CrossMark

Paradigm structure and predictability in derivational morphology

Olivier Bonami[1] · **Jana Strnadová**[2]

Abstract In this paper we address the usefulness of the notion of a paradigm in the context of derivational morphology. We first define a notion of *paradigmatic system* that extends conservatively the notion as it is used in inflection so as to be applicable to collections of structured families of derivationally-related words. We then build on this definition in an empirical quantitative study of derivational families of verbs in French. We apply information-theoretic measures of predictability initially designed by Ackerman et al. (2009) in the context of inflection. We conclude that key quantitative properties are common to inflectional and derivational paradigmatic systems, and hence that (partial) paradigms are an important ingredient of the study of derivation.

Keywords Inflection · Derivation · Paradigm · Morphological family · Predictability

1 Introduction

The notion that the lexicon is structured by a set of paradigmatic relations linking members of a morphological family and guiding word formation is a recurrent theme in research from the last four decades. In descriptive and theoretical morphology, modern interest in the issue was sparked by van Marle (1984) and further fuelled by Becker (1993), Bauer (1997), Booij (1997), and Pounder (2000). This led to the notable inclusion of paradigmatic relationships in the framework of Construction Morphology (Booij 2010). Independently, Bochner (1993) sketches a formal framework

✉ O. Bonami
olivier.bonami@linguist.univ-paris-diderot.fr

J. Strnadová
jstrnadova@google.com

1 Laboratoire de linguistique formelle, Université Paris Diderot, 5 rue Thomas Mann, 75205 Paris Cedex 13, France

2 Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA, USA

for representing inflectional and derivational paradigms alike, elaborating on Jackendoff (1975). The relevance of these developments was supported by evidence from psycholinguistics on the influence of the size and structure of morphological families on processing (see among many others Schreuder and Baayen 1997; de Jong et al. 2000; del Prado Martín et al. 2005), strongly suggesting the existence of paradigmatic organization in the mental lexicon, and by the development of methods for paradigmatic prediction of lexical units in computational linguistics (Pirrelli and Federici 1994; Pirrelli and Yvon 1999; Cotterell et al. 2017).

It is worth noting that there are two distinct ways in which an approach to word formation can be said to be paradigmatic, which correspond to two senses of the term 'paradigmatic' in modern linguistics. On the one hand, it may focus on paradigmatic relations between words by opposition to syntagmatic relations between words and word parts. Work in this area builds on a broad notion of paradigm, essentially coextensive to what Saussure (1916, p. 175) called an 'associative series'—a set of signs with some property in common,[1] and focuses on any situation where the inclusion of a word in a morphological family or derivational series has a noticeable effect. On the other hand, an approach to word formation may be 'paradigmatic' in the sense that it highlights properties of the word formation system that are parallel to properties exhibited by inflectional paradigms (Bauer 1997; Blevins 2001; Stump 2005; Štekauer 2014; Boyé and Schalchli 2016), or deploys in the context of word formation analytic techniques initially conceived for the study of inflectional paradigms (Kilbury 1992; Bochner 1993; Behrens 1995).

Interestingly, most recent work on derivational paradigms adopts the former approach, and hence exhibits only loose connections to the extensive literature on inflectional paradigms of the last three decades (see among many others Carstairs 1987; Anderson 1992; Stump 2001; Corbett 2007; Stump 2016). The goal of the present paper is to contribute to filling this gap. We will use analytic tools from Word and Paradigm approaches to inflection (Robins 1959; Matthews 1972; Blevins 2016) and assess to what extent these apply fruitfully to structured derivational families.[2] More specifically, we will apply to derivational families information-theoretic measures of predictability pioneered by Ackerman et al. (2009) and usually deployed in empirical studies of inflectional paradigms (see among others Ackerman and Malouf 2013; Bonami et al. 2014a, 2014b; Sims 2015). We will show that collections of structured derivational families exhibit properties in terms of predictability that are homogeneous with those observed for inflection. These properties argue in favor of the importance of paradigmatic organization, in the sense that a derived word's place in a structured network of morphologically related words is crucial to explaining some of its properties that cannot be reduced to the relation with a base.

The structure of the paper is as follows. Section 2 presents the conceptual groundwork on which the study is built. We define a notion of PARADIGMATIC SYSTEM that is a conservative generalization of the standard view of a system of inflectional paradigms. We show that this notion captures important parallels between inflectional paradigms and structured derivational families, and allows one to generalize to

---

[1]Matthews (2001, 50–51) notes that this use of 'paradigm' originates in Hjelmslev (1938).

[2]For readability we use 'derivational family' as a shortcut for 'morphological family of derivationally-related words'. See Sect. 2 for more precise definitions.

derivation the definition of key phenomena familiar from inflection, including inflection classes, heteroclisis, and syncretism. In Sect. 3 we turn to the empirical study. After presenting and illustrating the concept of IMPLICATIVE ENTROPY as a measure of the implicative structure of paradigms, we study the properties of families of French Verbs and related Action and Agent nouns. We show that these exhibit the key properties of DIFFERENTIAL PREDICTABILITY and JOINT PREDICTIVENESS discussed in previous literature in the context of inflection.

2 Paradigmatic systems

In this section we attempt to provide a sound conceptual basis for drawing parallels between inflection and word formation based on the notion of a paradigm. To this effect, we define a notion of PARADIGMATIC SYSTEM that is intended to be directly applicable both to families of inflectionally-related words and to families of derivationally-related words.[3] We then argue that this notion allows us to capture relevant parallels without committing us to ascribing problematic properties to derivational families.

The general outline of the construction is as follows: we take as a primitive a notion of morphological relatedness between (surface) words, and define (partial) morphological families as collections of morphologically related words. A paradigmatic system is then a collection of (partial) families that are aligned in terms of the content-based relations that their members entertain.

2.1 Definitions

To define paradigmatic systems, we start from the view that "Morphology is the study of systematic covariation in the form and meaning of words" (Haspelmath and Sims 2010, p. 2). This we take to mean that morphology deals with situations where some series of pairs of words co-vary both in form and in content.

(1)   Words have a CONTENT, a specification of their syntactic and semantic properties, and a FORM, a specification of their phonology and/or orthography. For any word w we note its content w_c and its form w_f.

(2)   Two words w and w' are MORPHOLOGICALLY RELATED if and only if

   – there exists a nontrivial content relation R_c relating the two words: $R_c(w_c, w'_c)$; and
   – there exists a nontrivial form relation R_f relating the two words: $R_f(w_f, w'_f)$; and
   – there are multiple pairs of words related by that same pairing of a content relation and a form relation.

Note that the definition of 'morphological relation' does not assume that all alternations of content should correspond to the same alternation of form, nor the other way

---

[3]Although we will use a bit of mathematical notation for the sake of explicitness, the following obviously cannot be taken to be a full formalization of the notion of a paradigmatic system.

around. Hence we may assume that the pairs (*replace*, *replaced*) and (*sing*, *sang*) instantiate the same relation of content, but not the same relation of form. Likewise the two nouns (*dog*, *dogs*) and the two verbs (*confer*, *confers*) instantiate the same relation of form but do not instantiate the same relation of content. The nontriviality requirement is intended to provide a sanity check on the relations under consideration: the pairs of words should be cohesive enough that semantically related but formally unrelated words (e.g. *dog* and *canine*) or formally related but semantically unrelated words (e.g. *broth* and *brother*) do not count as morphologically related. Of course there is an element of subjectivity here, and we leave it to further research whether this can be systematized in an operational way.[4]

From this we can now go to the definition of a morphological family:

(3)   A MORPHOLOGICAL FAMILY is a tuple $F = (w_1, \ldots, w_n)$ such that any member w_i of the family is morphologically related to any other member of w_j.[5]

(4)   A morphological family F is COMPLETE if there exists no larger morphological family that contains all members of F. A morphological family is PARTIAL if it is not complete.

Note that we depart from standard practice by not taking an exhaustive definition of morphological families: the word *sing* belongs to multiple families such as (*sing*, *sang*), (*sing*, *singer*), (*sing*, *sings*, *sang*, *sung*, *singing*), etc., although it belongs to a single complete family. We depart from standard terminology only for convenience: partial families will be much more important for our purposes than complete ones, hence we want to be able to use for them the simpler term 'family'. Also notice that a family may mix inflectionally and derivationally-related words (e.g. (*sing*, *sang*, *singer*) is a family), and complete families typically do: our definitions do not rely in any way on the distinction.

We may now turn to the definition of a paradigmatic system. This is based on aligned pairs of morphologically-related words.

(5)   Given two ordered pairs of morphologically related words (w_1, w_2) and (w_3, w_4), we say that the two pairs are ALIGNED if the same content relation holds between them: there is some content relation R_c such that $R_c(w_1, w_2)$ and $R_c(w_3, w_4)$. We call R_c the ALIGNING RELATION.

It is crucial here that the relevant notion of alignment is purely based on content, rather than form: pairs of words are aligned if they contrast in content in the same way. Whether they also contrast in form in the same way is beside the point. Hence for instance, the word pairs (*random*, *randomize*), (*class*, *classify*) and ($order_N$, $order_V$) are all aligned through the causative relation (Plag 1999). Also note that we take

---

[4]See Strnadová (2014, chap. 4) for relevant discussion of measures of regularity of the morphological relation between two words. Strnadová argues that the generalisability of a pattern across word pairs is a more relevant measure than similarity between the two words, although some combination of these two criteria is probably optimal.

[5]Taking families to be tuples rather than sets will be convenient when defining alignment below. Nothing however hinges on the order of the elements of the family.

no stance as to how general or particular the content relation used in the alignment must be: the definitions above are compatible both with coarse-grained and fine-grained views. Hence for instance for the purpose of some study we may say that (*random*, *randomize*) is aligned with (*hospital*, *hospitalize*) on the basis of a rather abstract relation "Verb denoting an action related to the noun" while other contexts may push us to focus on a more fine-grained classification where the two pairs instantiate different (causative vs. locative) relations.

We can now finally turn to the definition of a paradigmatic system. The idea is that a paradigmatic system combines families that are aligned in exactly the same way: for any pair of words in a family, there is a corresponding pair of words in each other family that is linked by the same content relation. The definition ensures that the exact same relations are used across pairs of words in all families.

(6)   A PARADIGMATIC SYSTEM is any set of morphological families of the same arity $P = \{(w_1^1, \ldots, w_n^1), \ldots, (w_1^m, \ldots, w_n^m)\}$ such that for any strictly positive $i, j \leq n$, all of $(w_i^1, w_j^1), \ldots, (w_i^m, w_j^m)$ are aligned pairwise by the same aligning relation.

