Géraldine Walther
University Paris Diderot[*]

UDK 81'366

MEASURING MORPHOLOGICAL CANONICITY

1. INTRODUCTION

The question of regularity within morphological paradigms has been formerly addressed within approaches falling in the scope of *Canonical Typology*.

This work follows in the tradition of those approaches. It adopts a *strictly lexicalist approach* to morphology (Karttunen 1989; Bresnan 1982), and more specifically a *lexeme-based* approach in the sense of (Matthews 1974). As defined in (Fradin 2003) and references therein, the lexeme is considered to be an abstract entity, defined by its morphophonological properties, its meaning and its morphosyntactic category.[1] Concrete *forms* are built by inflectional morphology. We in particular follow the *Word and Paradigm* approach, e.g. Matthews, Aronoff and Stump (Matthews 1972; Zwicky 1985; Anderson 1992; Aronoff 1994; Stump 2001) and adopt a representation of forms based on the notions of *stems* and *exponents* as used by (Robins 1959) and (Matthews 1974): stems are what remains of a form once all exponents have been removed from it. In our approach, (concrete) forms are built from (abstract) lexemes through *(form) realisation rules* which are applied to them. Such rules are defined in inferential-realisational models such as *Paradigm Function Morphology (PFM)* (Stump 2001) or *Network Morphology* (Corbett/Fraser 1993).

The aim of this paper is to provide a means for assessing the notion of morphological canonicity through original measures developed within our new morphological framework PARSLI. In particular, we introduce original measures for non-canonical phenomena such as *heteroclisis*, *deponency*, *defectiveness* and *overabudance*.

2 CANONICAL INFLECTION

The concept of *canonical typology* can be traced back to Corbett as in (Corbett 2003). It represents an attempt to better understand what exactly differs from a hypothetic ideal *canonical* stage in the different occurrences of non-canonical phenomena. Note that in this approach, *canonical inflection* must not be mistaken for *prototypical inflection*. Canonical inflection is rare. It corresponds to an ideal state, seldom, if ever, met, but that constitutes a purely theoretical space from which deviant phenomena can be formally distinguished (Corbett 2007).

[*] *Author's address:* Laboratoire de Linguistique Formelle, Université Paris Diderot, 175 rue du Chevaleret, 75013 Paris, France. Email: geraldine.walther@linguist.jussieu.fr

[1] In particular, we also consider morphosyntactic information such as *argument structure* to be specified within the lexicon.

Canonical inflection is supposed to represent complete regularity, as well as an ideal correspondence between form and function such that different forms can most efficiently be distinguished from each other. In particular, it is a notion that affects both the relation between the cells of a given lexeme's paradigm and the corresponding cells belonging to two different lexemes' paradigms. Canonical inflection is thus defined through the comparison of both the cells of one given lexeme and the lexemes themselves.

Each cell in a given paradigm canonically shares the lexeme's stem but varies in terms of exponence depending on the morphosyntactic feature structure[2] expressed by the given form. On the other hand, across lexemes, the stems will canonically vary from one paradigm to the other while the exponents will remain the same according to their expressed morphosyntactic feature structures. These relations are illustrated through Tables 1 and 2.

In this work, we consider an inflectional paradigm canonical if it satisfies the criteria given in Table 3 (Corbett 2007). To these criteria we add the ones in Table 4 that further define canonical paradigm shape.[3] As stated for example in (Corbett 2007), deviation from these criteria leads to non-canonical paradigmatic properties.

A paradigm is considered canonical if it matches above mentioned criteria. The more it deviates from these criteria, the less canonical the paradigm. However, existing work on canonicity does not provide quantitative means to assess the degree of canonicity of a lexeme's paradigm. Such quantitative measures are the new feature we propose in section 4.

We present measures for four types of non-canonical inflection phenomena, namely *deponency*, *heteroclisis*, *defectiveness* and *overabundance*. These measures are computed within the inferential realisational model for inflectional morphology PARSLI.

