DUVAR
black stone	*stone wall*	*black wall*
'black stone'	'stone wall'	'black wall'

Rightward modification is bound to the emergence of a suffix *–{(s)I(n)}* on the right periphery, which, according to Kornfilt (1997), Göksel & Kerslake (2005), and Ralli (2008) is a compound marker (COMP):

(2)a. *DUVAR TAŞ	b. DUVAR TAŞ -I	c. TAŞ KARA -SI
wall stone	*wall stone -COMP*	*stone black-COMP*
?	'wall stone/stone for wall'	'stone black'

2. *bir* insertion: *[X bir Y]* '[a(n) X Y] as X: While adjective-like and no-pref items can occur in the position X, which is the modifier of Y, nouny items cannot:

[*] corresponding author.

(3)a. KARA BİR TAŞ	b. TAŞ BİR DUVAR	c. *DUVAR BİR TAŞ
black one stone	*stone one wall*	*wall one stone*
'a black stone'	'a stone wall'	?

3. Suffixation with 'relational (REL)' *{-lI}* and 'privative (PRV)' *{-sIz}*: Adjective-like items do not allow combination of a REL or a PRV suffix, whereas such suffixation is grammatical for no-pref and nouny items:

(4)a. *KARA-LI	b. TAŞ -LI	c. DUVAR-LI
black-REL	*stone-REL*	*wall-REL*
'with/in black'	'with stone(s)'	'with wall(s)'
d. *KARA-SIZ	e. TAŞ -SIZ	f. DUVAR-SIZ
black-PRV	*stone-PRV*	*wall-PRV*
'without black'	'without stone(s)'	'without wall(s)'

4. Gradability: While adjective-like items can be graded with *daha* 'more' or *en* 'the most', nouny items cannot. Although questionable, non-pref items can be acceptable:

(5)a. DAHA KARA	b. ?DAHA TAŞ	c. *DAHA DUVAR
more black	*more stone*	*more wall*
'more black'	?	?
d. EN KARA	e. ?EN TAŞ	f. *EN DUVAR
the most black	*the most stone*	*the most wall*
'the most black'	?	?

5. Intensifying reduplication: While the meaning of adjective-like items can be intensified by a reduplicated syllable (INT), nouny items do not allow this type of intensification. No-pref items with reduplication are questionable, but can be acceptable:

(6)a. KAP-KARA	b. ?TAP-TAŞ	c. *DUP-DUVAR
INT-black	*INT-stone*	*INT- wall*
'dark/black as a pitch'	'all in stone'	?

As a result of the five tests described above, adjective-like items fulfill the criteria (1,2,4,5), nouny items fulfill the criterion (3), and no-pref items fulfill almost all criteria with a certain degree of ease. Therefore, concatenations cited as NN or AN compounds can be analyzed as combinations of the following types:

(7)a. [[no-pref][nouny]]	b. [[adj-like][nouny]]	c. [[adj-like][no-pref]]
TAŞ DUVAR	YENİ YIL	KARA TAHTA
stone wall	*new year*	*black board*
'stone wall'	'new year'	'blackboard '

As for the remaining combinations, i.e. [[nouny][nouny]], [[adj-like][adj-like]], and [[no-pref][no-pref]], they are not productively built in Turkish.

The question that arises now is whether there are true NN compounds in Turkish. We will claim that this category also exists, the members of which display a semantic non-compositionality or semi-compositionality, and have the following

characteristics, as illustrated in (8): a) internal structure negative to insertion, c) no *hangi* ('which') substitution of the non-head, d) no non-head ellipsis, and e) no transformation of the non-head as a modificational phrase [*X olan*…] '[the … which is (made of) X]':

(8)

	True Compounds		
CATEGORIES	**[[adj-like][nouny]]**	**[adj-like][no-pref]**	**[[no-pref][nouny]]**
	KABA KULAK *rough ear* 'mumps'	KARA DUL *black widow* 'black widow'	TAHTA KAFA *wood head* 'idiot'
TESTS			
Element Insertion	*KABA BİR KULAK *rough one ear* ?	*KARA BİR DUL *black one widow* ?	*TAHTA BİR KAFA *wood one head* ?
Substitution with *hangi* 'which'	-HANGİ KULAK? *which ear?* -* KABA KULAK *rough ear* ?	HANGİ DUL? *which widow?* -* KARA DUL *black widow* ?	-HANGİ KAFA? *which head?* -* TAHTA KAFA *wood head* ?
Ellipsis of the non-head	* KABA KULAK VE EL *rough ear and hand* ?	*KARA DUL VE TAHTA *black widow and board* ?	* TAHTA KAFA VE MASA *wood head and table* ?
Modificational phrase	* KABA OLAN KULAK *rough be-ADJ ear* ?	*KARA OLAN DUL *black be-ADJ widow* ?	* TAHTA OLAN KAFA *wood be-ADJ head* ?

We will further argue that compounds differ from other NN concatenations, which could also be distinguished into two different categories: *constructs* (Borer 1988, 2009) and *mere NPs*. In order to define the properties of the last two categories, we will compare them with compounds with respect to compound-proper properties according to the tests described in (8). The results of such a comparison are summarized in the following table:

(9)

	COMPOUNDS	**CONSTRUCTS**	**NPs**
	DEMİR EL *iron hand* 'squeezer'	KURŞUN KALEM *lead pen(cil)* 'pencil'	TAŞ DUVAR *stone wall* 'stone wall'
TESTS			
Element insertion	*DEMİR BİR EL *wood one hand* ? while, BİR DEMİR EL *one iron hand* 'a squeezer'	?KURŞUN BİR KALEM *lead one pen(cil)* 'a pencil' while, BIR KURŞUN KALEM *one lead pen(cil)* 'a pencil'	TAŞ BİR DUVAR *stone one wall* 'a stone wall' and, BİR TAŞ DUVAR *one stone wall* 'a stone wall'
Substitution with *hangi* 'which'	-HANGİ EL? *which hand?* -* DEMİR EL *iron hand* ?	-HANGİ KALEM? *which pen(cil)?* -KURŞUN KALEM *lead pen(cil)* 'pencil'	-HANGİ DUVAR? *which wall?* -TAŞ DUVAR *stone wall* 'stone wall'

Ellipsis of the non-head	* DEMİR EL VE MASA *iron hand and table* ?	*KURŞUN KALEM VE PARA *lead pen(cil) and coin* *'pen(cil) and coin which are (made of) lead'*	TAŞ DUVAR VE BINA *stone wall and building* 'stone wall and (stone) table'
Modificational phrase	* DEMİR OLAN EL *iron be-ADJ hand* ?	KURŞUN OLAN KALEM *lead be-ADJ pencil* 'the pencil which is (made of) lead'	TAŞ OLAN DUVAR *stone be-ADJ wall* 'the wall which is (made of) stone(s)

In accordance with Borer (1988, 2009) and Ralli & Stavrou (1998), who have identified the existence of constructs in Hebrew and Greek, respectively, it seems that their internal structure is visible to syntax only to a certain degree. Insertion of an element does not lead to ungrammaticality but to a slight change in meaning. The non-head of constructs can be substituted by *hangi* 'which', and transformed to a modificational phrase, but the ellipsis of the non-head triggers ungrammaticality. In contrast, the internal structure of NPs is totally visible to syntax, as they themselves are syntactic creations.

Finally, we propose that NN constructions are parts of a continuum, where the two poles are occupied by *compounds* and *NPs*, while *constructs* are placed in between. The internal structure of *compounds* is totally opaque to syntax, the structure of *constructs* exhibits a relative visibility to syntax, and that of *NPs* is fully accessible to syntactic operations. The existence of a continuum can also apply to nominals, the members of which range from more-nouny to more adjective-like, resolving the problems which are posed by accepting two radically distinct categories, nouns and adjectives.

Selected references

Bağrıaçık, M. (2010). *Onomatika sintheta stis Altaikes Ghlosses* (In Greek) [Nominal Compounding in Altaic Languages]. Unpublished MA Thesis. University of Patras.

Borer, H. (1988). On the morphological parallelism between compounds and constructs. In G. Booij, & J. van Marle (Eds.), *Yearbook of Morphology 1988* (pp. 45-65). Dordrecht: Kluwer.

Borer, H. (2009). Afro-Asiatic, Semitic: Hebrew. In R. Lieber & P. Stekauer (eds.) *The Oxford Handbook of Compounding* (pp. 491-511). Oxford: OUP.

Braun, F. & Haig, G. (2000). The noun/adjective distinction in Turkish: An empirical approach. In *Studies on Turkish and Turkic Languages: Proceedings of the Ninth International Conference on Turkish Linguistics*, A. Göksel & C. Kerslake (eds.). Wiesbaden: Horrossowitz, 85-92.

Gökdayı, H. (2007). Takısız ad tamlamaları (In Turkish). [Nominal compounds without a marker]. In L. Karoly (Ed.), *Turcology in Turkey; Selected Papers* (Paper selected by N. Demir and E. Yılmaz). *Studia Uralo-Altaica*, *47*, (pp. 243-252). Szeged: SzTE BTK Altjisztikai Tanszék.

Göksel, A. & Kerslake, C. (2005). *Turkish. A Comprehensive Grammar*. London: Routledge.

Kornfilt, J. (1997). *Turkish*. London: Routledge, Descriptive Grammars Series.

Ralli, A. (2008). Compound Markers and Parametric Variation. 2008. *Language Typology and Universals (STUF), 61*, 19-38.