Our construction of the notion of a paradigmatic system can be seen as a mathematically and conceptually sounder version of the view of lexical organization proposed by Bochner (1993, 66–74). Partial morphological families correspond to Bochner's 'cumulative' sets of lexical entries structured by pairwise bidirectional 'morphological rules'; his 'cumulative patterns' characterizing collections of cumulative sets can be seen as abstract descriptions of what we call paradigmatic systems.

2.2 Illustration

It should be obvious to the reader that canonical collections of inflectional paradigms for the same part of speech form a paradigmatic system as defined here. Consider French adjectives: these have four forms that contrast in gender and number. The collection of these four forms obviously forms a morphological family structured by 6 content relations ('*x* is the M.SG word corresponding to the M.PL word *y*', '*x* is the M.SG word corresponding to the F.PL word *y*', etc.). As each family is structured by the same set of relations, they give rise to pairwise alignments, and hence collections of such families form paradigmatic systems. This is depicted graphically in three dimensions in Fig. 1, where horizontal planes represent morphological families and vertical planes represent series of aligned pairs. The horizontal arrows highlight the fact that all members of a family are morphologically related, and the vertical dotted lines align words that participate in the same network of content relations and associate them with a name, here a combination of morphosyntactic properties.

The same reasoning applies to families of derivationally related words. Consider families consisting of a French verb and matching Agent noun and Action noun, as illustrated in Fig. 2. This time the relevant content relations are the following: '*x* is a noun denoting a set of individuals typically acting as agents of events of the type denoted by verb *y*' (Verb, Agent noun), '*x* is a noun denoting the same set of events as verb *y*' (Verb, Action noun), and '*x* is a noun denoting a set of individuals typically acting as agents of events of the type denoted by noun *y*' (Agent noun, Action noun).

Fig. 1 A French paradigmatic
system based on
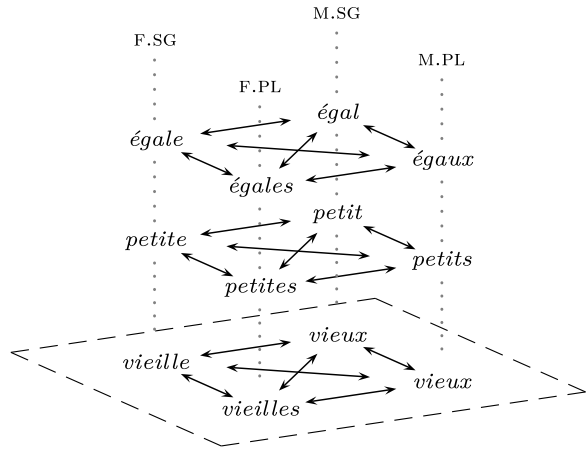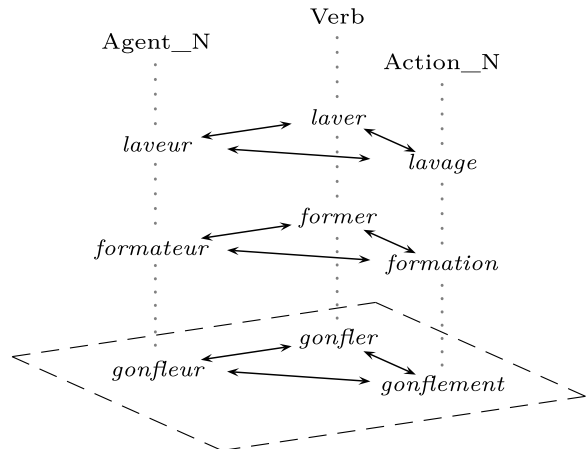inflectionally-related words

Fig. 2 A French paradigmatic
system based on
derivationally-related words

Note that our definitions are completely agnostic to the inflection-derivation distinction. Hence it is entirely possible to study as paradigmatic systems collections of families that mix derivational and inflectional relations. Fig. 3 provides a simple illustration.

Also note that, by design, we defined paradigmatic systems as PARTIAL both in the vertical and the horizontal directions: a system may consist of families that are not exhaustive, and the set of families is not necessarily exhaustive either. In this sense the illustrations in Figs. 1, 2 and 3 are illustrations of true (but small) paradigmatic systems, not of toy datasets. Partiality is crucial to being able to reason about paradigmatic organization while doing justice to the fine structure of complete morphological families. It is also convenient in multiple ways. One notable advantage is that it allows us to abstract away from the inflection-derivation opposition, while still focusing on inflection proper or derivation proper where relevant; for instance, when reasoning about derivational relationships, it makes full sense to look at morphological families consisting only of citation forms, as in Fig. 2.

Fig. 3 A French paradigmatic system mixing inflectional and derivational relations
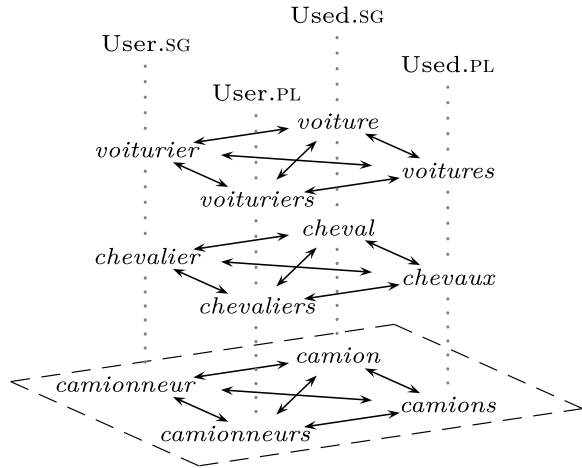
Hence there is basic plausibility to the claim that the notion of a paradigmatic system captures some basic intuitions of what constitutes a system of inflectional paradigms in such a fashion that these intuitions can be applied to any set of morphological families, whether the words in the families are inflectionnally related, derivationally related, or both. In the next subsection we address possible objections to that claim.

2.3 Discussion

In this paragraph we discuss limitations of our definition of paradigmatic systems and justify them. In each case, we argue that inflectional paradigms and derivational families are more similar than is sometimes suggested, and that paradigmatic systems capture the clear common properties while abstracting away from issues that are currently ill-understood.

2.3.1 Paradigm structure

Our definition of a paradigmatic system rests on the idea of a systematic contrast between all pairs of cells. Both *systematicity* and *contrast* are important here.

We take content-based contrast between words to be at the core of paradigm structure (Štekauer 2014). This is not highlighted in the Word and Paradigm traditions, where, following Matthews (1974, p. 136), paradigm cells are reified as sets of morphosyntactic properties, and the main locus of attention is the realization of those properties. However, there is no contradiction here. Consider the following representative definition (see also Carstairs 1987, 48–49):

(7)   "[...] we define the paradigm of a lexeme L as a complete set of cells for L, where each cell is the pairing of L with a complete and coherent morphosyntactic property set (MPS) for which L is inflectable."

(Stump and Finkel 2013, p. 9)

According to this definition, each cell in an inflectional paradigm contrasts with all the others by expressing a different morphosyntactic property set (MPS). The MPSs in effect provide a simple way of explicating the content-based aligning relations that structure a paradigmatic system. In fact this is exactly what we did when commenting the system of French adjectives in Fig. 1: we relied on morphosyntactic property sets to make explicit how aligned pairs of words contrasted. In effect then, any inflection system amenable to a description in terms of inflectional paradigms as defined in (7) trivially constitutes a paradigmatic system.

One interesting characteristic of Stump and Finkel's definition of paradigms is that it makes no commitment to a particular paradigm structure. Other authors differ here, and place a multidimensional organization of contrasts, in terms of morphosyntactic categories or features, at the heart of the concept. The quotation in (8) is representative.

(8)   "A paradigm is an n-dimensional space whose dimensions are the attributes (or features) used for the classification of word forms. In order to be a dimension, an attribute must have at least two values. The cells of this space can be occupied by word forms of appropriate categories."

(Wunderlich and Fabri 1995, p. 266)

If one takes such a multidimensional organization to be crucial to the notion of an inflectional paradigm, then it might seem that our notion of a paradigmatic system is missing something central. Moreover, under such a view, parallels between inflectional paradigms and structured derivational families may seem ill-advised, as derivational families are not standardly taken to have such shapes.

We contend that neither conclusion holds, and that the problem lies with the definition in (8). First, not all inflectional paradigms give rise to nice and clean multidimensional contrasts. This is typically the case for declensions in familiar Indo-European languages (with two dimensions of Case and Number for nouns, and a third dimension of Gender for adjectives), but systems with a single dimension (e.g. grade on adjectives, number on nouns) are common, as are systems where some contrasts are neutralized in some circumstances. This is the common situation in conjugation: in language after language, aspect distinctions are neutralized in the subjunctive, agreement distinctions are neutralized in the infinitive, etc. As a result, not all paradigm cells can be characterized as contrasting with other cells in *all* dimensions, although each paradigm cell contrasts with all the others holistically.

Second, some derivational patterns do support an organization in terms of multidimensional contrasts. Consider for instance families such as those in (9). These are clearly structured by contrasts between 'ideology' and 'advocate of ideology' on the one hand, and '*X*' and 'opposed to *X*' on the other hand.

(9)   (*communist*, *communism*, *anti-communist*, *anti-communism*)
        (*fascist*   , *fascism*   , *anti-fascist*   , *anti-fascism*   )

We conclude that organization into multidimensional contrasts is neither general of all inflectional paradigms nor exclusive to them. Hence it seems ill-advised to place that property at the heart of a definition of paradigms, and there is no downside to using the more permissive definition of a paradigmatic system proposed above for

purposes of describing inflection. Paradigm structure is an important question, but paradigm structures are diverse, thus the definition of a paradigm should not be too restrictive.

2.3.2 Doublets and overabundance

The definition of a paradigmatic system in (6) applies only to (partial) morphological families all of whose members are in pairwise contrast of content. As a result, we are bound to disregard situations of items with indistinguishable content within a family.[6] On the face of it, this seems problematic. On the side of inflection, situations of OVERABUNDANCE, where multiple forms fill the same cell of an inflectional paradigm, have received much attention in the recent past, and are found to be a lot more widespread than used to be thought (Thornton 2011, 2012, forthcoming). As Thornton discusses, rival overabundant forms may exhibit no contrast at all, or give rise to linguistically or sociolinguistically-conditioned variation. On the side of derivation, the study of affix rivalry has revealed the widespread existence of situations where doublets are lexicalized, with more or less specialization. For instance, a recent study by Fradin (in press) of rivalry between French *-age* and *-ment* contrasts situations of free variation (*rançonnage* vs. *rançonnement* 'ransoming'), partial specialization (*emballage* 'wrap' or 'wrapping' vs. *emballement* 'wrapping'), and complete specialization (*pliage* 'folding' vs. *pliement* 'bending').