In the following section, we first provide a short description of the major features of this model. A complete formal description can be found in Appendix B. Section 4 then goes on with presenting the measures of canonicity developed within PARSLI that allow for quantitatively assessing the canonicity of a given paradigm.

FEATURES	LEXEME 1		LEXEME 2	
1st p. Sing	stem1	*-ma*	stem2	*-ma*
2nd p. Sing	stem1	*-sa*	stem2	*-sa*
3rd p. Sing	stem1	*-ta*	stem2	*-ta*
1st p. Pl	stem1	*-mo*	stem2	*-mo*
2nd p. Pl	stem1	*-so*	stem2	*-so*
3rd p. Pl	stem1	*-to*	stem2	*-to*

▓ same

░ different

Table 1: Comparison over the cells of a given lexeme.

[2] Or what (Stump 2001) would refer to as morphosyntactic *property sets*.

[3] However, among the additional criteria, criterion 1 derives directly from criterion 2 in (Corbett 2007) and criterion 3 can be seen as derived from criterion 3 in (Corbett 2007).

FEATURES	LEXEME 1		LEXEME 2	
1st p. Sing	stem1	*-ma*	stem2	*-ma*
2nd p. Sing	stem1	*-sa*	stem2	*-sa*
3rd p. Sing	stem1	*-ta*	stem2	*-ta*
1st p. Pl	stem1	*-mo*	stem2	*-mo*
2nd p. Pl	stem1	*-so*	stem2	*-so*
3rd p. Pl	stem1	*-to*	stem2	*-to*

▓	same
░	different

Table 2: Comparison across lexemes.

	COMPARISON ACROSS CELLS OF A LEXEME	COMPARISON ACROSS LEXEMES
1 COMPOSITION/STRUCTURE	*same*	*same*
2 LEXICAL MATERIAL (≈ shape of stem)	*same*	*different*
3 INFLECTIONAL MATERIAL (≈ shape of inflection)	*different*	*same*
4 OUTCOME (≈ shape of inflected word)	*different*	*different*

Table 3: Criteria for Canonical Inflection according to (Corbett 2007).

	CANONICAL INFLECTION
1 STEMS AND FEATURES	*There is no "mismatch between form and function" (Baerman 2007).* *Each lexeme has exactly one stem that combines with a series of exponents.*
2 COMPLETENESS	*There exists exactly one form corresponding to the expression of a specific morphosyntactic feature structure.*
3 INFLECTION CLASS	*All forms of a lexeme are built from one single inflection class.*

Table 4: Additional criteria for Canonical Inflection.

3 INTRODUCING PARSLI

The name PARSLI stands for "PARadigm Shape and Lexicon Interface". PARSLI is a formal model designed for representing morphological information stored within the (morphological) lexicon on the one hand and (morphological) grammar on the other and giving a description of each lexeme of a given language with regard to its own paradigm structure. It is the paradigm structure that accounts for the various non-canonical inflectional phenomena mentioned above.

3.1 Defining the relevant notions

In PARSLI a lexeme is considered from the point of view of its formal participation in the inflectional process. Thus, we do not consider any specific semantics or possible derivational properties. In other words, we are here interested in the behaviour of what Fradin and Kerleroux refer to as *inflectemes* (Fradin/Kerleroux 2003), as opposed to lexemes, and for which a (very) simplified definition could be "a lexeme minus its semantic and argument-structural information."

At this stage of its development, PARSLI does not make any claims about how exactly the realisation of the forms should be modelled. It solely focuses on the distribution of the morphological information between the morphological lexicon and the morphological grammar. The realisation of the forms by the realisation rules contained within the morphological grammar can be represented by any suitable independent inferential-realisational formalism.