Ralli, A. & Stavrou, M. (1998). Morphology-syntax interface: A-N compounds vs. A-N constructs in Modern Greek. In G. Booij, & J. van Marle (Eds.), *Yearbook of Morphology 1997*, (pp. 243-264). Dordrecht: Kluwer

On the structure of reduplicants: Iconicity and preferred form in reduplication

Thomas Schwaiger (Karl-Franzens-Universität Graz)[1]

Cases of total reduplication perhaps most clearly demonstrate the special nature of the reduplication process in morphology. Nevertheless, considering the many attested forms of partial reduplication, they are by far not the only reduplicative patterns found in the languages of the world. Furthermore, these rather well known facts have time and again been supplemented by a range of data which indicate that the definitional systematic repetition of phonological material within a word for morphological purposes (cf. Rubino 2005a: 11) can often be countered by various dissimilatory tendencies that seem to avoid exact and/or too long sound sequences of the same kind within a reduplicated word form. Some examples follow:

Vietnamese, an Austro-Asiatic tone language, shows full segmental reduplication which at the same time very often displays differences on the tonal level: e.g. *to* 'large' – *to*∼*to* 'rather large' vs. *xám* 'gray' – *xam*∼*xám* 'grayish' (cf. Nguyen 1997: 27), where in the latter case a high/mid level tone contrasts with a high/mid rising tone after fully reduplicating the simplex form which carries the latter of the two tones. The Austronesian language Tawala forms its durative aspect by either an initial disyllabic reduplicant, an initial CV reduplicant, or an initial V.C[2] reduplicant, the respective form depending on the complex interaction between the structure of the simplex form on the one hand and a set of markedness constraints[3] prohibiting identical adjacent syllables on the other hand (see Hicks Kennard 2004). Finally, the reduplication patterns found in languages like Pangasinan and Chamorro are particularly interesting in the present context. In these two Austronesian languages one occurring type of reduplication does not strictly focus on the first (or sole) syllable of a simplex form but rather extracts a CV sequence as the reduplicant from it, ignoring all following or intervening consonants and thus forming an optimal syllable in the sense of Vennemann 1988: e.g. Pangasinan *plato* 'plate' – *pa*∼*pláto* 'plates' (cf. Rubino 2005a: 11) and Chamorro *dos* 'two' – *man-a-do*∼*dos* 'two-by-two'[4] (cf. Topping 1973: 168).

It is hypothesized that

> the process of copying phonological material in reduplication is accompanied by a tendency to optimize its resulting structure, i.e. the reduplicant.

This hypothesis stems from the observations and the study of the reduplication systems of a number of genetically and typologically diverse languages with the help of a typological online database.[5] A sample of 25 languages, including their genetic affiliation and a very condensed overview of the reduplication types (i.e. forms and functions) they possess, is given in the

[1]E-mail address: *thomas.schwaiger@edu.uni-graz.at*

[2]The period between the vowel and consonant symbols represents a syllable boundary.

[3]The analysis referred to here (i.e. Hicks Kennard 2004) is conducted within the framework of Optimality Theory.

[4]The two prefixes preceding the reduplicant are irrelevant for the present purpose.

[5]The database in question is the Graz Database on Reduplication *(gdr)*, accessible via http://reduplication.uni-graz.at/.

appendix. Owing to the greatest familiarity with the reduplication systems of the languages in question it is this small pilot sample from which the initial hypothesis has been derived and from which the conclusions further below will be drawn. Nevertheless, a first quick glance upon the larger dataset of the *gdr* seems to support the ideas described here.

The term *optimize* in the hypothesis stated above needs some elaboration: Reduplication is perceived as a genuine morphological process in the present context, albeit one with the special characteristic of being formally totally dependent on the particular simplex form that undergoes the process. Consequently, and especially in the light of the many possible varieties of partial reduplication (as opposed to straightforward total reduplication) that can be observed, chances are good that in describing the phenomenon functional motivations of morphological structure become intertwined with tendencies of reduction and simplification in phonology. It is proposed that the morphological process of reduplication should be divided into a morphological level and a phonological level. The former concerns all forms of full reduplication for it is certain units of morphology which are targeted here as bases (words, stems, roots, affixes, etc.). The latter concerns the many forms of partial reduplication ranging from foot reduplicants to single vowel or consonant reduplication. Obviously, the term *optimization* cannot mean the same thing on these two levels of description. Being concerned with the relationship between form and meaning (cf. Kiyomi 1995: 1152), it is suggested that on the morphological level the principle of iconicity is governing the configuration of reduplication systems; hence it is expected that full reduplications should always target linguistic units like words, stems, or roots[6] before, for example, targeting affixes. Indeed, affix reduplication is a rather rare phenomenon (e.g. Fijian *vanua* 'country' – *vei-vanua* 'various countries' – *vei∼vei-vanua* 'large number of countries'; Schütz 1985) and never seems to occur in a language that does not also show full reduplication of lexical units.[7] This is hardly surprising given the fact that the iconic consecutive (plurality, repetition/continuation) and cumulative (intensity) functions of reduplication (see Kiyomi 1995) are semantically more compatible with more concrete lexemes than with more abstract affixes. On the other hand, the phonological level seems to be driven by principles that are concerned with structure alone and do not care about the meaning of reduplication or the latter's relationship with its respective form. So far almost every language showing partial reduplication seems to necessarily possess some kind of CV pattern (the optimal syllable as hinted at above), this also being the case in languages like Classical Greek and Latin, where traces of a former productive reduplication process are found exclusively in such a pattern. To conclude, the optimization tendencies are concerned with optimal symbolization of linguistic categories on the morphological level and with optimal segmental and suprasegmental structures on the phonological level of reduplication (i.e. full and partial reduplication, respectively).

Some interesting consequences for reduplication research in historical linguistics and language typology arise from an investigation along the lines sketched above: Diachronic and cross linguistic work on the reduplication phenomenon over the past decades has yielded the common assumptions that partial reduplication always develops out of full reduplication and

[6] The delimitation of these units is a long standing problem in linguistics and no attempt will be made to settle it here.

[7] Affix recursions as in Italian *tavolinino* 'table:DIM:DIM' or French *il l'a re-re-répété* 'he has re-re-repeated it' are not counted as instances of genuine morphological reduplication in this paper.

that in no synchronic state of a language is there (productive) partial reduplication without full reduplication (see especially Bybee et al. 1994: 166–174, Niepokuj 1997, and the WALS map on reduplication in Rubino 2005b). Strictly adhering to these claims is tantamount to ignoring important structural facts about reduplication which manifest themselves in individual languages as well as across languages over and over again. A few languages in the sample below show partial reduplication but no full reduplication. Although the possibility remains that this is simply a matter of incomplete data, the less compliant way of doubting the infallibility of established views is chosen here. And not only the synchronic typological facts seem to opt for such a treatment; work like Hurch & Mattes 2005 has already shown that not all partial reduplication necessarily stems from a fully reduplicated predecessor diachronically. It is maintained here that while certain basic points of view on reduplication are definitely in need of modification and/or relaxation, functional and structural principles as the ones alluded to above can be adduced to give a systematic typological characterization of reduplication that is not arbitrary and which allows for important linguistic generalizations.

Appendix[8]

language (genetic affiliation)	full / partial RED	functions
Acoma (Keres)	stem / CV	PL, LE
Alamblak (Sepik-Ramu)	word, stem, root / –	PL, INT
Arapesh (Torricelli)	root / CV	INT
Bagirmi (Nilo-Saharan)	stem / CV	PL, INT
Berber, Middle Atlas (Afro-Asiatic)	– / C	PL
Burmese (Sino-Tibetan)	root / –	WCD
Cha'palaa (Barbacoan)	word / CV	PL, WCD
Chamorro (Austronesian)	– / CV, V	PL, INT, WCD
Chukchi (Chukotko-Kamchatkan)	stem / σ	CAS
Hmong Njua (Hmong-Mien)	word / –	PL, INT
Koasati (Muskogean)	– / CV	PL
Lavukaleve (East Papuan)	word / CV, F	PL, INT ...
Malagasy, Plateau (Austronesian)	root / CV, σ, F	PL, DIM, LE
Maung (Australian)	root / CV	PL
Maybrat (West Papuan)	stem / VC	INT
Ngiyambaa (Australian)	root / F	PL, DIM
Oromo, Eastern (Afro-Asiatic)	word / σ	PL, INT, WCD
Réunion Creole Fr. (Fr.-based Creole)	word, compound / –	PL, INT, DIM, WCD
Saisiyat (Austronesian)	stem / CV, σ	PL, DIM, LE, WCD ...
Somali (Afro-Asiatic)	word / σ, VC	PL
Swahili (Niger-Congo)	full / –	PL, INT, DIM, LE ...
Tiwi (Australian)	– / CV	PL
Tukang Besi (Austronesian)	stem / CV, F, V	PL, INT, DIM, LE
Vietnamese (Austro-Asiatic)	word / –	PL, INT, DIM
Yoruba (Niger-Congo)	word, compound / CV, F	PL, INT, DIM, WCD

[8]Uncommon abbreviations: CAS = case, LE = lexical enrichment, WCD = word class derivation.

Bibliography

Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.

Hicks Kennard, Catherine. 2004. Copy but don't repeat: The conflict of dissimilation and reduplication in the Tawala durative. *Phonology* 21. 303–323.

Hurch, Bernhard & Veronika Mattes. 2005. Über die Entstehung von partieller Reduplikation. In Gertraud Fenk-Oczlon & Christian Winkler (eds.), *Sprache und Natürlichkeit: Gedenkband für Willi Mayerthaler*, 137–156. Tübingen: Gunter Narr Verlag.

Kiyomi, Setsuko. 1995. A new approach to reduplication: A semantic study of noun and verb reduplication in Malayo-Polynesian languages. *Linguistics* 33. 1145–1167.

Nguyen, Dinh-Hoa. 1997. *Vietnamese: Tieng Viet Khong Son Phan* (London Oriental and African Language Library 9). Amsterdam & Philadelphia: John Benjamins Publishing Company.

Niepokuj, Mary. 1997. *The development of verbal reduplication in Indo-European* (Journal of Indo-European Studies Monograph 24). Washington: Institute for the Study of Man.