The preceding examples suggest that the situation is strikingly similar in inflection and derivation, and that a fully adequate theory of paradigmatic contrast in morphology should take into account the existence of such situations of absence of contrast or unsystematic contrast. Doing this at a satisfactory level of detail is beyond the scope of the present paper. However two remarks are in order.

First, the definition in (6) allows us to take into account overabundance and other doublets to some extent. It is crucial here that we have taken morphological families to be partial, and allowed ourselves to define aligning relations at varying levels of granularity. Given this, we can align multiple morphological families that have some words in common and hence treat doublets as parallel citizens within a paradigmatic system. For instance, the following collection of families forms a paradigmatic system.

(10)　　(*voler*　　　, *vol*　　　　　, *voleur*　　　)
　　　　(*rançonner*, *rançonnage*　　, *rançonneur*)
　　　　(*rançonner*, *rançonnement*, *rançonneur*)

It would then be a rather simple exercise to define a higher-order notion of a paradigmatic system where morphological families are tuples of sets of words (rather than

---

[6]Remember that our notion of content encompasses morphosyntactic properties: hence two words may be strict synonyms but still not have the same content. This is the case of contrasting forms of a lexeme differing only in the realization of contextual inflection, such as contrasting forms of French adjectives discussed above.

tuples of words) and the two last rows of (10) are collapsed into a single row where the central cell has two elements.[7]

(11)   ({*voler*}   , {*vol*}   , {*voleur*}   )
        ({*rançonner*}, $\left\{\begin{array}{l} \textit{rançonnage,} \\ \textit{rançonnement} \end{array}\right\}$, {*rançonneur*})

Second, the true difficulty here is to decide under what conditions exactly the properties of two words are similar enough that they should be considered to fill the same cell in a paradigm rather than distinct paradigm cells. This is a very hard discretization problem, where gradient and multidimensional notions of similarity and contrast need to be mapped to a categorical distinction. This problem arises in all extant approaches to both inflectional and derivational morphology. Hence, as unsatisfactory as the current proposal may be, it is no worse than alternatives.

2.3.3 Gaps and defectiveness

Another striking property of derivational families that is not directly addressed by the definition in (6) is their possible incompleteness: more often than not, two derivational families contrast by the absence of a relevant member in one of the two.

Consider the example of pairings of French nouns denoting a game or sport and nouns denoting a practitioner of that activity. Most games and sports do have a dedicated matching agent noun, but a few don't: *échequiste* is barely attested, the periphrastic expression *joueur d'échec* 'chess player' being overwhelmingly preferred; and no synthetic word is attested for a go player.

(12)   *bridge*   *bridgeur*
        'bridge'   'bridge player'
        belote   *beloteur*
        'belote'   'belote'
        boules   *bouliste*
        'boules'   'boule player'
        échec   *?échequiste*
        'chess'   'chess player'
        go   —
        'go'   'go player'

It is important to remember that such gaps are not particular to derivational families, but are also found in inflectional paradigms. The phenomenon of inflectional defectiveness is exactly parallel (Baerman et al. 2010; Sims 2015). In inflection as in derivation, gaps are sometimes clearly motivated, sometimes arbitrary. Likewise, in both domains, gaps have an epistemologically uncertain status, as attestations often can be found for words that native speakers judge dubious.

The definition in (6) does not accommodate within a single paradigmatic system datasets such as those in (12): by definition, all morphological families in a paradig-

---

[7]Note that this is exactly parallel to the way Bonami and Stump (2016) suggest to address overabundance in Paradigm Function Morphology.

matic system must have the same size. It does allow one to define smaller paradigmatic systems that either leave out an offending family (e.g. that of *go* in (12)) or an offending vertical series of words. As in the case of doublets, it would not be hard technically to accommodate gaps by defining a higher order notion of morphological family based on tuples of sets of words and allowing empty sets; for (12) we would have the (higher order) paradigmatic system in (13).

(13)　　({*bridge*}, {*bridgeur*}　)
　　　　({*boules*}, {*bouliste*}　　)
　　　　({*échec*}　, {*échequiste*})
　　　　({*go*}　　, { }　　　　　)

Again, the difficulty lies not in the technical definition, but in the clear delimitation of what constitutes a gap.

2.3.4 Suppletion

The definitions above all depend on a notion of 'morphological relatedness' which we deliberately left partially underspecified. However, the definition in (2) is purposefully specific enough to exclude situations of undisputable suppletion. Let us briefly justify this position.

　　Paraphrasing Mel'čuk (1976, 45), within the conceptual framework defined in this section, two words w and w' can be said to entertain a relation of suppletion if and only if (i) from the point of view of content, the pair (w, w') is aligned with some independently existing paradigmatic system: that is, there are pairs of words $(w_1, w'_1), \ldots, (w_n, w'_n)$ that are pairwise morphologically related and that enter the same content relation as (w, w'); but (ii) from the point of view of form, there is no reason to consider w and w' to be related. Notice that the identification of suppletion is dependent on the existence of non-suppletive pairs. This is the main conceptual reason for excluding suppletive pairs from initial consideration when defining a paradigmatic system.

　　As in the case of overabundance and gaps above, one may entertain the idea of a higher-order notion of a paradigmatic system, where morphological families in complementary distribution are combined in macro-families. For instance, consider the small dataset in (14) illustrating suppletion in French conjugation.

(14)

lave	*laves*	*lave*	*lavons*	*lavez*	*lavent*
'wash.1SG'	'wash.2SG'	'wash.3SG'	'wash.1PL'	'wash.2PL'	'wash.3PL'
finis	*finis*	*finit*	*finissons*	*finissez*	*finissent*
'finish.1SG'	'finish.2SG'	'finish.3SG'	'finish.1PL'	'finish.2PL'	'finish.3PL'
vais	*vas*	*va*	—	—	*vont*
'go.1SG'	'go.2SG'	'go.3SG'			'go.3PL'
—	—	—	*allons*	*allez*	—
			'go.1PL'	'go.2PL'	

From this data we can deduce the existence of a paradigmatic system with six coordinates comprising words in the families of *laver* 'wash' and *finir* 'finish' but not *aller*

'go' (15a), and two paradigmatic systems containing more families but fewer coordinates (15b–c). From this observation, and that of the existence of a suppletive relation between *v-* words and *all-* words, we may postulate the existence of a macro-family combining them.

(15)   a.   (*lave*, *laves*, *lave*, *lavons* , *lavez* , *lavent*)
              (*finis*, *finis* , *finit*, *finissons*, *finissez*, *finissent*)
      b.   (*lave*, *laves*, *lave*, *lavent*)
              (*finis*, *finis* , *finit*, *finissent*)
              (*vais*, *vas* , *va* , *vont*   )
      c.   (*lavons* , *lavez*)
              (*finissons*, *finissez*)
              (*allons*   , *allez*)

The construction sketched above would need to be made more precise to be fully convincing; but arguably it replicates the reasoning of a descriptive linguist deciding that *va* and *allez* belong to the same paradigm.

It is important to stress that, since we purposefully do not presuppose a distinction between inflection and derivation, suppletion as defined here applies equally to both domains (Dressler 1985; Mel'čuk 1994). As Boyé (2006, 297) stresses, suppletion presupposes paradigmatic organization, which is one reason that some morphologists are skeptical about its use in the domain of word formation. But since we are precisely extending paradigmatic organization over the inflection-derivation divide, there is no principled reason to maintain such skepticism. Hence the data in (16) provides evidence for considering *pierre* 'stone' and *pétrifier* 'petrify' to entertain a suppletive relation, and for postulating a macro-family containing both.

(16)   *gaz*   *gazeux*   *gazéifier*   *gazéification*
       'gas'   'gaseous'   'gasify'   'gasification'

       cancer   *cancéreux*   *cancériser*   *cancérisation*
       'cancer'   'cancerous'   'become cancerous'   'cancerization'

       pierre   *pierreux*   —   —
       'stone'   'stony'

       —   —   *pétrifier*   *pétrification*
                             'petrify'   'petrification'

The accomodation of suppletion composes with that of doublets in subtle ways. As a case in point, consider the situation in (17): the families of *prison* and *carcéral* align but are not in complementary distribution, with synonyms *emprisonner* and *incarcérer* filling the same slot. Accommodating such situations requires more elaboration on the conditions under which two basic families may combine in a macro-family.[8]

---

[8]Also note that combinations of suppletion and overabundance are also found in the context of inflection, as discussed in Grossman and Thornton (forthcoming), Thornton (forthcoming).

(17) *commerce* *commerçant* *commercial* *commercialiser*
      'shop'   'shopkeeper' 'commercial' 'to market'

      école   *écolier*   *scolaire*   *scolariser*
      'school' 'schoolboy' 'educational' 'to send to school'

      prison *prisonnier* —   emprisonner
      'prison' 'prisoner'   'to imprison'

      —   —   *carcéral*   *incarcérer*
                            'of prison'   'to imprison'

2.3.5 Stability of contrasts

The definition of paradigmatic systems presupposes that the contrast in content between pairs of words in different morphological families is stable enough that alignment of families based on content is possible. Conventional wisdom holds that this is obviously true for inflectionally related words (see e.g. Robins 1959; Matthews 1974; Wurzel 1989; Stump 1998): *birds* contrasts with *bird* in the same way as *cats* contrasts with *cat*, *committees* with *committee*, and *flights* with *flight*. On the other hand, it is generally accepted that this does not hold for derivationally-related words, whose meaning often drifts independently.

There certainly is some reality to the difference between inflection and derivation here. However one may question whether this is enough of a problem to make paradigmatic systems inoperative when speaking of derivationally related words.

First, recognizing that the difference is real does not entail that it is categorical. A moment's thought shows that the semantic stability of inflectional contrast is not quite absolute. Semantic drift sometimes does affect inflected words in their own right. Relevant here are cases of pluralia tantum emerging from idiomatization of plural nouns, e.g. *clothes*, originally the plural form of *cloth*. This has consequences for the stability of contrast idea. Granted, we normally take CLOTH and CLOTHES to be distinct lexemes—and hence the fact that *clothes* does not have the predictable content of a plural for *cloth* seems irrelevant. But the sole reason for assuming lexemic distinctness is the fact that there is no singular form with the meaning of *clothes*. In other words, if we presuppose that inflectionally-related forms must have parallel meanings, then *cloth* and *clothes* can't be inflectionally-related. But if we drop that presupposition, there is no obvious reason to reject the hypothesis that they are different forms of a polysemous lexeme. So we can't dismiss the (*cloth*, *clothes*) relation on the basis of lexemic identity without running into circular reasoning. This suggests that even within inflection, parallelism of contrast cannot be taken to be absolutely strict. In the absence of a quantitative measure allowing one to conclude that contrasts are *more* stable in inflection than in derivation, we have no reason to conclude to a clear difference between the two types of systems.