3.2 Describing the PARSLI model of inflectional morphology

PARSLI represents an inflecteme $\mathfrak{I}$[4] through seven defining elements:

1. the set of morphosyntactic feature structures $\mathfrak{I}$ can express,
2. the lexeme's morphosyntactic category,
3. an *inflection pattern*,
4. a *stem pattern*,
5. a *transfer rule* for stem selection,
6. a *transfer rule* for form realisation,
7. a pattern representing the paradigm

PARSLI relies on the concept of *inflection class*. Note that the definition of an inflection class in PARSLI is not the traditional one, that is a particular paradigm type. In PARSLI, an inflection class is defined as a function associating morphosyntactic feature structures with corresponding realisation rules, i.e., a way to apply specific exponents corresponding to a given morphosyntactic feature structure. Each inflection class is partitioned into one or more *inflection zones* which are the core of PARSLI's representation of inflection. As shown below, it is the selection of these inflection zones that determines an inflecteme's paradigm shape.

Inflection classes are the default associations of inflection zones that allow for computing the default paradigm structures of a language.[5] Using inflection zones from different inflection classes results in heteroclisis, as shown in Section 4 below.

Similarly, PARSLI also uses the concept of *stem class*, i.e., a function associating morphosyntactic feature structures with corresponding stem formation rules. Each stem class is partitioned into *stem zones*.

As will be shown below, these elements allow for the realisation of one of a given lexeme's form corresponding to a given morphosyntactic feature structure. In order

[4] I.e. the morphological part of a lexical entry.

[5] Usually this corresponds to the most frequent combination for a language's lexical items.

to illustrate the different steps of form realisation, we here outline the derivation of forms for the Italian adjectival inflecteme CARO *dear*. However, the complete process will be clearest after reading the formal definitions and illustrations in Appendixes B and C.

	MASC	FEM
SG	*karo*	*kara*
PL	*kari*	*kare*

Table 5: The paradigm of CARO dear in Italian.

The Italian inflecteme CARO can express four distinct morphosyntactic feature structures:

{GENDER *masc*, NUMBER *sg*}
{GENDER *masc*, NUMBER *pl*}
{GENDER *fem*, NUMBER *sg*}
{GENDER *fem*, NUMBER *pl*}

A traditional representation of that paradigm would be as in Table 5.

Stem formation

1. Each inflecteme being associated with a specific stem pattern, this stem pattern selects one particular stem zone corresponding to the morphosyntactic feature structure that is to be expressed. This stem zone is further used to obtain the stem formation rule.

   - *For CARO there is only one unique zone, associated with all four possible morphosyntactic feature structures.*

2. The (stem formation) transfer rule associated with the inflecteme computes the morphosyntactic feature structure that should be given as an input to the computed stem formation rule, given the morphosyntactic feature structure that is meant to be expressed for this given inflecteme;

   - *In the case of the inflecteme CARO, the stem formation transfer rule is the identity function, i.e., its output equals its input.*

3. The transformed feature structure is then associated with a specific stem formation rule through the inflection zone computed at step 1 above;

   - *This stem formation rule is the same for all morphosyntactic feature structures applicable to the inflecteme CARO. It will always compute the same stem regardless the morphosyntactic feature given as an input.*

4. The stem formation rule computes the correct stem form for the inflecteme (this rule may be expressed formally with any suitable realisation based formalism).

- *The inflecteme* CARO *has only one possible stem [kar].*

Inflection

1. In parallel, the inflection pattern associated with the inflecteme selects a specific inflection zone for the form realisation corresponding to a specific morphosyntactic feature structure;

- *Given an input feature* {GENDER *fem*, NUMBER *sg*} *this zone will also be the inflection zone associated with the plural forms of the feminine. Whether or not this zone is the same for the other forms of the paradigm depends on the general structure of the language.*[6]

2. The (form realisation) transfer rule associated with the inflecteme computes the morphosyntactic feature structure that should be given as an input to the computed realisation rule, given the morphosyntactic feature structure that is meant to be expressed for this given inflecteme;

- *In the case of the inflecteme* CARO, *the form realisation transfer rule is the identity function, i.e., its output equals its input, just as for the stem formation transfer rule.*

3. The transformed feature structure is then associated with a specific realisation rule through the inflection zone computed at step 1 above;

- *In the case of the inflecteme* CARO *for the input feature structure* {GENDER *fem*, NUMBER *sg*}, *this form realisation rule specifies the adding of the feminine singular exponent [a] to the stem.*

Form generation

1. Finally, the realisation rule obtained in *Inflection 3* is applied to the stem computed in *Stem formation 4* and the transformed morphosyntactic feature structure obtained in *Inflection 2*. It computes the correct form for a given input feature structure of the inflecteme. This realisation rule may be expressed formally with any suitable realisation based formalism.