Rubino, Carl. 2005a. Reduplication: Form, function and distribution. In Bernhard Hurch (ed.), *Studies on reduplication* (Empirical Approaches to Language Typology 28), 11–29. Berlin & New York: Mouton de Gruyter.

Rubino, Carl. 2005b. Reduplication. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 114–117. Oxford: Oxford University Press.

Schütz, Albert J. 1985. *The Fijian language*. Honolulu: University of Hawaii Press.

Topping, Donald. 1973. *Chamorro reference grammar* (PALI Language Texts: Micronesia). Honolulu: University of Hawaii Press.

Vennemann, Theo. 1988. *Preference laws for syllable structure and the explanation of sound change: With special reference to German, Germanic, Italian, and Latin*. Berlin et al.: Mouton de Gruyter.

Selkup denominal adjectives: a Generalized Paradigm Function analysis

Andrew Spencer, University of Essex
Irina Nikolaeva, School of Oriental and African Studies, London

Selkup (Uralic) has a great variety of morphological devices for creating transpositions from one lexical category to another without necessarily changing the lexical meaning at all. Kuznecova et al. (1980) explicitly demarcate this aspect of the grammar in their description under the heading of 'representation' ('reprezentacija'). This paper deals with 'adjectival representations of nouns', which behave as attributive adjectives, much like English *polar* from *pole* (e.g. *polar icecap*). However, in Selkup these retain significant traces of their nominal past and behave in many respects like inflected forms of nouns, with or without additional lexical content. These are therefore neither derivational nor inflectional in the usual senses of the terms and represent a poorly described type of lexical relatedness.

Selkup nouns share the general structure of Uralic nouns in having three suffix position slots, for number ([Number:{singular, dual, plural and collective}]), possessor agreement ([PossAgr:{person/number}]) and case ([Case:{nominative, accusative, genitive, instrumental, caritive, translative, coordinative, dative-allative, illative, locative, elative, prolative, vocative}]). A typical example of a fully inflected noun is shown in (1):

(1) qoː-iː-nyt-kɔːlyk
leader-PL-2PL.POSS-CAR
'without your (Pl) leaders (Pl)'

The three features are paradigmatic, i.e. the values of [Number], [PossAgr], [Case] are mutually exclusive.

In Paradigm Function Morphology (Stump, 2001) the inflectional paradigm of a lexeme is given by its Paradigm Function (PF). This takes pairing of <root/stem form, lexemic index> and a full set of features defining that word form (σ) and delivers the inflected form corresponding to that feature set, for all the permissible combinations of features applicable to that lexeme. The PF is defined in part in terms of sets of realization rules (RR), with the general form (2):

(2)

RR maps	to
<X, σ>	<Y, σ>
specific set of feature values (e.g Num:pl)	
lexeme class (e.g. N, V, …)	

where X, Y are (perhaps partially inflected) forms of the lexeme (X and Y may be identical).

The position class property is coded by organizing the RRs into Blocks. Each block corresponds to a position class and contains a realization rule that applies if its feature content is nondistinct from the feature content, σ, of the full word form. Only one realization rule per block can apply and this guarantees the paradigmatic relationship. For instance, we can propose the rules in (3) to represent the form given in (1) (with obvious simplifications):

(3)

a.	I:	Num(X, {Pl})	= Xiː
b.	II:	PossAgr(X, {2pl})	= Xnyt
c.	III:	PossAgr(X, {Caritive})	= Xkɔːlyk

Given the root *qoː* as our starting point these rules allow us to infer the form in (1) for the paradigm cell defined by the feature set {LEADER [Num: Pl, PossAgr: 2pl, Case: Caritive]}, as shown in (4):

(4) a. I: Num(qoː, {Pl}) = qoːiː
b. II: PossAgr(qoːiː, {2pl}) = qoːiːnyt
c. III: PossAgr(qoːiːnyt, {Caritive}) = qoːiːnytkɔːlyk

Default zero-marked forms such as ‘Num:sg’, ‘PossAgr:unpossessed’ lack their own RR and the form is defined by the Identity Function Default (IDF).

In addition to these clearly inflectional forms, there are three major ‘adjectival representation of nouns’. These are denominal forms derived by suffixation:

(5)

associative representation	kana-l′	‘dog’s, pertaining to dogs’
similitudinal representation	alako-ššal′	‘similar to a boat’
locative representation	mɔːt-qyl′	‘located in the/a house’

The adjectival representations of nouns serve for attributive modification. Unlike canonical inflection, similitudinal and locative representations adds semantic content to the noun denotation, essentially creating a representation of the form SIMILAR_TO(N) and LOCATED_IN(N). The semantics of these predicates means that such a word will denote a property as well as denoting an object. Note that derived adjectives with a similar meaning are widespread in the languages of the world, cf. English *boy-ish*, *god-like* and Tundra Nenets *war°-xi°*‘(located) on the shore’ < *war°* ‘shore’. The associative representation is similar to relational adjectives in English and other languages: it is a transposition from the noun class to the adjective class without the addition of a semantic predicate to the content or semantic representation of the lexeme. It denotes some loosely (contextually/pragmatically) defined association between two entities.

However, Kuznecova et al. (1980) make a clear distinction between true adjectives and adjectival representations of nouns in terms of their morphosyntax. The two types are similar in that both can function only as modifiers and do not differ in their external distribution, but adjectival representations are analyzed as part of the nominal paradigm (and hence, are in a sense ‘inflectional’). The crucial difference is that, unlike true adjectives, adjectival representations have (inflectional) possessive forms. Thus, in addition to the associative form of the unpossessed noun *qaqly* sledge’, *qaqly-l′* ‘pertaining to a sledge’, we have forms such as *qaqly-nɪː-l′* ‘pertaining to our.DU sledge’ and *qaqly-ntyty-l′* ‘pertaining to their.PL sledge’, where *-nɪː-* and *-ntyty-* are possessive affixes. Simple denominal adjectives are not compatible with possessive affixes, cf.:

(6) a. *(mat) mɔːt-ny-**symyl′** qum
I.GEN house-1SG-**PROP** man
‘man with my house’

b. (mat) pɔːra-ny-**šal′** qum
I.GEN size-.1SG-.**SIM** man
‘man of my size (lit. man similar to my size)’

Example (6a) illustrates a denominal adjective which is formed with the suffix *-symyl′* and has a proprietive meaning (‘with N’). Such adjectives cannot be derived from inflected word forms, including possessive forms, as is expected of a derivational process. The behaviour of the similitudinal suffix *-šal′* in (6b) is different: it can attach to a noun bearing possessor agreement morphology, as is typical of inflectional morphology.

A further intriguing observation is that the suffixes of the three adjectival representations are incompatible with case markers. In the case of the locative representation, this is not surprising, given the meaning. However, it remains puzzling that the associative and similitudinal suffixes should be incompatible with at least the meaningful case markers, given that they combine with

possessor inflection. In other words, it is not difficult to conjure up meanings for the relational adjectives which would include cases. For instance, it is unclear why we cannot have similitudinal forms derived from cases other than nom, acc, voc, say ‘similar to knife-GEN (blade)’ = ‘(blade which is) similar to a knife’s’ or ‘pertaining to river-LOC (mist)’ = `(mist which is associated with being) at the river’. But such forms are impossible.

So the similitudinal suffix (and those of the other two adjectival representations) are not only compatible with inflected forms of lexemes, but are also in paradigmatic opposition to case suffixes. Although it is not itself a case suffix in any obvious sense of the term, it belongs functionally to the same set of suffixes as the case suffixes. Classical PFM is not designed to handle such categorial mixing. We therefore adopt an extension of PFM under which a Generalized Paradigm Function (GPF) is capable of mapping a complete lexical representation (consisting of <FORM, SYNTAX, SEMANTICS, LEXEMIC INDEX> attributes) to another complete lexical representation. The LEXEMIC INDEX (LI) is an arbitrary integer individuating all and only the distinct lexemes in the lexicon. For straightforward inflection the SYN, SEM, LI attributes are not mentioned in the GPF, but are defined by the Identity Function Default and the mapping is effected by three blocks of realization rules applying to the FORM value. This is equivalent to the classical PF analysis, as seen schematically in (7):

(7) GPF(LEADER, {<pl, 2pl, car>})

maps		to
FORM:	qoː-	qoːˈ-iː-nyt-kɔːlyk
SYN: N	N (by IDF)	
SEM:	[LEADER(x)]	[LEADER(x)] (by IDF)
LI:	LEADER	LEADER (by IDF)

Derivational morphology involves a non-trivial change in all four attributes (as with the Selkup proprietive adjective, (6a)).

A pure (a-semantic) transposition is effected by a GPF which alters the FORM and SYN attributes but not the SEM or LI attributes. We represent syntactic categories as argument structure representations following, essentially, Higginbotham (1985), Spencer (1999). A noun has the semantic function role ‘R’, and an attributive adjective has the role $A^*_i(x_i)$. Attribution consists of coindexing the ‘A’ and the ‘R’ roles: *tall*<$A^*_i(x_i)$ *tree*<R*>. A relational adjective such *polar* has the derived role <$A^*_i(x_i)$,<R_1>>: *polar* <$A^*_i(x_i)$,<R_1>> *icecap*<R^*_2> ‘icecap bearing some relation to the pole’:

(8)

FORM:	pole-ar
SYN:	<$A^*_i(x_i)$,<R>>
SEM:	[POLE(x)]
LI:	POLE

The associative representation is effectively ‘inflectional transposition’, i.e. a member of the inflected paradigm of the noun which nonetheless acquires the morphosyntax of an adjective:

(9) a. GPF(SLEDGE, {<sg, unpossd, associative>})

maps		to
FORM:	qaqly-	qaqly-l′
SYN:	<R>	<$A^*_i(x_i)$,<R>>
SEM:	[SLEDGE(x)]	[SLEDGE(x)]
LI:	SLEDGE	SLEDGE

b. GPF(SLEDGE, {<sg, 2pl, associative>}) ‘pertaining to your(pl) sledges’

maps		to
FORM:	qaqly-	qaqlynty-ty-l′
SYN:	<R>	<A*$_i$(x_i),<R>>
SEM:	[SLEDGE(x)]	[SLEDGE(x)]
LI:	SLEDGE	SLEDGE

The GPF can also add a semantic predicate to such an "inflectional transposition" without changing the lexemic status of the output, as with the similitudinal:

(10) GPF(SLEDGE, {<sg, 2pl, similitudinal>}) 'similar to your(pl) sledge'

maps		to
FORM:	qaqly-	qaqlyn-ty-ššal′
SYN:	<R>	<A*$_i$(x_i),<R*>>
SEM:	[SLEDGE(x)]	[SIMILAR_TO(x, y)[SLEDGE(y)]]
LI:	SLEDGE	SLEDGE

This is similar to Booij's (1996) 'inherent inflection', which also (generally) adds semantic content, but in the case of the similitudinal representation, the process is not purely inflectional because it changes the lexical category, rather like a transposition.