Second, we have been explicit about the fact that paradigmatic systems rely on aligning relations that may make more or less fine-grained contrasts. Thus the study of paradigmatic systems does not force us to establish perfect parallelism of content between morphological families, but only parallelism that is good enough to be able to align them in a specific way. This seems possible in practice in many instances,

Table 1 Partial inflectional paradigms of a few Czech nouns

	NOM.SG	GEN.PL	Translation
(a)	hrad	hradů	'castle'
(b)	žena	žen	'woman'
(c)	táta	tátů	'dad'
(d)	stavení	stavení	'building'

Table 2 Partial inflectional paradigms of a few French toponyms and related demonyms

	AREA			INHABITANT		
	Orthography	Phonology	Translation	Orthography	Phonology	Translation
(a)	France	fʁɑ̃s	'France'	Français	fʁɑ̃sɛ	'French'
(b)	Russie	ʁysi	'Russia'	Russe	ʁys	'Russian'
(c)	Albanie	albani	'Albania'	Albanais	albanɛ	'Albanian'
(d)	Corse	kɔʁs	'Corsica'	Corse	kɔʁs	'Corsican'

even if there are some hard cases. We thus conclude that the observed instability of contrasts in derivational families does not invalidate the study of their paradigmatic properties.

2.4 Fruitful analogies between inflection and derivation

Having justified that the notion of a paradigmatic system may be applied to collections of derivational families without creating conceptual problems, we now show that such an application allows one to draw useful parallels between inflection and derivation, and to extend to derivational families useful analytic concepts first deployed in the context of inflection. By way of illustration, we will compare properties of partial paradigms of Czech nouns in the NOM.SG and GEN.PL, as illustrated in Table 1, and the system formed by French toponyms and their related demonyms, as illustrated in Table 2.[9] The choice of these two datasets is entirely opportunistic: although the rest of the paper is focused on French, we use Czech rather than French to illustrate typical inflectional patterns, as French inflection is notoriously odd.

Inflectional systems of any complexity usually exhibit INFLECTION CLASSES: different lexemes use different inflection patterns, and lexemes may be grouped in classes on the basis of the pattern they instantiate. The Czech data in Table 1 readily illustrate this: each row is illustrative of a different inflection class. The property leading to the position of multiple inflection classes we call DIFFERENTIAL EXPONENCE: different lexemes signal the same morphosyntactic contrast by different pairings of exponents: in the present instance we have the alternations $X \sim X\mathring{u}$ vs. $Xa \sim X$ vs. $Xa \sim X\mathring{u}$ vs. $X \sim X$.

---

[9]We rely on conventional orthography in the case of Czech, since it is transparent enough. For French we use phonemic transcriptions, as mute orthographic letters introduce confusion as to the nature of the patterns of alternation.

It is worth noting that the paradigmatic systems of derivationally-related words also have that property. In the case of the French system in Table 2, we see the alternations $X \sim X\varepsilon$ vs. $Xi \sim X$ vs. $Xi \sim X\varepsilon$ vs. $X \sim X$. These are directly parallel to those exhibited by Czech nouns. Thus we conclude that, although it would make little sense to say that derivational paradigms have inflection classes, the underlying notion of differential exponence applies to both inflectional and derivational paradigmatic systems.

The two small systems in Tables 1 and 2 exhibit remarkable distributions of exponents that have been of interest to inflectional morphologists but are seldom commented on in the context of derivation. First, both exemplify DIFFERENTIAL ZERO EXPONENCE: depending on which family is under examination, either one or the other of a pair of contrasting words may carry no exponent. The contrast between NOM.SG and GEN.PL in Slavic declensions is a celebrated example of that situation, at least since Jakobson (1939). It is worth remembering that the same situation is often found in derivational pairs, as exhibited in Table 2 with the contrast between the $X \sim X\varepsilon$ and the $Xi \sim X$ pattern.

Second, inflection systems often exhibit HETEROCLISIS (Stump 2006): some paradigms use an exponence strategy that is a hybrid of two others. Thus, in Table 1, the contrast between NOM.SG and GEN.PL for TÁTA is marked by the hybrid combination of exponents from the first two patterns: $Xa \sim X\mathring{u}$. An analogous heteroclite strategy can be observed also in case of word formation. In Table 2, the contrast between Area and Inhabitant for ALBANIE is marked by the pattern $Xi \sim X\varepsilon$ which combines the exponents of the first two rows.

Finally, we note that inflection and derivation give rise to comparable patterns of SYNCRETISM (see e.g. Baerman et al. 2005): the same form is used in multiple cells in a paradigm. This clearly characterizes both the (*stavení*, *stavení*) pair in Table 1 and the (*Corse*, *Corse*) pair in Table 2.

In inflection, different paradigms often give rise to different patterns of syncretism. To see this, we need to examine larger paradigmatic systems. Consider the system in Table 3. We observe several distinct situations: for HOST, syncretism appears in cells GEN and ACC and also in cells DAT and LOC with multiple syncretic forms in each cell; For LINGVISTA and VĚTA, it is distributed in cells DAT and LOC even though the marking is different as the two nouns belong to different inflection classes; For MOST and KOST, the syncretism occurs between NOM and ACC and also between cells GEN, DAT and LOC, but for MOST LOC can have two different forms out of which only one is syncretic with other wordforms in the paradigm. Finally, in case of MĚSTO, there is syncretism between NOM and ACC similarly to KOST and MOST, but the second syncretism in the paradigm between DAT and LOC is more similar to the LINGVISTA or HOST case, but again it has a second wordform in LOC that is not syncretic with any other form.

A comparable situation is found in paradigmatic systems of derivationally-related words (Strnadová 2014). Table 4 presents a few examples of nouns denoting an Institution and a Member of that institution, each morphologically related to a relational adjective (which we paraphrase as 'of Institution' and 'of Member'). For ACADÉMIE, SÉNAT and MINISTRE, the adjectives related to these nouns are syncretic (have the same form with a contrast in meaning). In case of ÉCOLE, the adjective related to the

Table 3 Partial inflectional paradigms of a few singular Czech nouns

NOM	GEN	DAT	ACC	LOC	INS	Translation
host	hosta	hostovi, hostu	hosta	hostovi, hostu	hostem	'guest'
lingvista	lingvisty	lingvistovi	lingvistu	lingvistovi	lingvistou	'linguist'
most	mostu	mostu	most	mostu, mostě	mostem	'bridge'
věta	věty	větě	větu	větě	větou	'sentence'
kost	kosti	kosti	kost	kosti	kostí	'bone'
město	města	městu	město	městě, městu	městem	'city'

Table 4 Partial derivational paradigms of a few French nouns and adjectives

INSTITUTION		MEMBER	OF INSTITUTION	OF MEMBER
académie	'academy'	académicien	académique	académique
sénat	'senate'	sénateur	sénatorial	sénatorial
ministère	'ministry'	ministre	ministériel	ministériel
école	'school'	écolier	scolaire	écolier
lycée	'high school'	lycéen	lycéen	lycéen
parlement	'parliament'	parlementaire	parlementaire	parlementaire

institution is SCOLAIRE, while there is a syncretism between the nouns for member and the related adjective: ÉCOLIER. Finally, a different pattern of the distribution of syncretism is presented by LYCÉEN and PARLEMENTAIRE which is the form used for a member, and the two adjectives of institution and of member.

This short exploration thus indicates that the formal makeup of paradigmatic systems is on the whole very similar, whether these consist of inflectionally or derivationally-related words.

3 Predictability in derivational paradigms

In the previous section we established that analytic concepts familiar from inflection were useful in the description of paradigmatic relations between derivationally-related words. In this section we turn to a different aspect of paradigm structure, the extent and nature of predictability relations between words. We take as our starting point Ackerman et al.'s (2009) celebrated PARADIGM CELL FILLING PROBLEM (PCFP):

(18)　　What licenses reliable inferences about the inflected (and derived) surface forms of a lexical item?

(Ackerman et al. 2009, 54)

Note that, although Ackerman et al. explicitly take the PCFP to be relevant for both inflectionnally and derivationally related forms, most of the ensuing literature has focused on inflection. Here we explicitly address the PCFP for derivationally related

words, and argue that the same type of paradigmatic organization is apparent in the data that may help speakers solve the problem. We first review the extant literature, and highlight two central properties of inflectional paradigms that have been uncovered by previous research: DIFFERENTIAL PREDICTABILITY and JOINT PREDICTIVENESS. We then present a dataset consisting of a large paradigmatic system of French derivationally related words, and apply to this dataset the very same measures originally designed for the study of inflection. The conclusion is that derivational paradigms exhibit the very same properties identified in inflection.

3.1 Addressing the PCFP in inflection

3.1.1 Quantitative assessment of implicative structure

As formulated above, the PCFP is a question about the structure of morphological systems.[10] Ackerman et al. argue that the crucial aspect is what Wurzel ([1984]) calls the IMPLICATIVE STRUCTURE of paradigms: there are reliable implicative statements linking paradigm cells that allow one to predict the wordform filling one cell from the wordform filling another. Two sample implicative statements are given in ([19]).

(19) a.   If the Present 3SG form of an English verb is $X + s$, then the present participle of that verb is $X + ing$

  b.   If the base form of an English verb is X, then the past form of that verb is $X + ed$

These two examples illustrate two important properties of implicative structure. First, implicative statements may hold of any pair of cells in the paradigm; there is no reason *a priori* to restrict attention to prediction from a 'base form', as speakers will sometimes be in a situation where they know only a non-base form and need to infer another form. Second, reliable implicative statements need not be categorical: while not all English past tenses end in -*ed*, ([19]b) still constitutes a useful resource for inference in the absence of definite information on a verb's past form. These two observations suggest that a detailed study of the PCFP should examine how it applies to the prediction of any paradigm cell from any other cell, and use quantitative modeling to address the variable reliability of implicative statements.

The analytic tool of choice to address this question has been the information-theoretic concept of CONDITIONAL ENTROPY. Given two variables A and B, the conditional entropy of B knowing A, written $H(B \mid A)$, evaluates as a positive number how informative the value of A is on the value of B. A conditional entropy of 0 indicates that B is categorically predictable from A. Ackerman et al. ([2009]) and later

---

[10]Note that the PCFP is formulated as a question of linguistic structure (what licences inferences) rather than a question of psycholinguistic processing (what inferences are actually drawn by speakers). There is a long tradition of addressing the psycholinguistic question: wug tests and morphological priming tasks can both be seen as assessing the role of paradigmatic inferences in processing. However, the research questions in the service of which these tasks have been put often have little to do with the PCFP. In this paper we focus on the linguistic issue and on methods stemming from Ackerman et al.'s seminal paper.