- *In the case of the inflecteme* CARO *for the input feature* {GENDER fem, NUMBER sg}, *the realised form is thus [kara].*

Note that transfer rules most often default to the identity function. Whenever they differ from the identity function, they express “a mismatch between form and function“ as (Baerman 2007) puts it. They are used for modelling deponency.

[6] For a more detailed representation thereof, see the representation of heteroclisis in Section 4.

Inflection zones used by a given inflecteme $\mathfrak{I}$, i.e., the set of zones associated with it by its inflection rule, are called its *inflection pattern*. They build the inflecteme's paradigm. The set of $\mathfrak{I}$'s stem zones is called its *stem pattern*.

Inflection classes are defined as *the most natural combination of inflection zones*, i.e., those that are used together by a majority of inflectemes. They are default inflection patterns.

Sometimes a given morphosyntactic feature structure can be associated with more than one stem zone by a stem pattern and more than one inflection rule by the inflection pattern. However, in such a case, nothing enforces that each stem zone can be equally combined with each inflection zone. The situation is even worse when transfer rules differ from the identity function.

Therefore, we need a way to express the possible combinations of stem zones and inflection zones: the combinations are what we call *subpatterns*. These subpatterns are 4-tuples consisting of a stem zone, an inflection zone and two transfer rules. They express the possible combinations for a given inflecteme. A subpattern requires that the sets of morphosyntactic feature structures associated with the two zones have a non-empty intersection. The set of a given inflecteme's subpatterns is the inflecteme's *pattern*.

In the following section, we will show how the measures developed within PARSLI allow for measuring the canonicity of paradigms in terms of y *deponency*, *heteroclisis*, *defectiveness* and *overabundance*.[7]

4 EXPRESSING AND MEASURING NON-CANONICAL PARADIGM SHAPES WITH PARSLI

In this section we present non-canonical phenomena affecting paradigm structures and the associated measures.

4.1 Stem alternations, allomorphy and suppletion

Suppletion comes in two types: *stem suppletion* and *form suppletion* (Boyé 2006). Stem suppletion occurs whenever, inside a paradigm, the forms' exponents remain regular, but their stems vary. This is for example the case for the French verb ALLER *to go* which has four different stems, *all-*, *v-*, *i-* and *aill-*. Form suppletion corresponds to cases where a whole form is inserted in a paradigm cell that should canonically be filled by a certain stem and the exponent corresponding to this specific cell. Form suppletion is described in (Bonami/Boyé 2002) for the French verbe ÊTRE *to be* in the present indicative. For this verb, the 1^{st} person plural form *sommes*, for example, is unique in not using the regular 1^{st} person plural exponent *-ons* that canonically appears with corresponding forms of other verbs (see Table 6).

[7] For a list of the symbols used in the more formal definitions, please refer to Appendix A.

	SINGULAR	PLURAL
p1	***suis***	***sommes***
p2	*es*	*êtes*
p3	*est*	***sont***

Table 6: Form suppletion in the present indicative paradigm of French *être* 'to be'

4.1.1 Formal definition of allomorphy

Let $\mathfrak{I}$ =($\mathscr{K}_{\mathfrak{I}}$, $C_{\mathfrak{I}}$, $s_{\mathfrak{I}}$, $f_{\mathfrak{I}}$, $\mathscr{T}_{\psi_{\mathfrak{I}}}$, $\mathscr{T}_{\chi_{\mathfrak{I}}}$, $P_{\mathfrak{I}}$) be an inflecteme, its stem pattern $s_{\mathfrak{I}}$ associates (at least) one stem zone $\zeta_{s_{\mathfrak{I}},\varkappa}$ to a morphosyntactic feature structure $\varkappa \in \mathscr{K}_{\mathfrak{I}}$.[8]

A stem selection rule s allows for formally representing morphomic (in the sense of (Aronoff 1994)) structures in stem selection, such as can be observed for Latin verbs.