Depending on language-specific (indeed, construction-specific) principles the 'R' role of the original noun may or may not be accessible to morphosyntactic principles. But since the adjectival representations of Selkup nouns are essentially like case-marked forms they often behave like nouns in the syntax, taking their own modifiers/specifiers, as in (6b).

The Selkup system of 'representacija' thus demonstrates the need for a model of lexical relatedness that can be fully integrated into a model of derivation, and is flexible enough to account for the complex interactions which give rise to the kinds of non-canonical inflection illustrated here, and which also provides sufficiently rich representational structure to allow us to reflect the mixed categorial status of these types of word.

References

Booij, G. 1996. Inherent vs. contextual inflection and the split morphology hypothesis. *YoM* 1995: 1-16.

Kuznetsova, A. I. et al. 1980. *Ocherki po sel'kupskomu jazyku*. Nauka.

Higginbotham, J. 1985. On semantics. LI 16, 547—621.

Spencer, A. 1999. Transpositions and argument structure. YoM 1998:73—102.

Stump, G. 2001. *Inflectional Morphology*. CUP.

Évaluation et pluriactionnalité : une mise au point théorique qui s'impose

Dejan STOSIC, Université d'Artois, Grammatica EA 4521
Dany AMIOT, Université de Lille 3, STL UMR 8163

De nombreuses langues font appel à des procédés morphologiques comme l'affixation (1-2) ou la réduplication (3) pour exprimer des sens souvent qualifiés d'itératifs, de distributifs ou de fréquentatifs :

(1) fra. *taper / tapoter*
(2) srp. *vaditi* 'sortir qqch de' */ po-vaditi* 'sortir une par une plusieurs choses'
(3) egy. *wn* 'aller' / *wn-wn* 'aller d'avant en arrière' (repris à Tovena & Kihm 2008)

Ces phénomènes de pluralisation d'événements sont souvent mis en relation avec la notion de diminution (l'itération impliquant dans ce cas la subdivision de l'événement dénoté par la base en une pluralité de petits sous-procès de même type) ; dès lors deux grands types de travaux ont vu le jour pour tenter d'appréhender ces faits : les travaux portant sur la sémantique du pluriel (de loin les plus nombreux ; cf. par ex. Cusic 1981, Lasersohn 1995, Collins 2001, Wood 2007, Greenberg soumis, etc.) et certains travaux, plus rares, de morphologie (Dal 1999, Sörés 1999, Plénat 1999, Amiot 2002). De façon générale, ces recherches se font sans réelle référence l'une à l'autre : les premières ont souvent pour cadre les travaux sur l'aspect, la pluralité et la quantification des événements alors que les secondes constituent en quelque sorte un sous-domaine de la morphologie, celui de la morphologie évaluative : ce sont en effet généralement les mêmes procédés qui servent à construire des lexèmes à sens évaluatif (diminutif, augmentatif, etc.) et les sens itératifs dont il est question ici. Le seul article, à notre connaissance, qui fasse, d'une certaine manière, le lien entre les deux approches est celui de Tovena & Kihm (2008).

Cependant, la lecture de ces deux types de travaux, ainsi que des études en morphologie fondées sur un corpus relativement conséquent de verbes déverbaux du français soulève un certain nombre de questions : les spécificités de traitement qui apparaissent au fil des lectures sont-elles à rapporter à de simples différences terminologiques ou sont-elles le signe de divergences plus profondes ? Est-ce que l'une des deux approches possède davantage de pouvoir explicatif que l'autre ?, etc. Pour tenter de répondre à ces questions, nous allons d'abord expliciter les principes et les notions centrales de chaque approche (d'abord la morphologie, § 1, puis la pluriactionnalité, § 2.), pour ensuite (§ 3) « tester » chacune d'elles sur un corpus d'environ 170 verbes déverbaux suffixés par l'un des suffixes évaluatifs du français que nous avons répertoriés, i.e. *-eter* (*voler / voleter*), *-oter* (*siffler / siffloter*), *-asser* (*rêver / rêvasser*), *-oner* (*mâcher / mâchonner*), *-iner* (*trotter / trottiner*), *-Vcher* (*pleurer / pleunicher*), *-iller* (*sauter / sautiller*), *-Viller* (*crier / criailler*), le corpus ayant été établi à partir du *TLFi*. Précisons que nous avons retenu les verbes dans lesquels figurent l'une des formes suffixales mais qui ne peuvent être considérés comme ayant été régulièrement construits, aucune base n'étant identifiable en synchronie (cf. par ex. *barbouiller*, qui ne peut avoir pour base *barber* 'ennuyer', ou *gribouiller*, auquel ne semble correspondre aucun verbe ??*griber*).

En ce qui concerne la morphologie, nous ferons le point, de façon relativement rapide vu l'importance de la littérature sur le sujet, sur les grands principes qui sous-tendent ce type de morphologie ; pour cela nous prendrons principalement appui sur Dal (1997, 1999), Fradin (2003), Fradin & Montermini (2009), Grandi (2002), Juravsky (1996), Dressler & Merlini

Barbaresi (1994), Merlini Barbaresi (2006). De ces travaux se dégagent quelques grandes lignes de force en ce qui concerne tant les aspects formels de la morphologie évaluative (types de procédés, catégories mises en jeu, etc.) que les aspects proprement sémantiques, ceux qui nous intéressent au premier chef. Sémantiquement, la morphologie évaluative semble avoir pour caractéristique première de construire des lexèmes qui manifestent, d'une manière ou d'une autre, un écart par rapport à une norme supposée, écart qui se réalise à de nombreux niveaux : purement sémantique, par rapport à la représentation prototypique du référent du nom de base, mais aussi pragmatique ou discursif (cf. notamment sur ce dernier point Dressler & Merlini-Barbaresi (1994) et Merlini-Barbaresi (2006). L'une des conséquences, sur laquelle s'accordent généralement les auteurs cités, est la grande variété interprétative des dérivés évaluatifs (diminution, augmentation, affectivité, péjoration, appropriation, familiarité, itérativité, etc.) ; toutefois, les manières de traiter cette diversité divergent : certains (comme Juravsky) considèrent qu'il existe un sens fondamental dont les autres seraient dérivés, d'autres (par ex. Dal) font l'hypothèse d'un sens abstrait fondamental se réalisant de différentes façons en fonction de différents paramètres (par ex. la catégorie lexicale), d'autres enfin (Fradin, Fradin & Montermini entre autres) ne font pas réellement l'hypothèse d'un sens unique mais insistent sur le caractère multipolaire des interprétations, celles-ci pouvant d'ailleurs se cumuler. Nous nous sommes cependant heurtés à une difficulté réelle : si les sens itératifs sont souvent mentionnés, ils ne sont que très peu étudiés, la majorité des travaux portant sur les noms (dénominaux) ; quelques exceptions notables cependant : Sörés (1999), Plénat (1999), Amiot (2002), Tovena & Kihm (2008), même si pour ces derniers il n'existe pas à proprement parler de suffixes évaluatifs verbaux. Ces derniers travaux nous permettront de faire le lien entre ce qui est dit de l'évaluation en général et les interprétations plus particulières qui apparaissent dans les verbes déverbaux.

La seconde partie sera donc consacrée aux approches liées à la pluriactionnalité. Les études effectuées dans ce type de travaux se sont, elles, au contraire, développées à partir de l'étude de la catégorie verbale (le titre d'un des ouvrages fondateurs, celui de Cusic 1981, s'intitule d'ailleurs *Verbal Plurality and Aspect*), même si des parallèles ont ensuite été établis avec la pluralisation nominale (cf. par ex. Landman 1996), mais sans que l'évaluation y soit réellement évoquée. Nous verrons qu'une des caractéristiques principales, et un des intérêts majeurs de cette notion est qu'elle permet de conceptualiser de façon unitaire des interprétations / effets de sens qui peuvent sembler à première vue très dissemblables. La pluralisation peut en effet concerner le procès lui-même (*mordiller*), mais aussi d'autres paramètres impliqués par le procès : les actants / participants (*se baisoter* 'se donner des baisers nombreux'), la localisation, spatiale (*voleter* 'Voler ça et là en se posant souvent') ou temporelle (*dansoter* 'danser un peu'), etc. ; cf. la définition de Lasersohn (1995 : 240) « Pluractional markers attach to the verb to indicate a multiplicity of actions, whether involving multiple participants, times or locations ». Nous chercherons aussi à rendre compte (i) de distinctions qui se révèleront utiles pour notre étude de corpus, notamment la distinction entre pluralité interne et pluralité externe, et (ii) de ce qui est généralement affirmé à propos de l'aspect des verbes pluriactionnels ; ceci pour préparer notre troisième partie, sur l'étude du corpus.