Table 5 A small sample of French adjectival paradigms

	M.SG	M.PL	F.SG	F.PL
BANAL 'banal'	/banal/	/banal/	/banal/	/banal/
RIVAL 'rival'	/ʁival/	/ʁivo/	/ʁival/	/ʁival/
CHAUD 'hot'	/ʃo/	/ʃo/	/ʃod/	/ʃod/
FAUX 'false'	/fo/	/fo/	/fos/	/fos/

Table 6 Relevant random variables inferred from the data in Table 5

(a) M.SG ∼ M.PL		(b) M.SG$_{M.SG∼M.PL}$		(c) Conditional probabilities		
Value	Prob.	Value	Prob.	P(col | row)	$X ∼ X$	$Xal ∼ Xo$
$X ∼ X$	³/₄	$\{X ∼ X\}$	¹/₂	$\{X ∼ X\}$	1	0
$Xal ∼ Xo$	¹/₄	$\{X ∼ X, Xal ∼ Xo\}$	¹/₂	$\{X ∼ X, Xal ∼ Xo\}$	¹/₂	¹/₂

studies (Ackerman and Malouf 2013; Blevins 2016; Sims 2015; Mansfield 2016) apply the notion of conditional entropy to the PCFP by evaluating how hard it is to predict the exponent used in some paradigm cell on the basis of the exponent used in some other paradigm cell. Bonami and colleagues (Bonami and Boyé 2014; Bonami and Luís 2014; Bonami and Beniamine 2016) have argued that it is more satisfactory not to rely on a preexisting analysis of exponence, as such an analysis usually presupposes knowledge of the very paradigmatic contrasts that one is trying to deduce. They propose instead to examine the conditional entropy of the PATTERN relating cell *A* and cell *B* given aspects of the SHAPE of the form filling cell *B* that are relevant to pattern satisfiability. This they call the IMPLICATIVE ENTROPY from cell *A* to cell *B*. Here we quickly present and illustrate the central concepts on a toy dataset, and refer the reader to Bonami and Beniamine (2016) for detailed motivation, and a more precise presentation of the algorithms used to evaluate implicative entropy.

Let us consider the small sample of adjectives in Table 5, and assume that this is representative of the type frequency distribution of inflectional patterns in the language. We want to assess how difficult it is to predict the M.PL form of an adjective from its M.SG form. To this effect, we can define two relevant random variables, whose values are presented in Table 6.'M.SG ∼ M.PL' is a random variable over pairs of matching M.SG and M.PL forms, and characterizes the distribution of alternations between these forms. In this case two alternations are found: identical forms in 3 examples out of 4, and an alternation between a form ending in /al/ and a matching form ending in /o/ in 1 example out of 4. These observed frequencies are used to estimate the probability of each alternation type as indicated in Table 6: given what has been observed, if we pick a new adjective at random, we expect a 0.75 probability of choosing an adjective with identical M.SG and M.PL forms.

From this variable over pairs of forms we may now infer a new random variable over M.SG forms, labeled 'M.SG$_{M.SG∼M.PL}$'. The intuition here is the following. We want to classify possible shapes of M.SG adjectives on the basis of phonological characteristics that are relevant to determining which shape it could have in the M.PL. To this effect, we examine which of the values of the variable 'M.SG ∼ M.PL' are compatible with the phonological shape of each M.SG adjective. From this point of view,

it turns out that the only relevant distinction opposes adjectives with a M.SG ending in /al/, whose paradigm may instantiate either of the two values of 'M.SG $\sim$ M.PL', and adjectives whose M.SG does not end in /al/, which may possibly instantiate only the value $X \sim X$ of 'M.SG $\sim$ M.PL'. Given our observed lexicon, it appears that each of these two possibilities has an equal probability of 0.5.

Given these two random variables and their associated probability distributions, we can now compute conditional probabilities indicating how the shape of an M.SG form conditions which pattern of alternation it entertains with the M.PL. These are indicated in part (c) of Table 6. As can be seen there, if the M.SG does not end in *-al* (i.e. the value of 'M.SG$_{\text{M.SG}\sim\text{M.PL}}$' is '$\{X \sim X\}$'), only one outcome is possible, and hence a probability of 1 is attributed to the pattern $X \sim X$: we know for sure that the M.PL is identical to the M.SG. On the other hand, if the M.SG does end in *-al*, two patterns are available, $X \sim X$ and $Xal \sim Xo$. Because in our toy dataset each choice is witnessed exactly once, we attribute to each alternative an equal part of the probability mass, that is, $1/2$.

We are now in a position to compute the implicative entropy from the M.SG to the M.PL in our toy dataset, which is basically just a summary of the conditional probability distribution just discussed. In general, the conditional entropy of variable Y given variable X is the average of the entropy of Y given each possible value x of X, weighted by the probabilities of these values. Notice the crucial role played by conditional probabilities in the equation.

(20) $\qquad H(Y \mid X) = - \sum_{x \in X} P(X) \sum_{y \in Y} P(y \mid x) \log_2 P(y \mid x)$

In our particular case, we define the implicative entropy from M.SG to M.PL, noted $H(\text{M.SG} \Rightarrow \text{M.PL})$, as the conditional entropy between our two variables.

(21) $\qquad H(\text{M.SG} \Rightarrow \text{M.PL}) = H(\text{M.SG} \sim \text{M.PL} \mid \text{M.SG}_{\text{M.SG}\sim\text{M.PL}})$

$$= -\left(\tfrac{1}{2} \times \left(\tfrac{1}{2} \log_2 \tfrac{1}{2} + \tfrac{1}{2} \log_2 \tfrac{1}{2}\right) + \tfrac{1}{2} \times (1 \times \log_2 1)\right)$$

$$= -\left(\tfrac{1}{2} \log_2 \tfrac{1}{2} + 0\right) = \tfrac{1}{2}$$

This is then our estimation of how difficult it is on average to predict the M.PL of an adjective from its M.SG in the toy version of French encapsulated in Table 5. Conceptually, we reduced this question to the question of guessing the pattern of alternation relating the M.SG and the M.PL given the shape of the M.SG and statistical knowledge on the distribution of shapes in paradigms in the extant lexicon.

3.1.2 Differential predictability

An important property of inflection systems that emerges from all studies of the PCFP on different datasets is the property of DIFFERENTIAL PREDICTABILITY. Quality of prediction does not distribute in any simple way in the paradigm.[11] More generally,

---

[11]This can already be seen using the toy dataset in Table 5. Computations parallel to the ones in the preceding section reveal that predicting the M.SG from the M.PL is harder than the other way around: $H(\text{M.SG} \Rightarrow \text{M.PL}) = \tfrac{3}{4} \log_2 3 - \tfrac{1}{2} \approx 0.69$. This is due to the fact that in this dataset, there are more

Table 7 Unary implicative entropy from raw to column in the Flexique database

$H(\text{row} \Rightarrow \text{col})$	F.SG	F.PL	M.SG	M.PL
F.SG	—	0	0.213	0.231
F.PL	0	—	0.213	0.231
M.SG	0.641	0.641	—	0.018
M.PL	0.666	0.666	0.041	—

Table 8 Paradigms of a sample of French adjectives

	F.SG	F.PL	M.SG	M.PL
BANAL 'banal'	/banal/	/banal/	/banal/	/banal/
LOYAL 'loyal'	/lwajal/	/lwajal/	/lwajal/	/lwajo/
RÉGLO 'reliable'	/ʁeglo/	/ʁeglo/	/ʁeglo/	/ʁeglo/
VRAI 'true'	/vʁɛ/	/vʁɛ/	/vʁɛ/	/vʁɛ/
PRÊT 'ready'	/pʁɛt/	/pʁɛt/	/pʁɛ/	/pʁɛ/
FRAIS 'fresh'	/fʁɛʃ/	/fʁɛʃ/	/fʁɛ/	/fʁɛ/
NIAIS 'stupid'	/njɛz/	/njɛz/	/njɛ/	/njɛ/

a paradigm cell may be a good overall predictor or a good overall predictee, but, more often than not, cell X will be a good predictor of cell Y but a poor predictor of cell Z. For illustration, consider now the more realistic dataset consisting of the 11,252 French adjectives whose paradigms are documented in the *Flexique* database (Bonami et al. 2014b). Table 7 indicates the implicative entropy for each pair of cell in the four cell paradigm.

As calculations of implicative entropies rest on the examination of patterns of alternation and the statistical distribution of these patterns, we can examine the trace of the calculations and identify the underlying reasons for contrasting entropy values. The sample data in Table 8 illustrates the underlying reasons for the contrasts in implicative entropy for various pairs of cells. The feminine singular and plural are always identical except for the use of a final /z/ in so-called *liaison* contexts (not represented explicitly here). Hence they are entirely predictable from one another (entropy value 0), and have the same predictiveness and predictability for other cells. Predicting the M.PL from the M.SG is easy, but not trivial, because adjectives ending in *-al* in the singular may end in either *-al* (see BANAL) or *-o* (see LOYAL in the plural). This leads to uncertainty and an implicative entropy higher than zero. However, since adjectives in *-al* are not very numerous, and prediction is categorical for all other endings, the entropy is quite low at 0.018. The same situation holds when predicting the M.SG from the M.PL, except that this time the source of unpredictability is plurals in *-o*. The implicative entropy is a bit higher at 0.041, because plurals in *-o* are more prevalent than singulars in *-al*. Continuing to examine the table by rising entropy values, the M.SG forms are hard to predict from the feminine because of the diverging behavior of consonant-final feminines: either the masculine form is identical to the

---

instances where a M.PL has a form that does not fully determine that of its M.SG than the other way around.

feminine, or it differs by the absence of the final consonant. Since most feminine adjectives end in a consonant and the distribution of patterns is rather balanced, this leads to significantly higher entropy (0.213). The entropy is a tad higher still when predicting the M.PL instead of the M.SG, because of the added difficulty of predicting the outcome with *-al* endings. Finally, the highest entropy values are found when predicting feminine forms from masculine ones. This is due to the fact that, when a masculine form ends in a vowel, the feminine may be identical to the masculine or introduce any of a number of different final consonants. The existence of a choice with more than two possible outcomes explains the much higher entropy values.

This simple example illustrates the main lessons of studies of the PCFP relying on conditional entropy. First, traditional citation forms are not necessarily the best predictor cell, even where, as in French, they are both the morphosyntactically unmarked cell (used e.g. in situations of default agreement) and the most frequent cell. Second, the question of predictability needs to be asked separately for each pair of a predictor cell and a predicted cell: sources of uncertainty differ from case to case, and hence levels of predictability can also vary in drastic ways. Hence there is no evidence for one paradigm cell to be a credible 'unitary base' (Albright 2002) nor even for a hierarchy of such cells to be inferable in general, *contra* Bonami and Boyé (2002).