4.1.2 Example: Latin verb stems

In Latin, the distribution of the three existing stems available for all Latin verbs is morphomic in the sense that all verbs use the same stem pattern. This stem pattern is partitioned into three stem zones. Tables 7 and 8 give a schematic representation of the three stem zones.

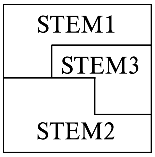

Table 7: Stem zones in the Latin active (sub-) paradigm

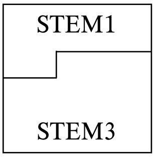

Table 8: Stem zones in the Latin passive (sub-) paradigm

STEM	ACT. SUBPARADIGM	PASS. SUBPARADIGM
STEM1	*imperf. finite*	*imperf. finite*
STEM2	*perf. finite*	
STEM3	*active future part.*	*passive past part.* *perf. finite (periphr.)*

Table 9: Morphomic combinations between morphosyntactic features and Latin verb stems

[8] Usually $\zeta_{s_{\mathfrak{I}},\varkappa}$ associates all compatible morphosyntactic feature structures with one unique stem formation rule.

4.2 Deponency

Some Croatian nouns use singular forms to express plural, as shown in the data presented in (Baerman 2006). This mismatch between form and function is what, following Baerman (Baerman 2007), we name *deponency*.

As shown in (Baerman 2006), Croatian nouns are inflected according to a number of different declension classes. Some classes that are relevant for our discussion are shown in Table 10. The data shows that the nouns *dete* 'child' and *tele* 'calf' inflect in the plural according to the singular pattern of respectively the A-STEM and I-STEM inflection classes. Using singular inflection to express the plural results in this mismatch between form and function.[9]

	(FEMININE) A-STEM		(FEMININE) I-STEM	
	žena 'woman'		*stvar* 'thing'	
	SINGULAR	PLURAL	SINGULAR	PLURAL
NOM	*žen-a*	*žen-e*	*stvar*	*stvar-i*
ACC	*žen-u*	*žen-e*	*stvar*	*stvar-i*
GEN	*žen-e*	*žen-a*	*stvar-i*	*stvar-i*
DAT	*žen-i*	*žen-ama*	*stvar-i*	*stvar-ima*
INS	*žen-om*	*žen-ama*	*stvar-i*	*stvar-im*

Table 10: Croatian noun inflection

	(FEMININE) A-STEM		(FEMININE) I-STEM	
	žena 'woman'		*stvar* 'thing'	
	SINGULAR	PLURAL	SINGULAR	PLURAL
NOM	*dete*	*deca*	*tele*	*telad*
ACC	*dete*	*decu*	*tele*	*telad*
GEN	*deteta*	*dece*	*teleta*	*telad*
DAT	*detetu*	*deci*	*teletu*	*teladi (ma)*
INS	*detetom*	*decom*	*teletom*	*teladi (ma)*

Table 11: Croatian deponent noun inflection

4.2.1 Formal definition of deponency

Let $\mathfrak{I}$=($\mathscr{K}_{\mathfrak{I}}$, $C_{\mathfrak{I}}$, $\mathrm{s}_{\mathfrak{I}}$, $\mathrm{f}_{\mathfrak{I}}$, $\mathscr{T}\psi_{\mathfrak{I}}$, $\mathscr{T}\chi_{\mathfrak{I}}$, $P_{\mathfrak{I}}$) be an inflecteme.

The “mismatch between form and function” stated by (Baerman, 2007) to be the definition of deponency occurs whenever the morphosyntactic features expressed by a given inflecteme's form $\mathfrak{f}$ do not match the morphosyntactic features $\varkappa$ usually expressed by the realisation rule φ used to build that form of this given inflecteme.