Dans cette dernière partie, nous pensons fournir une analyse aussi précise que possible du rôle de la suffixation évaluative pour les verbes de notre corpus, et ceci en fonction de différents paramètres :
– la pluralisation, et s'il y a pluralisation, nous indiquerons quelle dimension est affectée : pluralisation du procès, des participants, etc. ;

– l'évaluation, et là aussi, s'il y a évaluation, nous indiquerons la dimension affectée : la dimension mesurative (diminution / l'augmentation), affective (hypocoristique, péjoratif / mélioratif, etc.) ou discursive (par ex. la minoration de la prise en charge du locuteur). Pour parvenir à mettre au jour ces types d'interprétations, notamment le dernier, nous aurons en général recours à des emplois contextualisés.
– l'aspect lexical, du verbe de base et du verbe suffixé.

Par cette étude, nous souhaitons mettre en évidence à la fois la complémentarité des deux notions (il ne s'agit donc pas d'une simple question de variation terminologique) et leur relative autonomie. Il est en effet important de noter qu'il peut y avoir évaluation sans pluralisation (*neigeoter, traînasser*) et pluriactionnalité sans évaluation, dans des verbes simplex (par ex. *mâcher* ou *taper*), mais aussi, même si les cas sont peu fréquents, dans les verbes dérivés (cf. *lanciner* 'Se faire sentir par élancements douloureux / Causer à quelqu'un des douleurs par élancements'.). Ces deux types d'interprétation existent en effet indépendamment l'un de l'autre, bien qu'elles soient fréquemment associées dans la mesure où elles exploitent les mêmes procédés morphologiques.

Cette étude permettra aussi de mieux saisir un autre lien qui s'établit entre évaluation et pluriactionnalité, cette fois au niveau conceptuel. En effet, les verbes pluriactionnels étudiés manifestent, du fait même de la pluralisation interne des procès qu'ils dénotent, la propriété caractéristique des lexèmes construits par la morphologie évaluative évoquée au début de ce résumé, à savoir l'expression d'un écart par rapport à une norme (pour ne prendre que quelques exemples simples : *tapoter* 'taper de façon légère et répétée', *courailler* 'courir de côtés et d'autres'). La pluralisation interne ne constitue, bien entendu, qu'une des façons possibles de rendre le procès non-conforme par rapport à la norme.

Bibliographie

Amiot, D. (2002), « Re-, préfixe aspectuel ? » *Chronos* 10, 1-20.

Collins, C. (2001), Aspects of Plurality in ǂHoan, *Language* 77/3, 456-476.

Cusic, D. (1981), *Verbal Plurality and Aspect*, PhD dissertation, Stanford University.

Dal, G., (1997), *Grammaire du suffixe –et*, Paris, Didier Érudition.

Dal, G. (1999), « Suffixation par *–et(te)* et bases verbales », In D. Corbin, G. Dal, B. Fradin, B. Habert, F. Kerleroux, M. Plénat & M. Roché éds, *La morphologie des dérivés évaluatifs*, *Silexicales* 2, Université de Lille III, 37-47.

Dressler, W. U. & Merlini Barbaresi, L. (1994). *Morphopragmatics. Diminutives and Intensifiers in Italian, German, and other Languages*, Berlin, Mouton de Gruyter.

Fradin, B., (2003), Le traitement de la suffixation en *–et*, *Langages* 152, 51-77.

Fradin, B. & Montermini, F. (2009), La morphologie évaluative », in Fradin, B., Kerleroux, F. & Plénat, M. (éds), *Aperçus de morphologie du français*, Presses Universitaires de Vincennes, 231-266.

Grandi, N. (2002), *Morfologie in contatto. Le construzioni valuative nelle lingue del Mediterraneo*, Milan, FrancoAngeli.

Greenberg, Y. (soumis), « Event Internal Pluractionality in Modern Hebrew: A Semantic Analysis of One Verbal Reduplication Pattern ».

Juravsky, D. (1996), « Universal tendencies in the semantics of the diminutive ». *Language*, 72, 3: 533–578.

Merlini Barbaresi, L. (2006), « Morphopragmatics ». *Encyclopedia of Language and Linguistics*. 2nd Edition. K. Brown. Oxford, Elsevier Ltd: 332-335.

Landman, F. (1996), "Plurality." In Sh. Lappin (ed), *The Handbook of Contemporary Semantic Theory*. Oxford: : Blackwell, 425-457.

Lasersohn, P. (1995), *Plurality, Conjunction and Events*. Kluwer, Dordrecht.

Plénat, M. (1999), Distribution des suffixes évaluatifs verbaux en français », In D. Corbin, G. Dal, B. Fradin, B. Habert, F. Kerleroux, M. Plénat & M. Roché éds, *La morphologie des dérivés évaluatifs*, *Silexicales* 2, Université de Lille III, 179-188.

Sörés, A. (1999), « Comment un procès peut-il être évalué ? Étude sémantique de quelques dérivatifs verbaux en hongrois », In D. Corbin, G. Dal, B. Fradin, B. Habert, F. Kerleroux, M. Plénat & M. Roché éds, *La morphologie des dérivés évaluatifs*, *Silexicales* 2, Université de Lille III, 243-251.

Tovena, L. & Kihm, A. (2008), « Event internal pluractional verbs in some Romance languages », in Tovena L. (ed.), *Aspect et pluralité d'événements. Recherches linguistiques de Vincennes* 37 : 9-30.

Wood, E. J. (2007), *The semantic typology of pluractionality*. Unpublished Ph.D. dissertation, UC, Berkeley.

Anna M. Thornton (Università dell'Aquila)
Towards a Typology of overabundance

The paper is couched in the canonical approach to morphological typology (Corbett 2005, 2007a,b). In this approach, paradigms are canonical if they conform to all the conditions stated in table 1.

Table 1 Canonical Inflection (Corbett 2007a: 9; cf. also Corbett 2007b: 23-24)

	COMPARISON ACROSS *CELLS* OF A LEXEME	COMPARISON ACROSS *LEXEMES*
COMPOSITION/STRUCTURE (≈ means of exponence)	same	same
LEXICAL MATERIAL (≈ shape of stem)	same	different
INFLECTIONAL MATERIAL (≈ shape of inflection)	different	same
OUTCOME (≈ shape of inflected forms)	different	different

There are then several ways in which paradigms may deviate from canonicity; the possible deviations can be seen within a paradigm, i.e., by comparing forms in different cells of a single lexeme's paradigm, or between paradigms, i.e., by comparing forms in the corresponding cell(s) of different lexemes.
The different kinds of deviations from canonicity are presented in Tables 2 and 3 (from Corbett 2007b).

Table 2 Deviations from canonical behavior in single lexemes

	canonical behaviour	**deviant behaviour**	**type of deviation**
COMPOSITION/STRUCTURE (of the inflected *word*)	same	different	fused exponence periphrasis
LEXICAL MATERIAL (≈ shape of *stem*)	same	different	alternations suppletion
INFLECTIONAL MATERIAL (≈ shape of *affix*)	different	same	syncretism uninflectability

Table 3 Deviations from canonical behavior across lexemes

	canonical behaviour	**deviant behaviour**	**type of deviation**
COMPOSITION/STRUCTURE (of the inflected *word*)	same	different	defectiveness overdifferentiation anti-periphrasis
LEXICAL MATERIAL (≈ shape of *stem*)	different	same	homonymy
INFLECTIONAL MATERIAL (≈ shape of *affix*)	same	different	inflectional classes heteroclisis deponency

This presentation will address a further kind of deviation, called **overabundance**, which occurs when a cell in a paradigm is filled by two synonymous forms which realize the same set of morphosyntactic properties. A very well-known example is given in (1):

(1) English 'burn.PST' = *burnt* / *burned*

Overabundance does not appear in Tables 2 and 3. There is a question about where this phenomenon might appear. In Table 3, it should probably appear alongside with defectiveness, as a deviation in the composition of a paradigm that can be detected by comparing paradigms from different lexemes: where Lexeme X has a single form in a given cell, Lexeme Y has two (or more) forms in the same cell. In a certain sense, overabundance is a mirror image of defectiveness: while defective paradigms lack forms in one or more cells, overabundant paradigms have more than one form in one or more cells.

The forms that inhabit the same cell of a paradigm can be called cell-mates, as proposed by Thornton (to appear). This term is to be preferred to the term *doublets*, which is sometimes used, because there can be more than two cell-mates in a given cell. Examples of such overcrowded cells are given in (2):

(2) Latin *fecĕrunt / fecērunt / fecēre* 'do.3PL.PRF.IND'
MS Arabic *sāriqūn / sāriqāt / saraqa / surrāq / sawarīqu* 'thief.PL'
(sg. *sāriq;* Kaye 2007:235)

It is not possible to insert overabundance in Table 2, as the presence of cell-mates adds a dimension to the ones considered in Table 2. In fact, a specific form in a set of cell-mates can differ from its mate(s) in terms of all the factors considered in the rows of Table 2: composition or structure of cells (a formula that in Corbett's system refers to the "means of exponence" by which a cell is filled, cf. Corbett 2007b: 24), lexical material and inflectional material. Examples for all these cases are given in Table 4:

Table 4

	Type of deviation	**Examples**
COMPOSITION/STRUCTURE (≈ means of exponence)	forms built according to different means of exponence in the same cell	Dutch *drukste / meest drukke* 'busy:SUPERL' Arabic *sāriqūn / surrāq* 'thief.PL'
LEXICAL MATERIAL (≈ shape of stem)	forms built on two different stems in the same cell	Italian *devo / debbo* 'must.1SG.PRS.IND' English *cacti / cactuses* 'cactus.PL'
INFLECTIONAL MATERIAL (≈ shape of inflection)	forms with two different inflectional endings in the same cell	Latin *fecĕrunt / fecērunt / fecēre* 'do.3PL.PRF.IND'

The two Dutch Superlatives in Table 4 differ from each other because one form is suffixed while the other is periphrastic[1]; the two Arabic noun plurals differ from each other because the first is realized

[1] Reviewer n. 1 observes: "I don't understand the role (and the relevance) of periphrastic constructions in the whole picture". Following the suggestion by the organizers to revise the abstract to take into account the comments of the reviewers, I will try to clarify the mention of periphrasis here. Table 4 offers examples of cell-mates that differ from one another along the three dimensions identified by Corbett as relevant to classify types of deviant, non-canonical behaviour in paradigms. As shown in Table 2, periphrasis is a type of deviation from canonicity which has to do with the structure of the material in a given cell: canonical inflectional forms are non-periphrastic, and periphrastic forms are non-canonical. In a pair of cell-mates in which one mate is periphrastic and the other one isn't, there are two kinds of deviations from canonicity: having two forms realizing the same cell is a deviation, and having a periphrastic form realizing the cell is another deviation.

by suffixing *-ūn* to the singular, while the second is a "broken" plural, realized by fitting the stem in a specific prosodic template with a prespecified vocalic melody.[2] The other examples in Table 4 show cases of cell-mates built on different stems of a lexeme, or by using different inflectional endings after one and the same stem.