3.1.3 Joint predictiveness

In the preceding paragraph we focused on a subcase of the PCFP where inference relies on a single premise: we looked at how predictive some cell X is of some cell Y. In general though, speakers may be attuned to multiple forms of a lexeme, and the simultaneous knowledge of these multiple forms may allow for new inferences that are not available when looking at separate forms. Hence the JOINT PREDICTIVENESS of two forms may exceed the predictiveness of each of these forms in isolation.

To assess the joint predictiveness of multiple paradigm cells, Bonami and Beniamine (2016) generalize the definition of implicative entropy and the algorithms used to evaluate it to the case of n predictors. The intuition behind their definition is simple: predicting from multiple cells amounts to predicting on the basis of the joint knowledge of the shape of each predictor and the relationship between these predictors. In the binary case this amounts to the definition in (22), which generalizes trivially to the n-ary case.[12]

(22) $H(X, Y \Rightarrow Z) = H(X \sim Z, Y \sim Z \mid X_{X \sim Z}, Y_{Y \sim Z}, X \sim Y)$

Examining French and European Portuguese conjugation, Bonami and Beniamine conclude that in both systems, on average, joint knowledge of $n + 1$ paradigm cells

---

[12]To see how this works intuitively, consider again the toy dataset in Table 5, and prediction of the M.SG from joint knowledge of the F.SG and M.PL. Predicting the pattern relating M.SG and M.PL (i.e., the value of 'M.SG ∼ M.PL' on the basis of just the M.PL is hard, because plurals in /o/ may correspond to two kinds of singulars; that is, there is one value of 'M.PL$_{M.SG \sim M.PL}$' which corresponds to two possible values of 'M.SG ∼ M.PL'. However, if we add in knowledge of the relationship between M.PL and F.SG, no uncertainty remains: all lexemes with an Xal ∼ Xo alternation between M.SG and M.PL exhibit the same alternation between F.SG and M.PL.

Table 9 Average implicative
entropy for French adjectives
with *n* predictors

1 predictor	0.2966
2 predictors	0.1443
3 predictors	0.0044

Table 10 Partial sample datapoints from *Démonette* (format adapted for present purposes)

Lexeme 1		Lexeme 2		Formal relation	Relation type	Morphosemantic relation
Citation form	Stem	Citation form	Stem			
CAMPER 'to camp'	/kɑ̃p/	CAMPEMENT 'camp'	/kɑ̃pəmɑ̃/	$s_2 = s_1 + $/əmɑ̃/	$s_1 \rightarrow s_2$ by suffixation	L_2 is the Action noun of L_1
SCANDALISER 'to scandalize'	/skɑ̃daliz/	SCANDALE 'scandal'	/skɑ̃dal/	$s_1 = s_2 + $/iz/	$s_2 \rightarrow s_1$ by suffixation	L_2 is the Action noun of L_1
AIDER 'to help'	/ɛd/	AIDE 'help'	/ɛd/	$s_1 = s_2$	conversion pair	L_2 is the Action noun of L_1

leads to significantly higher predictiveness than joint knowledge of *n* cells. We illustrate here with our running example of French adjectives. As Table 9 shows, there is a sizeable drop of average entropy when going from one to two or from two to three predictors.

The usefulness of joint prediction provides a strong argument to the effect that paradigms are first class citizens of the morphological universe. What this result shows is that there are useful inferences to be drawn on the basis of simultaneous knowledge of two words, an inherently paradigmatic notion. It is thus particularly interesting to see whether such structure can be found in paradigmatic systems of derivationally related words.

3.2 The dataset

In the remainder of this paper we deploy the analytic tools presented above on a dataset of derivationally related words, and attempt to establish that the same quantitative properties observed in inflectional paradigms hold in the context of derivation. The data under consideration consists of French morphological families consisting of a Verb, an Action noun, and a masculine Agent noun. We first detail how the paradigmatic system under consideration was constructed.

The main source for the dataset is *Démonette*, a database of 20,493 derivational relations in the French lexicon (Hathout and Namer 2014). *Démonette* takes the form of a large table, where each row documents the relation between two derivationally related lexemes, and in particular encompasses a classification of formal morphological relations between the stems of the lexemes under examination and broad morphosemantic classification documenting whether a pair instantiates a (Verb, Action noun) relation, a (Verb, masculine Agent noun) relation, a (Verb, Result noun) relation, etc. Crucially for our purposes, these two classifications are independent of one another.

So for instance, as Table 10 illustrates, the same morphosemantic relation may be instantiated by affixation in either direction or by a conversion pair, in which case di-

Table 11 Sample families from the paradigmatic system extracted from *Démonette*

Verb	Action Noun	Agent Noun
{*abaisser*}	{*abaissement*}	{*abaisseur*}
'lower'	'lowering'	'depressor'
{*abandonner*}	{*abandon, abandonnement*}	{*abandonneur, abandonnateur*}
'abandon'	'abandonment'	'who abandons'
{*abattre*}	{*abattement, abattage*}	{*abatteur*}
'bring down/slaughter'	'reduction', 'felling/slaughter'	'slaughterer'
{*affamer*}	{ }	{*affammeur*}
'starve'		'responsible for starvation'
{ }	{*agriculture*}	{*agriculteur*}
	'agriculture'	'farmer'

rectionality of derivation is usually indeterminate in French (Tribout 2010). Another important observation is the broadness of the morphosemantic classification. For instance, doublets with different lexical meanings, such as *abattement* 'reduction' and *abattage* 'felling', both derived from *abattre* 'bring down', are classified in the same way. Likewise, strict agent nouns and deverbal instrument nouns in *-eur* (e.g. *moteur* 'motor') are not differentiated.

From *Démonette* we extracted all rows documenting the relation between a Verb and an Agent noun, a Verb and an Action noun, or an Agent noun and an Action noun. We then merged rows into structured families by fusing cells containing the same wordform labeled with the same morphosemantic type. This resulted in a paradigmatic system of 5,414 families, with many empty cells and many doublets. Table 11 exhibits some examples.

As the notion of implicative entropy is defined only for complete paradigms with no doublets, we reduced the dataset as follows. First, we dropped all morphological families containing a gap. Second, if a cell contained multiple fillers, we examined the relative frequency of the fillers in the FrWac web corpus (Baroni et al. 2009). If one of the cell members accounted for more than two thirds of the distribution, we kept that word as the 'main realization' of the cell, and dropped the other ones. If the relative frequency of the cell members was more balanced, we dropped that morphological family altogether. 1,331 families survived this selection process.

Finally, we relied on the *GLÀFF* lexicon (Hathout et al. 2014) derived from the French Wiktionary to provide phonemic transcriptions for the words in the dataset. As not all lexemes documented in *Démonette* are also present in the *GLÀFF*, this led to dropping more families. Our final dataset thus consists of 913 families of three derivationally-related words in phonemic transcription. Note that we used the citation forms of lexemes for prediction. A more thorough study would have examined the formal relations between all inflected forms of all lexemes, and determined whether different inflected forms contrast in predictiveness of derivationally related words. We doubt that this would have led to significantly different results in this particular instance.

Table 12 Unary implicative entropy for the French (Verb, Action Noun, Agent Noun) paradigmatic system

$H(\text{row} \Rightarrow \text{col})$	Verb	Action Noun	Agent Noun
Verb	—	1.115	0.709
Action Noun	0.101	—	0.269
Agent Noun	0.264	1.114	—

Table 13 Sample data from the French (Verb, Action Noun, Agent Noun) paradigmatic system

Verb		Action Noun		Agent Noun	
laver	/lave/	*lavage*	/lavaʒ/	*laveur*	/lavœʁ/
'wash'		'washing'		'washer'	
bâtir	/batiʁ/	*bâtiment*	/batimɑ̃/	*bâtisseur*	/batisœʁ/
'build'		'building'		'builder'	
contrôler	/kɔ̃tʁole/	*contrôle*	/kɔ̃tʁol/	*contrôleur*	/kɔ̃tʁolœʁ/
'control'		'control'		'controller'	
corriger	/koʁiʒe/	*correction*	/koʁɛksjɔ̃/	*correcteur*	/koʁɛktœʁ/
'correct'		'correction'		'corrector'	
former	/fɔʁme/	*formation*	/fɔʁmasjɔ̃/	*formateur*	/fɔʁmatœʁ/
'train'		'training'		'trainer'	
écrire	/ekʁiʁ/	*écriture*	/ekʁityʁ/	*scripteur*	/skʁiptœʁ/
'write'		'writing'		'writer'	
gonfler	/gɔ̃fle/	*gonflement*	/gɔ̃fləmɑ̃/	*gonfleur*	/gɔ̃flœʁ/
'inflate'		'inflating'		'inflater'	

3.3 Results

3.3.1 Differential predictability

We first comment results on prediction of one paradigm cell from another paradigm cell. Implicative entropy values are given in Table 12. As is immediately apparent, predictability is highly variable. Action nouns are the best predictors and Verbs are the best predictees. Predicting an Action noun from a Verb or from an Agent noun is almost exactly as hard, but predicting an Agent noun from a Verb is a lot harder than predicting it from an Action noun.

As in the case of inflection, we can now examine the patterns identified by the algorithm and their statistical distribution in the lexicon, so as to understand the underlying reasons for contrasts in predictability. Going through full tables of patterns would be tedious, so we will focus on the illustrative data in Table 13 and comment impressionistically on the prevalence of various patterns.

First, let us note that there is a clear reason why Action nouns are hardest to predict. In French as in many languages, there is a wide variety of strategies for contrasting verbs and Action nouns. As we saw in Table 10, the noun may be derived from the verb, or the verb derived from the noun, or they may form a conversion pair. More important in numerical terms is the fact that there is a sizeable number of distinct

highly productive ways of deriving Action nouns from verbs, including *-age* suffix-ation, *-ment* suffixation, *-(at)ion* suffixation, and conversion. Although the strategies are not quite equiprobable, and the suffixes exhibit some amount of phonotactic se-lectivity,[13] there is thus a lot of uncertainty, with an unpredictable choice to be made between the 4 main patterns $Xe \sim Xa\math!z$, $Xe \sim X\partial m\tilde{a}$, $Xe \sim Xasj\tilde{\partial}$ and $Xe \sim X$. Pre-dicting the Action noun from the Agent noun is barely easier: although the overall distribution of patterns is not exactly the same, sources of opacity and transparency balance one another.