[9] For more data on deponency, the reader may refer to the large database put together by the Surrey Morphology Group: http://www.smg.surrey.ac.uk/deponency.

As mentioned above, within PARSLI this means that the transfer rule $\mathscr{T}_{\chi_{\mathfrak{I}}}$ differs from the identity function, or, in other words, the morphosyntactic feature structure $\mathscr{T}_{\chi_{\mathfrak{I}}}(\varkappa)$ expressed by $\mathfrak{f}$ differs from the morphosyntactic feature structure $\varkappa$ that has been associated by the appropriate inflection zone ξ through $\mathrm{f}_{\mathfrak{I}}$.

More precisely, a given inflecteme $\mathfrak{I}$ is said to be *deponent* iff there exists at least one form $\mathfrak{f}$ in its paradigm $\mathfrak{P}$ built in way such that

$$\varkappa \neq \mathscr{T}_{\chi_{\mathfrak{I}}}(\varkappa).$$

Let $\mathscr{P}_{\mathscr{K}_{\mathfrak{I}}} = \{\mathscr{K}_{\mathfrak{I},1}, \ldots, \mathscr{K}_{\mathfrak{I},n}\}$ be the smallest partition of $\mathscr{K}_{\mathfrak{I}}$ such that $\mathrm{f}_{\mathfrak{I}}$ associates each $\varkappa \in \mathscr{K}_{\mathfrak{I},i}$ with the same $\mathscr{K}_{\mathfrak{I},i}$.[10]

An inflecteme is considered to be *semi-deponent* iff for at least one element of $\mathscr{K}_{\mathfrak{I},i} \subset \mathscr{P}_{\mathscr{K}_{\mathfrak{I}}}$ but not all, the restriction $\mathscr{T}_{\chi_{\mathfrak{I}}|\mathscr{K}_{\mathfrak{I}i}}$ of $\mathscr{T}_{\chi_{\mathfrak{I}}}$ to $\mathscr{K}_{\mathfrak{I}i}$ is the identity function.

4.2.2 Deponency Index

For the non-canonical phenomenon of deponency we can thus compute a measure of canonicity. We call this measure the *deponency index*. The deponency index $\mathfrak{D}_{\mathfrak{I}}$ of an inflecteme $\mathfrak{I}$ is defined as the number of elements of the form $\mathscr{K}_{\mathfrak{I},i}$ in $\mathscr{P}_{\mathscr{K}_{\mathfrak{I}}}$ such that $\mathscr{T}_{|\mathscr{K}_{\mathfrak{I},\mathrm{i}}} \neq \mathrm{id}$:

$$\mathfrak{D}_{\mathfrak{I}} = |\{\mathscr{K}_{\mathfrak{I},\mathrm{i}} \in \mathscr{P}_{\mathscr{K}_{\mathfrak{I}}} | \mathscr{T}_{|\mathscr{K}_{\mathfrak{I},\mathrm{i}}} \neq \mathrm{id}\}|$$

Hence, an inflecteme is deponent iff $\mathfrak{D}_{\mathfrak{I}} > 0$.
An inflecteme is semi-deponent iff $|\mathscr{P}_{\mathscr{K}_{\mathfrak{I}}}| > \mathfrak{D}_{\mathfrak{I}} > 0$.
Conversely, a non-deponent inflecteme veries $\mathfrak{D}_{\mathfrak{I}} = 0$.

4.2.3 Example: Croatian nouns

The Croatian data presented above can be modelled within PARSLI with a transfer rule. Hence, an inflecteme building its paradigm as described above entails a transfer rule $\mathscr{T}_{\chi_{\mathfrak{I}}}$ within its definition for which

$$\mathscr{T}_{\chi_{\mathfrak{I}}}(\{\textsc{number}\ \textit{plural}\})=\{\textsc{number}\ \textit{singular}\}.$$

Hence it is a semi-deponent noun: the transfer rule differs from the identity function for the morphosyntactic feature structures containing the attribute-value pair {NUMBER *plural*}. For those containing {NUMBER *singular*}, $\mathscr{T}_{\chi_{\mathfrak{I}}} = \mathrm{id}$.