Canonical overabundance is defined as a situation in which two or more forms that realize the same cell (i.e., the same set of morphosyntactic features) in a lexeme's paradigm can be used interchangeably, with the choice of one or the other form subject to no condition. As is expected in the canonical approach, the definition is likely to be matched by no actual case.[3] In many cases two cell-mates will differ according to some kind of condition concerning dimensions of language variation, or according to grammatical conditions.

Contrasts between relatively unconditioned (and therefore more canonical) overabundance and overabundance conditioned by some dimension of variation can be illustrated by the different sets of cell-mates realizing 1SG.PRES.IND of some Italian verbs, listed in (3):

(3) a. *possiedo / posseggo* 'possess.1SG.PRS.IND'

highly canonical; both forms are attested with almost equal frequency in contemporary Italian, and do not seem subject to diaphasic, diastratic or diatopic conditioning

b. *chiedo / chieggo* 'ask.1SG.PRS.IND'

non-canonical: *chieggo* is extinct in contemporary Italian, and is therefore diachronically conditioned, being found only in texts produced up to the beginning of the 20th century

c. *vado / vo* 'go.1SG.PRS.IND'

non-canonical: *vo* in contemporary Italian is used only by speakers of the Tuscan variety, or in poetry and other text types subject to metrical constraints (proverbs, songs) and is therefore diatopically and diaphasically conditioned

Contrasts between sets of unconditioned cell-mates vs. sets conditioned by some kind of system-internal linguistic factor can be illustrated by data on double Preterit forms in English verbs (Quirk 1970, Kempson & Quirk 1971)[4]:

(4) a. *burnt / burned*

highly canonical; both forms used in similar measure in all kinds of contexts

b. *spoilt / spoiled, dreamt / dreamed*

non canonical, aspectually conditioned: (particularly in British English) *-t* forms are more likely in perfective contexts, *-ed* forms are more likely in imperfective / durative contexts (Quirk 1970)

c. *wet / wetted*

non canonical, semantically conditioned: according to Kempson & Quirk (1971: 551) *wet* is used

[2] Reviewer n. 1 observes: "the abstract does not touch upon […] the issue of inflectional class changes". This issue was not explicitly mentioned in the abstract for lack of space: however, many examples (such as the Arabic and the English ones in (2) and (4)) involve cell-mates which differ because they follow different inflectional classes (strong vs. weak verbs in English, different types of sound vs. broken plurals in Arabic). Overabundance in a cell is often a stage along the path of inflectional class change: a lexeme starts as a regular (i.e., non-overabundant) lexeme of class x, then in an intermediate stage it displays some inflected forms that follow class x and some that follow class y (constituting pairs of cell-mates), and eventually (often, but not always) forms of class x go out of usage, only forms of class y are used and class change is completed.

[3] Corbett (2007a: 9) observed: "The canonical instances, that is, the best, clearest, indisputable (the ones closely matching the canon), are unlikely to be frequent. Rather, they are likely to be rare or even nonexistent".

[4] These forms can also show contrasts due to factors of variation: for example, *leaped* is more used in the US and *leapt* in the UK.

both in the general sense 'moistened' and in the more restricted sense 'urinated', *wetted* is used only in the sense 'moistened'; some speakers, however, don't have this disticntion (Grev Corbett, personal communication): for these speakers, then, *wet / wetted* are rather canonical cell-mates.[5]

In the presentation, results of a preliminary survey on the extent of canonical and non-canonical overabundance in the world's languages will be presented.

Topics that will be addressed include:

a) word classes in which overabundance is attested;
b) correlations between overabundance and morphosyntactic features;
c) correlations between overabundance and morphomic partition classes.
d) correlations between overabundance and frequency.

Concerning a), preliminary data have shown cases of overabundance in verbs, nouns, adjectives, pronouns (cf. Cappellaro 2010); data on possessives and demonstratives are currently being investigated.

Concerning b), certain feature specifications seem to be particularly prone to overabundance; for example, in nouns overabundance is common in the plural, while rare in the singular.

Concerning c), Thornton (to appear) has shown that in Italian verb paradigms overabundance occurs in cell that belong to independently established morphomic partition classes; it is a question for future research whether this is a general tendency or a language-specific fact.

Concerning d), data must still be studied carefully. The few data that have been studied so far show contrasting tendencies: while in Italian verbs overabundance seems to occur in fairly frequent verbs, in Italian pronouns it occurs rather in the least frequent cells (Cappellaro 2010).[6]

References

Cappellaro, Chiara (2010). 'Pronominal overabundancy in Italo-Romance. A canonical approach', paper presented at the 14th IMM, Budapest, May 2010.

Corbett, Greville G. (2005). 'The canonical approach in typology', in Z. Frajzyngier, A. Hodges, D. S. Rood (eds.), *Linguistic Diversity and Language Theories*. Amsterdam / Philadelphia, Benjamins, 25-49.

Corbett, Greville G. (2007a). 'Canonical typology, suppletion, and possible words', *Language* 83.1, 8-41.

Corbett, Greville G. (2007b). 'Deponency, Syncretism, and What Lies Between', *Proceedings of the British Academy* 145, 21-43.

Kaye, Alan S. (2007). 'Arabic Morphology', in A. S. Kaye (ed.), *Morphologies of Asia and Africa*, Winona Lake (Indiana), Eisenbrauns, 211-247.

Kempson, Ruth M. & Randolph Quirk (1971). 'Controlled Activation of Latent Contrast', *Language* 47.3, 548-572.

Quirk, Randolph (1970). 'Aspect and Variant Inflection in English Verbs', *Language* 46.2, 300-311.

Thornton, Anna M. (2010). 'Diachronic paths to reduction and maintenance of overabundance in Italian verb paradigms', paper presented at the 14th IMM, Budapest, May 2010.

Thornton, Anna M. (to appear). Overabundance (multiple forms realizing the same cell): a non-canonical phenomenon in Italian verb morphology. In Goldbach, M. *et al.* (eds.) *Morphological Autonomy: Perspectives from Romance Inflectional Morphology*, Oxford: Oxford University Press.

[5] Reviewer n. 1 observes: "the abstract does not touch upon important issues which have been hotly debated in the literature (for instance parallel forms like *brothers / brethren*)". In as far as *brothers / brethren* can be considered cell-mates, they would be highly non-canonical mates, heavily semantically conditioned. But I would prefer to consider *brothers* and *brethren* as the plural forms of two different lexical items, which are homophonous in their singular form *brother*, because the semantic difference between 'male sibling' and 'fellow member of a profession, society or sect' is big enough to recognize two different items. But of course the border between homophony and polysemy is indeed hotly debated in the literature, and a typology of overabundance will have to address this problem.

[6] The relation between frequency and overabundance in Italian verb paradigms has been investigated, after the submission of this abstract, by Thornton (2010).

What semantic content for a Lexeme-Formation Rule ?

The case of noun to verb conversion in French

Delphine Tribout – LLF & Université Paris Diderot

In this paper I will discuss noun to verb conversion in French. I will particularly focus on the semantic part of the rule. After having presented the different meanings converted verbs may have, I will argue that neither an underspecified semantic instruction nor several fully specified ones constitute a sastifying way of handling the semantic part of the rule. I will then propose an output oriented way of considering the semantics of conversion.

1 Definition of conversion

Conversion is a lexeme formation pattern characterized on the one hand by the phonological identity of the base lexeme and the derived lexeme[1], and on the other hand by the fact that the two lexemes necessarily belong to two different parts of speech, as the french examples in (1) show.

(1) SCIE$_{N}$ 'saw' > SCIER$_{V}$ 'to saw'
DANSER$_{V}$ 'to dance' > DANSE$_{N}$ 'a dance'
CALME$_{A}$ 'calm' > CALMER$_{V}$ 'to calm down'

In French the inflectional marks, and particularly the verbal ones, can conceal the identity of the two lexemes. But these marks must not be taken into account (see (Corbin, 1987) on that subject).

In the lexeme-based theory of morphology adopted here ((Matthews, 1972), (Aronoff, 1994)) the unit of the morphology is the lexeme, which is defined as a complex, multidimensional object, having at least a form, a meaning and a syntactic category. Therefore a Lexeme-Formation Rule must specify each of these dimensions on the input side as well as on the output side. In the case of noun to verb conversion, we only need to say on the form level that the two lexemes are identical, and on the categorial level that the base is a noun and the derived lexem is a verb. I will thus focus on the semantic dimension of the conversion rule.

2 Semantic types of denominal converted verbs in French

According to (Plag, 1999) there are eight semantic types of denominal converted verbs in English : locative, ornative, resultative, performative, similative, instrumental, privative and stative verbs. These eight semantic types are found with french converted verbs too. But another type will be necessary to account for all the data.