Second, we note that predicting an Agent noun from an Action noun is a lot easier than predicting it from a Verb. This contrast is linked to the existence of two com-peting strategies for forming Agent nouns in French, and its interaction with the for-mation of Action nouns (Bonami and Boyé 2005, 2006; Bonami et al. 2009). Agent nouns may either be formed on the verb's basic stem, the stem normally apparent in the present participle, or on a Latinate stem that is the descendant of the Latin 'third stem' (Aronoff 1994).[14] In the latter case, the Latinate stem will most often end in *-at*, so that the two most prevalent patterns relating verbs to Agent nouns are $Xe \sim X\alpha\alpha$ and $Xe \sim Xat\alpha\alpha$; other possibilities exist but are not prevalent enough to have a strong influence on the entropy. Since the choice between these two possi-bilities is not predictable (e.g. the Verb noun *former* could just as well have led to an agent noun **formeur*), and they are close to be equiprobable, we find an implicative entropy close to 1. If we now look at prediction from Action nouns, the situation is radically different. Action nouns also may be formed on either the verb's basic stem or the Latinate stem, but suffixes select one or the other: *-ion* always attaches to the Latinate stem, *-age* and *-ment* select the basic stem or one of its allomorphs, never the Latinate stem. More importantly for our purposes, there is a strong correlation between the formation strategy used for the Action noun and that for the Agent noun: when both exist, with very few exceptions, the Agent noun is formed on the Latinate stem if and only if the Action noun also is. As a result, the patterns dominating the distribution are $Xa\math!z \sim X\alpha\alpha$, $X\partial m\tilde{a} \sim X\alpha\alpha$, $X \sim X\alpha\alpha$, and $Xj\tilde{\partial} \sim X\alpha\alpha$.[15] Since these patterns are mutually exclusive (a noun can't simultaneously end in /aʒ/ and in /əmã/ or in /aʒ/ and in /jɔ̃/), they do not lead to uncertainty. The residual level of entropy is due to rare situations of unpredictability, such as that which arises with *écriture~scripteur*.

Finally, consider prediction of Verbs from Agent nouns. There is little uncertainty here: if the noun ends in /atœʁ/, there is a high likelihood that it is formed on a Latinate stem, and that the verb ends in /e/.[16] If it has any other ending, it is very likely

---

[13]If the 4 strategies were strictly equiprobable, the implicative entropy would be higher than 2, the entropy associated with choosing the value of a variable with 4 equiprobable outcomes.

[14]The two formations also contrast in being matched with feminine nouns in *-euse* /øz/ or *-rice* /ʁis/ respectively.

[15]Note that this pattern applies both in regular cases like *formation~formateur*, where the Latinate stem itself is predictable from the verb's basic stem, and in cases like *correction~correcteur*, where the Latinate stem is unpredictable. This also contributes to making prediction of Agent nouns from Action nouns easy: Agent and Action nouns do not contrast in terms of unpredictable patterns of allomorphy.

[16]There are only five exceptions in the dataset, including *constateur* 'in charge of reporting' related to *constater* 'take notice'.

Table 14 Average implicative entropy for 1 or 2 predictor cells in the French (Verb, Action noun, Agent noun) paradigmatic system

1 predictor	0.595
2 predictors	0.196

Table 15 Binary implicative in the French (Verb, Action Noun, Event Noun) paradigmatic system

Predictors	Predicted	Binary entropy	Unary entropy	
			from 1st predictor	from 2nd predictor
Verb, Action_N	Agent_N	0.138	0.709	0.269
Verb, Agent_N	Action_N	0.444	1.115	1.114
Agent_N, Action_N	Verb	0.006	0.264	0.101

that it instantiates the pattern $X \alpha \text{в} \sim Xe$. Exceptions mostly occur with non-first conjugation verbs, which take different infinitive endings (e.g. *bâtisseur* could just as well have been related to a verb **bâtisser*), but these are rare enough to contribute little uncertainty: 88% of the verbs in our dataset are first conjugation. The situation is very similar when predicting Verbs from Action nouns. There is more diversity of patterns, since Action nouns instantiate diverse suffixes, but these patterns tend to be mutually exclusive; hence the only sizeable source of uncertainty is the verb's conjugation class.

In conclusion then, we see that paradigmatic systems of derivationally-related words give rise to exactly the same situation as those of inflectionally-related words in terms of differential predictabilities: predictability varies because of the exact relation between the predictor cell and the predicted cell, in such a fashion that each prediction relation gives rise to its own challenges.

3.3.2 Joint predictiveness

We finally turn to the issue of joint predictiveness. As Table 14 indicates, on average, predicting the third word from knowledge of the other two is about three times easier than predicting from a single word.

The averages in Table 14 are not quite enough to conclude however. Predicting from two cells can never be harder than it is to predict from the most predictive of the two cells; the minimum of the two corresponding unary implicative entropies is an upper bound on the binary entropy. In Table 15, we thus look at the binary entropy from each pair of predictor cells to the remaining predicted cell, and compare it to the corresponding unary entropies. In all three cases, the binary entropy is dramatically lower than even the lowest of the two unary entropies.

A striking result of Table 15 is near categoricity of predictability of verbs from pairs of associated Agent and Action nouns. This is due to the conjugated effects of patterns identified earlier. For instance, although the Agent noun *constateur* is a poor predictor of the corresponding verb, which could have been either *constater* or **conster*, the Action noun *constat* disambiguates. In fact, close examination of the results shows that the very small residual uncertainty is due to the existence of a few

verbs derived by conversion from nouns in -*ion* (Lignon and Namer 2010)—there are 6 such verbs in our dataset. Although they are quite rare, the existence of such verbs entails that, when predicting from an Action noun in -*ion* and an Agent noun, one must sometimes entertain the possibility that the verb be in -*ionner*, cf. (*inspection*, *inspecteur*, *inspecter*) vs. (*fonction*, *foncteur*, *fonctionner*).

This confirms that simultaneous knowledge of both predictors is a lot more useful than knowledge of even the most predictive of the two. This result is congruent with what we observed for inflection, and provides a strong argument to the effect that paradigmatic structure is just as important to the PCFP in derivation as it is in inflection. In the context of derivation, it takes an even stronger value, because of the expectation that a derived word's properties should follow from the properties of its base, modulo unpredictable idiosyncrasies due to idiomatization. We see here that, quite to the contrary, properties of a derived word that are unpredictable from its conventional base may be predicted when one takes other members of the morphological family into account.

4 Conclusions

Wurzel (1989) places predictability and implicative structure at the heart of inflectional paradigmatic systems:

(23)   The inflectional paradigms are, as it were, kept together by implications. There are no paradigms (except highly extreme cases of suppletion) that are not based on implications valid beyond the individual word, so that we are quite justified in saying that inflectional paradigms generally have an implicative structure, regardless of deviations in the individual cases.

<div align="right">(Wurzel 1989, 114)</div>

In this paper we have argued that the same property holds of paradigmatic systems of derivationally-related words. After defining, justifying and exemplifying a notion of paradigmatic systems applicable to both inflectional and derivational families of words, we have argued that derivational paradigmatic systems have the same kind of implicative structure as inflectional paradigmatic systems. Our central argument is the observation that the two types of systems share the two important properties of DIFFERENTIAL PREDICTABILITY and JOINT PREDICTIVENESS. First, each pair of paradigm cells is tied by a different set of implicative relations, giving rise to a specific pattern of partial predictability, measurable by unary implicative entropy. Second, we established empirically that, for any cell in a paradigm, predicting the content of that cell from joint knowledge of two other cells is significantly easier than predicting from either of the two other cells. These properties both show that the predictive structure of derivational families goes well beyond prediction of a derived word from a unique base.

The present study was limited in its ambition in at least two ways. On the conceptual side, we made no claim as to the fact that the lexicon as a whole constitutes a paradigmatic system, or that there are no relevant differences between inflection and derivation. We rather argued that some subparts of the lexicon consisting of either

inflectionally or derivationally-related words can soundly be seen as paradigmatic systems and exhibit comparable properties.

On the empirical side, the present study is limited by the representative character of the dataset it is based on. Although this does not affect our overall results, there is a lot of room for improvement here. First, the *Démonette* database of derivational relations does not document the lexicon in a well-defined corpus, but was compiled from a variety of resources with width rather than representativity in mind. Hence the statistical distribution of patterns in the dataset might have a complex relation to the true distribution. Second, as impressive as the coverage of that database is, it still covers quite a limited number of morphosemantic relations, which led us to focus on 3-cell paradigms. Third, we used a very crude selection method to reduce the number of doublets in the data, by assuming that the more frequent variant was the 'real' member of the paradigm. This again might lead to bias in the statistical distributions, since the frequency of variants might have a complex relation with the nature of their morphosemantic status.

We take these observations to suggest that significant progress in the documentation of morphological families needs to be made. Right now, in addition to sheer size, the bottleneck in the construction of such resources is the large-scale fine-grained classification of morphosemantic relations: in the long run we should not be satisfied with treating e.g. *équipage* 'crew' and *équipement* 'gear' as standing in the same morphosemantic relation to *équiper* 'equip'. We hope that the construction of finer and larger resources will allow in the near future detailed explorations of the fine structure of derivational paradigms.

References

Ackerman, F., & Malouf, R. (2013). Morphological organization: the low conditional entropy conjecture. *Language*, *89*, 429–464.

Ackerman, F., Blevins, J. P., & Malouf, R. (2009). Parts and wholes: implicative patterns in inflectional paradigms. In J. P. Blevins & J. Blevins (Eds.), *Analogy in grammar* (pp. 54–82). Oxford: Oxford University Press.

Albright, A. C. (2002). *The identification of bases in morphological paradigms*. PhD thesis, University of California, Los Angeles.

Anderson, S. R. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.

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

Baerman, M., Brown, D., & Corbett, G. G. (2005). *The syntax–morphology interface: a study of syncretism*. Cambridge: Cambridge University Press.

Baerman, M., Corbett, G. G., & Brown, D. (Eds.) (2010). *Defective paradigms*. Oxford: Oxford University Press.

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The wacky wide web: a collection of very large linguistically processed web-crawled corpora. In *Language resources and evaluation* (Vol. 43, pp. 209–226).

Bauer, L. (1997). Derivational paradigms. In G. Booij & J. van Marle (Eds.), *Yearbook of morphology 1996* (pp. 243–256). Dordrecht: Kluwer.

Becker, T. (1993). Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology. In G. Booij & J. van Marle (Eds.), *Yearbook of morphology 1993* (pp. 1–25). Dordrecht: Kluwer.

Behrens, L. (1995). Lexical rules cross-cutting inflection and derivation. *Acta Linguistica Hungarica*, *43*(1/2), 33–65.