In other words, for both these two nouns, the smallest partition of $\mathscr{K}_{\mathfrak{I}}$ such that $\mathrm{f}_{\mathfrak{I}}$ associates each $\varkappa \in \mathscr{K}_{\mathfrak{I},i}$ with the same $\mathscr{K}_{\mathfrak{I},i}$ consists of two subsets of $\mathscr{K}_{\mathfrak{I}}$, one for singular feature structures (for which no deponency occurs) and one for plural ones.

[10] In the case of overabundant inflectemes, this does not concern a unique zone but the same set of zones.

Thus, the deponency index for these nouns is equal to 1. Since the total number of elements of that partition is 2, these lexemes are semi-deponents.

	MASCULINE ANIMATE		MASCULINE INANIMATE		MASCULINE HETEROCLITE	
	CHLAP 'boy'		*DUB 'oak'*		*OROL 'eagle'*	
	SINGULAR	PLURAL	SINGULAR	PLURAL	SINGULAR	PLURAL
NOM	*chlap*	*chlap-i*	*dub*	*dub-**y***	*orol*	*orl-**y***
GEN	*chlap-a*	*chlap-ov*	*dub-a*	*dub-ov*	*orl-a*	*orl-ov*
DAT	*chlap-**ovi***	*chlap-om*	*dub-u*	*dub-om*	*orl-**ovi***	*orl-om*
ACC	*chlap-a*	*chlap-ov*	*dub*	*dub-**y***	*orl-a*	*orl-**y***
LOC	*chlap-**ovi***	*chlap-och*	*dub-e*	*dub-och*	*orl-**ovi***	*orl-och*
INS	*chlap-om*	*chlap-mi*	*dub-om*	*dub-mi*	*orl-om*	*orl-ami*

Table 12: Heteroclisis in Slovak masculine animal names inflection

4.3 Heteroclisis

Heteroclisis refers to the phenomenon where a lexeme's paradigm is built out of (at least) two, otherwise seperate, inflection classes.

Examples of heteroclisis are (some) Slovak animal nouns. Indeed, in Slovak, most masculine animal nouns are inflected as masculine animate nouns in the singular, whereas they may (and for some lexemes, must) inflect as masculine inanimate nouns in the plural (except in specific cases, such as personification, which triggers the animate inflection even for plural forms) (Zauner 1973). Compare for example the inflection of *CHLAP 'boy'*, *DUB 'oak'* and *OROL 'eagle'* in Table 12.[11]

4.3.1 Definition

If all inflection zones associated with a given inflecteme $\mathfrak{I}$ belong to the same inflection class $\mathcal{F}$, the inflecteme is canonical in the dimension of inflection class constitution. Conversely, if its inflection zones belong to at least two distinct inflection classes, the inflecteme is said to be *heteroclite*.

4.3.2 Heteroclicity Index

We define an inflecteme's *heteroclicity index* $\mathcal{H}_{\mathfrak{I}}$ as the number of zones used to build an inflecteme's paradigm that are partitions of distinct inflection classes. In other words, this represents the number of inflection classes involved in the building of that inflecteme's paradigm.

More precisely, $\mathcal{H}_{\mathfrak{I}}$ is defined in the following way:

$$\mathcal{H}_{\mathfrak{I}} = |\{\tilde{\xi} \,|\, \xi \in X_{\mathfrak{I}}\}| - 1$$

[11] Both *CHLAP* and *DUB* have a regular inflection: *CHLAP* belongs to the standard inflection class for masculine animate stems ending with a consonant, whereas *DUB* belongs to the standard inflection class for masculine inanimate stems ending with what is called a hard or neutral consonant in the Slavic linguistic tradition.

where $\tilde{\xi}$ stands for the inflection class to whose partition ξ belongs, and $\mathsf{X}_{\mathfrak{J}}$ for the set of zones associated with $\mathfrak{J}$ through its inflection pattern.