Locative verbs are movement verbs. They derive from nouns which denote the final location of the event denoted by the verbs, as in the examples in (2).

[1] In fact, the identity between the two lexemes is not merely phonological and concerns the different stems of both lexemes. I simplify here for the sake of clarity.

(2) CADRE 'frame' > CADRER 'to frame' ≈ 'to put into a frame'
CLOÎTRE 'cloister' > CLOÎTRER 'to put into a cloister'

Ornative verbs are movement verbs too. They denote the movement of an entity which is denoted by the base noun. Some examples are shown in (3).

(3) POIVRE 'pepper' > POIVRER 'to pepper' ≈ 'to put some pepper to'
SELLE 'saddle' > SELLER 'to saddle' ≈ 'to put a saddle on'

Resultative verbs are verbs deriving from nouns which denote the resulting object of the event denoted by the verbs, as shown in (4).

(4) BAVE 'drool' > BAVER 'to drool' ≈ 'to produce some drool'
PRÉFACE 'preface' > PRÉFACER 'to write a preface' ≈ 'to produce a preface'

Performative verbs derive from nouns denoting an event, and not an object. They denote the same event as their base nouns. Some of them are presented in (5).

(5) ANALYSE 'analysis' > ANALYSER 'to analyse' ≈ 'to do/perform an analysis'
PARTAGE 'sharing' > PARTAGER 'to share' ≈ 'to do/perform a sharing'

Similative verbs denote an event whose agent is acting as if it were what is denoted by the base noun. The base noun can denote an animate, human or not, as well as an inanimate, as the examples in (6) show.

(6) SINGE 'ape' > SINGER 'to ape' ≈ 'to act like an ape'
BALLON 'balloon' > BALLONNER 'to swell' ≈ 'to do do what a balloon would do'

Instrumental verbs derive from nouns denoting an object but not necessarily an instrument. They refer to an activity in which the object denoted by the noun is used one way or another, as in the case of the verbs in (7).

(7) BOUTON 'button' > BOUTONNER 'to button up' ≈ 'to use a button'
CLOU 'nail' > CLOUER 'to nail' ≈ 'to use a nail'

Privative verbs are another type of movement verbs, but unlike ornative verbs their base nouns denote an object which is removed from another object, as in the examples in (8).

(8) ÉCAILLE 'scale' > ÉCAILLER 'to scale' ≈ 'to remove the scale of'
ÉCORCER 'bark' > ÉCORCER 'to bark' ≈ 'to remove the bark of'

Stative verbs refer to a situation in which the agent is or constitutes what the base nouns denote. The verbs in (9) are some examples of this kind of verbs.

(9) PRÉLUDE 'prelude' > PRÉLUDER 'to be a prelude to'
STÉRÉOTYPE 'stereotype' > STÉRÉOTYPER 'to stereotype' ≈ 'to be a stereotype'

In order to account for all the french denominal converted verbs we need to add one semantic type : the causative verbs. These verbs derive from nouns denoting a state, and refer to a situation in which the sate denoted by the noun is caused, as shown in (10).

(10) CONFUSION 'confusion' > CONFUSIONNER 'to cause confusion'
ATROPHIE 'atrophy' > ATROPHIER 'to atrophy' ≈ 'to cause atrophy'

3 The semantic content of the rule

Having presented nine semantic types of denominal converted verbs in French, the question of the semantic content of the rule arises. There are two ways of considering the question.

Underspecified semantic instruction within the rule

The first one is to consider the semantic instruction of the rule underspecified. This is for instance the position of Aronoff (1980) according to whom the noun to verb conversion rule in English has no more semantic content than the one sketched in (11).

(11) Do something with N

To Aronoff's point of view nothing semantically constraints the conversion rule, so that every verb denoting an activity which has some connection with the entity denoted by the noun is possible. According to him, only pragmatic reasons can prevent or constraint the formation of a denominal converted verb. However, this semantic only holds for verbs which derive from nouns denoting an object. It cannot account for performative and causative verbs. Moreover, even when the noun denotes an object it cannot predict resultative, similative and stative meanings of the verbs.

Following the same idea but in a different way, Plag (1999) proposed that the conversion rule has the underspecified Lexical-Conceptual Structure in (12) as semantic part (where the dash underlined clause is optional).

(12) $\underline{\text{CAUSE}}(\lfloor \_ \rfloor_i, [\text{GO}([_{Property,Thing}\ ]_{Theme/Base}\ ; [\text{TO}\ [_{Property,Thing}\ ]_{Base/Theme}\ ])])$

According to Plag this Lexical-Conceptual Structure can account for the first five types of verb meanings. The structure GO([x] [TO[y]]) denotes a transfer and accounts for both change-of-place verbs (i.e. locative and ornative verbs) and change-of state verbs (i.e. resultative verbs) depending on whether the arguments are a thing or a property. The performative verbs are considered to be equivalents to the ornative ones with the meaning 'apply N (to something)', and the similative verbs to be transfer verbs like the resultative ones. Then, given one and the same semantic instruction, the correct meaning of a verb is assessed on encyclopaedic knowledge grounds. However, Plag does not say how this Lexical-Conceptual Structure can account for the instrumental, privative or stative verbs so that it seems unable to correctly predict all the possible meanings of the verbs. Conversely, this semantic instruction seems able to predict meanings that do not exist, like the change-of-state meaning we find with the *-ify* suffixed verbs such as *momifier* 'mummify' ≈ 'to transform into a mummy'.

Fully specified semantic instructions

Instead of one underspecified semantic instruction within the rule, we could think of nine fully specified semantic instructions. But then other questions arise.

Are there as many conversion rules as there are semantic instructions ? That is, are there nine noun to verb conversion rules depending on the meaning of the output verb ? And if so, how can we choose the correct rule to be applied, given one noun ?

Or, can one and the same rule have many semantic instructions ? And in that case, how is a particular semantic instruction chosen between the nine available when deriving

a new verb from a noun ?

4 Constraints on the outputs ?

Neither the underspecified semantic instruction of the rule, nor the fully specified ones, seem to be a satisfying way of considering the semantic content of the conversion rule. On the one hand underspecified semantic instructions like the two in (11) and (12) cannot predict some meanings of the verbs. Simultaneously they are so much underspecified that they could predict non attested meanings. Besides, they are so much underspecified that we can question the exact semantic content of such instructions. On the other hand, fully specified semantic instructions question the unity of a Lexeme-Formation Rule.

A third solution may be to consider the output verbs to be semantically fully specified. Focusing then on the output, we can think of the rule as constraining possible outputs rather than specifying them fully. We can thus consider nine different types of denominal converted verbs corresponding to the nine one presented in section 2. To each type is associated one constraint which stipulates :

i) the kind of noun referent (i.e. object, animate, state, event...) and

ii) the meaning of the derived verb.

In such a constraints based model of morphology, there is not one conversion rule but nine semantic constraints on conversion. As a result, a well formed denominal converted verb is a verb which satisfies one of these nine constraints on conversion.

Such a conception of conversion fits well with a declarative model of language, and can thus easily be represented in a declaractive, feature structure based formalism like the Sign-Based Construction Grammar framework (Sag, 2010).

References

Aronoff, M. (1980). Contextuals. In T. Hoekstra, H. Van der Hulst, and M. Moortgat (Eds.), *Lexical Grammar*, pp. 263–285. Dodrecht : Foris Publications.

Aronoff, M. (1994). *Morphology by Itself.* Cambridge : The MIT Press.

Corbin, D. (1987). *Morphologie dérivationnelle et structuration du lexique.* Tübingen : Max Niemeyer Verlag.

Matthews, P. H. (1972). *Inflectional morphology.* Cambridge : Cambrige University Press.

Plag, I. (1999). *Morphological Productivity. Structural Constraints in English Derivation.* Berlin/New York : Mouton de Gruyter.

Sag, I. A. (2010). Sign-Based Construction Grammar : an informal synopsis. In H. Boas and I. A. Sag (Eds.), *Sign-Based Construction Grammar.* Stanford : CSLI Publications.

Paradigmatic Generalization of Morphemes

Jochen Trommer, Department of Linguistics, University of Leipzig

1 Introduction

It is a pervasive feature of natural language morphology that formatives which have a well-defined meaning (or function) in a specific part of an inflectional paradigm seem to take on a slightly different meaning in other paradigms or subparadigms. Take as a simple example the Ainu agreement prefix *eci-* which is consistently used to mark 2nd person plural for subjects and objects:

(1) **Ainu Intransitive/Transitive Verb Agreement** (Tamura, 2000)

	1sg	**1pl**	**2sg**	**2pl**	**3sg**	**3pl**	–
1sg			eci-	eci-	ku-	ku-	ku-
1pl			eci-	eci-	ci-	ci-	-as
2sg	en-	un-			e-	e-	e-
2pl	ecien-	eciun-			eci-	eci-	eci-
3sg	en-	un-	e-	eci-	Ø-	Ø-	Ø-
3pl	en-	un-	e-	eci-	Ø-	Ø-	Ø-

Ascribing *eci-* the feature specification [+2+pl] concisely captures its distribution in the whole paradigm except for the paradigm cells for 1st person subject and 2nd person object, where *eci-* appears to cross-reference all 2nd person objects, whether they are singular or plural. Hence in this part of the paradigm, we seem to deal with a slightly different instance (or use) of *eci-* which doesn't denote [+2+pl], but more generally [+2].