Blevins, J. P. (2001). Paradigmatic derivation. *Transactions of the Philological Society*, *99*(2), 211–222.

Blevins, J. P. (2016). *Word and paradigm morphology*. Oxford: Oxford University Press.

Bochner, H. (1993). *Simplicity in generative morphology*. Berlin: de Gruyter.

Bonami, O., & Beniamine, S. (2016). Joint predictiveness in inflectional paradigms. *Word Structure*, *9*(2), 156–182.

Bonami, O., & Boyé, G. (2002). Suppletion and stem dependency in inflectional morphology. In F. Van Eynde, L. Hellan, & D. Beerman (Eds.), *The proceedings of the HPSG '01 conference* (pp. 51–70). Stanford: CSLI Publications.

Bonami, O., & Boyé, G. (2005). Construire le paradigme d'un adjectif. *Recherches Linguistiques de Vincennes*, *34*, 77–98.

Bonami, O., & Boyé, G. (2006). Subregular defaults in French conjugation. In *12th international morphology meeting*, Budapest.

Bonami, O., & Boyé, G. (2014). De formes en thèmes. In F. Villoing, S. Leroy, & S. David (Eds.), *Foisonnements morphologiques. Etudes en hommage à Françoise Kerleroux* (pp. 17–45). Presses Universitaires de Paris Ouest.

Bonami, O., & Luís, A. R. (2014). Sur la morphologie implicative dans la conjugaison du portugais: une étude quantitative. In J. L. Léonard (Ed.), *Morphologie flexionnelle et dialectologie romane. Typologie(s) et modélisation(s), no. 22 in Mémoires de la Société de Linguistique de Paris* (pp. 111–151). Leuven: Peeters.

Bonami, O., & Stump, G. T. (2016). Paradigm function morphology. In A. Hippisley & G. T. Stump (Eds.), *Cambridge handbook of morphology* (pp. 449–481). Cambridge: Cambridge University Press.

Bonami, O., Boyé, G., & Kerleroux, F. (2009). L'allomorphie radicale et la relation flexion-construction. In B. Fradin, F. Kerleroux, & M. Plénat (Eds.), *Aperçus de morphologie du français* (pp. 103–125). Saint-Denis: Presses de l'Université de Vincennes.

Bonami, O., Boyé, G., & Tseng, J. (2014a). An integrated analysis of French liaison. In P. Monachesi, G. Jäger, G. Penn, & S. Wintner (Eds.), *Proceedings of the 9th conference on formal grammar* (pp. 29–45). Stanford: CSLI Publications.

Bonami, O., Caron, G., & Plancq, C. (2014b). Construction d'un lexique flexionnel phonétisé libre du français. In F. Neveu, P. Blumenthal, L. Hriba, A. Gerstenberg, J. Meinschaefer, & S. Prévost (Eds.), *Actes du quatrième congrès mondial de linguistique française* (pp. 2583–2596).

Booij, G. (1997). Autonomous morphology and paradigmatic relations. In G. Booij & J. van Marle (Eds.), *Yearbook of morphology 1996* (pp. 35–53). Dordrecht: Kluwer.

Booij, G. (2010). *Construction morphology*. Oxford: Oxford University Press.

Boyé, G. (2006). Suppletion. In K. Brown (Ed.), *The encyclopedia of language and linguistics* (2nd ed., pp. 297–299). Oxford: Elsevier.

Boyé, G., & Schalchli, G. (2016). The status of paradigms. In A. Hippisley & G. Stump (Eds.), *The Cambridge handbook of morphology* (pp. 206–234). Cambridge: Cambridge University Press.

Carstairs, A. (1987). *Allomorphy in inflection*. London: Croom Helm.

Corbett, G. G. (2007). Canonical typology, suppletion and possible words. *Language*, *83*, 8–42.

Cotterell, R., Vylomova, E., Khayrallah, H., Kirov, C., & Yarowsky, D. (2017). Paradigm completion for derivational morphology. arXiv:170809151.

Dressler, W. U. (1985). Suppletion in word formation. In J. Fisiak (Ed.), *Historical semantics, historical word formation* (pp. 97–112). Berlin: de Gruyter.

Fradin B. (forthcoming). Competition in derivation: what can we learn from duplicates? In: F. Gardani, H. C. Luschützky, & F. Rainer (Eds.), Competition in Morphology, Springer, Berlin, u. L'Acquila

Grossman, M., Thornton, A. M. (forthcoming). Overabundance in Hungarian accusative pronouns. In B. Baldi, L. Franco, M. Grimaldi, & R. Lai (Eds.), Structuring variation in Romance linguistics and beyond. Studies in honour of Leonardo M. Savoia, Amsterdam: Benjamins.

Haspelmath, M., & Sims, A. (2010). *Understanding morphology*. London: Routledge.

Hathout, N., & Namer, F. (2014). Démonette, a French derivational morpho-semantic network. *Linguistic Issues in Language Technology*, *11*(5), 125–168.

Hathout, N., Sajous, F., & Calderone, B. (2014). GLÀFF, a large versatile French lexicon. In *Proceedings of LREC*, 2014.

Hjelmslev, L. (1938). Essai d'une théorie des morphèmes. In *Actes du quatrième congrès international de linguistes* (pp. 140–151). Copenhague: Munksgaard.

Jackendoff, R. (1975). Morphological and semantic regularities in the lexicon. *Language*, *51*, 639–671.

Jakobson, R. (1939). Signe zéro. In *Mélanges de linguistique offerts à Charles Bally*, Geneva: Georg et Cie. reprinted in Jakobson (1971).

Jakobson, R. (1971). *Selected writings II*. The Hague: Mouton.

de Jong N. H., Schreuder, R., & Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and cognitive processes*, *15*(4–5), 329–365.

Kilbury, J. (1992). Paradigm-based derivational morphology. In *Konvens* (Vol. 92, pp. 159–168). Berlin: Springer.

Lignon, S., & Namer, F. (2010). Comment conversionner les v-ion ? ou la construction de v-ionnerverbe par conversion. In *Actes du 2eme congrès mondial de linguistique française* (pp. 1009–1028).

Mansfield, J. (2016). Intersecting formatives and inflectional predictability: how do speakers and learners predict the correct form of Murrinhpatha verbs? *Word Structure*, *9*(2), 182–214.

van Marle, J. (1984). *On the paradigmatic dimension of morphological creativity*. Dordrecht: Foris.

Matthews, P. (2001). *A short history of structural linguistics*. Cambridge: Cambridge University Press.

Matthews, P. H. (1972). *Inflectional morphology. A theoretical study based on aspects of Latin verb conjugation*. Cambridge: Cambridge University Press.

Matthews, P. H. (1974). *Morphology*. Cambridge: Cambridge University Press.

Mel'čuk, I. A. (1976). On suppletion. *Linguistics*, *170*, 45–90.

Mel'čuk, I. A. (1994). *Suppletion. Studies in Language*, *18*, 339–410.

Pirrelli, V., & Federici, S. (1994). Derivational paradigms in morphonology. In *Proceedings of the 15th conference on computational linguistics* (pp. 234–240).

Pirrelli, V., & Yvon, F. (1999). The hidden dimension: a paradigmatic view of data-driven NLP. *Journal of Experimental & Theoretical Artificial Intelligence*, *11*(3), 391–408.

Plag, I. (1999). *Morphological productivity*. Berlin: de Gruyter.

Pounder, A. (2000). *Process and paradigms in word-formation morphology* (Vol. *131*). Berlin: de Gruyter.

del Prado Martín F. M., Deutsch, A., Frost, R., Schreuder, R., Jong, N. H. D., & Baayen, R. H. (2005). Changing places: a cross-language perspective on frequency and family size in Dutch and Hebrew. *Journal of Memory and Language*, *53*(4), 496–512.

Robins, R. H. (1959). In defense of WP. *Transactions of the Philological Society*, 116–144.

Saussure, F. (1916). *Cours de linguistique générale*. Paris: Payot.

Schreuder, R., & Baayen, R. H. (1997). How complex simple words can be. *Journal of Memory and Language*, *37*, 118–139.

Sims, A. (2015). *Inflectional defectiveness*. Cambridge: Cambridge University Press.

Strnadová, J. (2014). *Les réseaux adjectivaux: Sur la grammaire des adjectifs dénominaux en français*. PhD thesis, Université Paris Diderot et Univerzita Karlova V Praze.

Stump, G. T. (1998). Inflection. In A. Spencer & A. Zwicky (Eds.), *The handbook of morphology* (pp. 13–43). London: Blackwell.

Stump, G. T. (2001). *Inflectional morphology. A theory of paradigm structure*. Cambridge: Cambridge University Press.

Stump, G. T. (2005). Referrals and morphomes in Sora verb inflection. In G. Booij & J. van Marle (Eds.), *Yearbook of morphology 2005* (pp. 227–251). Dordrecht: Springer.

Stump, G. T. (2006). Heteroclisis and paradigm linkage. *Language*, *82*, 279–322.

Stump, G. T. (2016). *Inflectional paradigms*. Cambridge: Cambridge University Press.

Stump, G. T., & Finkel, R. (2013). *Morphological typology: from word to paradigm*. Cambridge: Cambridge University Press.

Thornton, A. M. (2011). Overabundance (multiple forms realizing the same cell): a non-canonical phenomenon in Italian verb morphology. In M. Maiden, J. C. Smith, M. Goldbach, & M. O. Hinzelin (Eds.), *Morphological autonomy: perspectives from romance inflectional morphology* (pp. 358–381). Oxford: Oxford University Press.

Thornton, A. M. (2012). Reduction and maintenance of overabundance. a case study on Italian verb paradigms. *Word Structure*, *5*, 183–207.

Thornton, A. M. (forthcoming) Overabundance: a canonical typology. In F. Rainer, F. Gardani, H. C. Luschützky, & W. U. Dressler (Eds.), *Competition in morphology*, Dordrecht: Springer.

Tribout, D. (2010). How many conversions from verb to noun are there in French? In *Proceedings of the HPSG 2010 conference* (pp. 341–357). Stanford: CSLI Publications.

Štekauer, P. (2014). Derivational paradigms. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of derivational morphology* (pp. 354–369). Oxford: Oxford University Press.

Wunderlich, D., & Fabri, R. (1995). Minimalist morphology: an approach to inflection. *Zeitschrift für Sprachwissenschaft, 14*(2), 236–294.

Wurzel, W. U. (1984). *Flexionsmorphologie und Natürlichkeit. Ein Beitrag zur morphologischen Theoriebildung*. Berlin: Akademie-Verlag, translated as Wurzel (1989).

Wurzel, W. U. (1989). *Inflectional morphology and naturalness*. Dordrecht: Kluwer.