The literature knows two major approaches to "exceptional syncretism" of this type. First, the behavior of *eci-* might be captured by a rule of referral (Zwicky, 1985; Stump, 2001) which requires that the paradigm cells for 1sg:2sg (1st person singular subject and 2nd person singular object) are identical to the corresponding cells for 1sg:2pl forms. The problem with this approach is that it is completely stipulative, and doesn't capture the fact that the use of *eci-* in 1:2 forms although distinct from its more general use is still closely related to the "original" meaning of the prefix in the rest of the paradigm. Thus a rule of referral could with equal ease require that the completely unrelated 1:3 forms take over *eci-*. A second kind of approach which avoids this problem is familiar from the Distributed-Morphology literature (Noyer, 1992; Halle & Marantz, 1993): We could assume that *eci-* is maximally underspecified and denotes [+2] throughout. This immediately captures the close connection between *eci*'s different uses, but runs potentially into another problem: *eci-* is now predicted to crossreference *all* 2nd person arguments, whether singular or plural. Since it doesn't actually appear in 2sg:3 and 3:2sg forms it must be blocked for these cases in some way.[1] The same point can also be made from a more conceptual perspective: maximal underspecification combined with blocking doesn't capture the intuition that *eci-* seems to occur in a more regular and basic form for 2pl arguments compared to its extension to 1:2sg forms. Here, I propose a third, alternative approach to the behavior of *eci-* and similar cases, assuming that the formal entry for

[1] An obvious possibility is to assume that *eci-* is blocked by the presence of the more specific 2sg marker *e-*, but this implies that there is an explicit singular feature, an assumption which is controversial.

eci- is actually [+2+pl], but that it is generalized to [+2] in the context of transitive 1st person subjects.

2 The Framework

I'm assuming a paradigmatic approach to morphological exponence (cf. Wunderlich & Fabri, 1994; Stump, 2001) where cells of paradigms are defined as sets of fully specified feature structures such as {[Nom+1–2–pl], [Acc+2–1–pl]}. Morphemes are ordered pairs consisting of a phonological operation Φ and a (possibly underspecified) set of feature structures. Thus the Ainu prefix *ku-* which marks 1pl subjects in the context of 3rd person definite objects can be represented as (*Prefix*(ci), {[Nom+pl], [Acc+3]}). Since I will discuss here only concatenative (and especially affixal) morphology, I will use the simplified notation for morphemes illustrated in (2), where *ci-* is represented by (2-f). Throughout the paper I will use simple privative features for case augmented by the feature specification [+intr(ansitive)] singling out intransitive forms.

(2) **Ainu Agreement Morphemes**

a.	en-	[Acc+1–pl]	e.	ku-	[Nom+1–pl][+3]
b.	un-	[Acc+1+pl]	f.	ci-	[Nom+1+pl][+3]
c.	eci-	[+2+pl]	g.	-as	[Nom+1+pl+intr]
d.	e-	[Nom+2–pl][+3]			

A morphological grammar of a language is defined by a specification of possible paradigm cells (cf. Stump, 2001) which I won't discuss further here), a set of morphemes such as the one in (2) for Ainu, and a list of generalization rules such as the ones in (3):

(3) **Generalization Rules for Ainu**

a.	[+pl]	→	Ø	/ [Nom+1][Acc+2]	(eci-)
b.	[+3]	→	Ø	/ [Nom–3–pl+intr]	(ci- & e-)
c.	[Nom]	→	Ø	/ [Nom+3][+2–pl]	(e-)

A generalization rule basically deletes features from a morpheme - (3-a) effectively replaces the [+2+pl] of *eci* by [+2] in the context of a specific paradigmatic cell - thus (3-a) applies only to paradigm cells containing a feature structure subsumed by [Nom+1] and [Acc+2]. The realization of a paradigm cell P in a given language L is now achieved as follows ((4) abstracts away from the linear order of phonological operations) :

(4) **Realization of Paradigm Cells**

a. Create a Copy C of the morpheme set of L
b. Apply the generalization rules of L to C (with respect to P)
c. For every morpheme M in C:
   If for every feature structure F in M there is some feature structure in P such that F subsumes P,
   then apply the phonological operation of M to the stem of P

For example, ku-:[Nom+1–pl][+3] is inserted into the 1sg:3pl cell (specified [Nom+1–pl][Acc+3–pl]) since no generalization rule applies to the morpheme *ku-* in this context, [Nom+1–pl] in the morpheme trivially subsumes [Nom+1–pl] in the paradigm cell, and [+3] subsumes [Acc+3–

pl]. For the 1sg cell ([Nom–3–pl+intr]), the morpheme ku-:[Nom+1–pl][+3] is generalized by the rule in (3-b) to [Nom+1–pl] which subsumes [Nom+1–pl+intr].

Note that the specifications of morphemes in the approach proposed here are underspecified, but only to a very limited extent. Specifications of morphemes are chosen in a way that the morpheme is not simply restricted to occur *only* in cells which are compatible with (subsumed by) it, but also occurs in *all* cells compatible with it. Thus in the terminology of Pertsova (2009), the morphemes by themselves show the most simple type of underspecification (type 1). Apparently more complex underspecification is achieved by generalization rules. Hence even though *eci-* occurs in paradigm cells containing [+2–pl] (especially 1sg:2sg and 1pl:2sg), it could not be specified as [+2] because it does not occur in *all* cells containing [+2–pl] (thus it doesn't occur in the 2sg:1pl cell). As shown by Pertsova, type1 underspecification is actually the unmarked and typologically most widespread type of underspecification in the languages of the world, an observation which is directly reflected in the structure of morphological grammars as they are proposed here. That morphemes are factually restricted to this type of underspecification follows from the learning algorithm I assume, which is spelled out below.

3 Learning

The formalism developed here lends itself to an easy learning algorithm which combines the learning of segmentation and "meaning", i.e.assigning morphosyntactic features to morphemes. The algorithm takes a paradigm with affix strings and morphosyntactic annotation for each affix cell as its input. It successively removes parts of the affix strings from the paradigm, adding corresponding morphemes and generalization rules to the grammar until all cells contain empty strings. The basic (iteratively repeated) procedure is shown in (5):

(5) **Basic Learning Procedure**

1. Compute the set of Potential Morphemes SPM for all free forms in the paradigm and choose the Optimal Potential Morpheme OPM from SPM
2. Add OPM to the morpheme inventory of G and remove all instances of OPM from the paradigm
3. Recompute potential morphemes for free_form(OPM) from the reduced paradigm
4. Add generalization rules for all resulting PMs to the grammar and remove all instances of these PMs from the paradigm

The *free forms* of a paradigm are the (types of) prefix or suffix strings which occur in isolation (without other affixal material) in paradigmatic cells. For Ainu the relevant free forms are shown in (6):

(6) **Free Forms in Ainu:** en-, ecien-, un-, eciun-, ku-, ci-, e-, -as

Potential morphemes are obtained by computing all possible intersections of the morphosyntactic annotations for cells containing specific free forms F and cells with affix strings containing F and retaining only type-1 feature specifications in the sense of Pertsova (2009). Thus intersecting the features of the 2sg:1sg cell (with the free form *en* and the annotation [Nom+2–pl][Acc+1–pl]) and the 2pl:1sg cells (with the string *ecien* containing *en* and the annotation [Nom+2+pl][Acc+1–pl]) results in the potential morpheme [Nom+2][Acc+1–pl].

In the following I will show how one iteration of (5) derives the correct segmentation and feature assignment for *eci-*. This is an especially interesting case since instances of *eci-* in

Ainu could potentially be segmented differently into independently occurring *e-* and *ci-*. (7) shows for every free form in Ainu the PMs with the maximal paradigmatic range (i.e. covering the maximal number of paradigmatic cells). I disregard *-as, ecien-*, and *eciun-* because these occur only once in the paradigm (whether viewed in isolation or as substrings of other affix strings):

(7) **Ainu Potential Morphemes with Maximal Paradigmatic Range (1st Cycle)**

Free Form	**Potential Morphemes**		**Range**	**Cardinality of Range**
en-	([Nom–1])	[Acc+1–pl]	2sg:1sg, 2pl:1sg,3sg:1sg, 3pl:1sg	4
un-	([Nom–1])	[Acc+1+pl]	2sg:1pl, 2pl:1pl,3sg:1pl, 3pl:1pl	4
ku-	[Nom–1–pl]	[Acc+3]	1sg:3sg,1sg:3pl	2
ci-	[Nom–1+pl]	[Acc+3]	1pl:3sg,1pl:3pl	2
eci-	[+2+pl]		2pl,2pl:1sg,2pl:1pl, 2pl:3sg,2pl:3pl 1sg:2pl,1pl:2pl,3sg:2pl,3pl:2pl	9
e-	[Acc+2]		1sg:2sg, 1sg:2pl, 1pl:2s, 1pl:2pl, 3sg:2sg, 3sg:2pl, 3pl:2s, 3pl:2pl	8

The optimal potential morpheme from a list of PMs is chosen by the following criteria (ranked in this order):

(8) **Selection of the Optimal Potential Morpheme**

Choose the PM with . . .

1. . . . the maximal paradigmatic range
   (covering the maximal number of paradigmatic cells)
2. . . . the minimal recall
   (instances of the free form in the paradigm not covered by the PM)
3. . . . the minimal number of feature structures
4. . . . the biggest number of segments

From the list in (7), eci:[+2+pl] is chosen as OPM since it has the maximal range, and the maximal range criterion outranks all other criteria. Consequently, this entry and the generalization rule in (3-a) are added to the grammar and all instances of *eci-* are removed from the paradigm (step 3 & 4 of (5)) resulting in the paradigm in (9):

(9) **Ainu Paradigm with all Instances of *eci* Removed**

	1sg	**1pl**	**2sg**	**2pl**	**3sg**	**3pl**	–
1sg			Ø-	Ø-	ku-	ku-	ku-
1pl			Ø-	Ø-	ci-	ci-	-as
2sg	en-	un-			e-	e-	e-
2pl	en-	un-			Ø-	Ø-	Ø-
3sg	en-	un-	e-	Ø-	Ø-	Ø-	Ø-
3pl	en-	un-	e-	Ø-	Ø-	Ø-	Ø-