Thus, $\mathfrak{J}$ is heteroclite iff $\mathcal{H}_{\mathfrak{J}} > 0$.

4.3.3 Example 1: Slovak animal nouns

For the Slovak animal nouns described in Table 12, the inflection zone used for building the singular forms of the noun *OROL* 'eagle' is an element of the partition of the inflection class associated with animate nouns such as *CHLAP* 'boy', while the inflection zone used for the plural forms of such animal nouns belongs to the partition of the inflection class associated with inanimates like *DUB* 'oak'.

The heteroclicity index of such nouns is

$$\mathcal{H}_{OROL} = |\{\tilde{\xi}\,|\,\xi \in \mathsf{X}_{OROL}\}| - 1 = |\tilde{\xi}_{F_{animate},\ \mathrm{pl}}| + |\tilde{\xi}_{F_{animate},\ \mathrm{sg}}| - 1 = 1 + 1\ - 1 = 1$$

4.3.4 Example 2: Croatian nouns

Similarily, the Croatian nouns from Table 11 show inflection patterns producing the inflection zones listed in Table 13. These tables show that the corresponding Croatian nouns are not only deponent, but also heteroclite.

Thus, several-canonical phenomena may sometimes occur simultaneously in non-canonical paradigms.

INFLECTION CLASS	A: NEUTER	B: (FEMININE)	C: (FEMININE)
	STEM IN -ET	STEM IN -A	STEM IN -I
DETE 'child'	SG: $\xi_{\mathrm{A,sg}}$	PL: $\xi_{\mathrm{B,sg}}$	
TELE 'veal'	SG: $\xi_{\mathrm{A,sg}}$		PL: $\xi_{\mathrm{C,sg}}$

Table 13: Nominal inflection of Croatian heteroclite nouns

4.4 DEFECTIVENESS

Defectiveness (Baerman et al. 2010) refers to lexemes which display empty (missing) cells in their paradigm. Sometimes languages contain lexemes for which expected forms are simply unexisting; native speakers would always try avoiding having to build the corresponding forms. This is for example what we can observe with some French verbs such as PAÎTRE *to graze* for which there are no past tense forms available apart from the imperfect. Another example are the *pluralia tantum* described below.

4.4.1 Formal definition

A paradigm is considered *defective* iff there is at least one morphosyntactic feature structure $\varkappa$ belonging to the set $\mathscr{K}_{C_{\mathfrak{J}}}$ of the morphosyntactic feature structures of the category of an inflecteme $\mathfrak{J}$ which $\mathrm{f}_{\mathfrak{J}}$ does not associate with any inflectional zone ξ.

One can also say that an inflecteme $\mathfrak{J}$ is defective iff the set $\mathscr{K}_{\mathfrak{J}}$ of its morphosyntactic feature structures does not cover the set $\mathscr{K}_{C_{\mathfrak{J}}}$ of the morphosyntactic feature structures of its category.

$$\mathscr{K}_{\mathfrak{J}} \subsetneq \mathscr{K}_{C_{\mathfrak{J}}}$$

Example: Pluralia Tantum Another example are the nouns called *pluralia tantum* which only exist in the plural, cf. English TROUSERS, French VIVRES *food supplies* or Slovak VIANOCE *Christmas*.

Let us take the example of the French *pluralium tantum* $\mathfrak{I}$ VIVRES food supplies. If we only consider the number features, we get the following defined morphosyntactic feature structures:

$$\mathscr{K}_{\mathfrak{I}} = \{\text{NUMBER } \textit{plural}\}$$

while

$$\mathscr{K}_{C_{\mathfrak{I}}} = \mathscr{K}_{\text{nom}} = \{\text{NUMBER } \textit{singular}, \text{ NUMBER } \textit{plural}\}$$

and

$$\mathscr{K}_{\mathfrak{I}} \subsetneq \mathscr{K}_{C_{\mathfrak{I}}}$$

4.5 Overabundance

The obvious counterpart to defectiveness is the concept of *overabundance*. Overabundance occurs when cells of a paradigm contain more than one form. The notion has been introduced by Thornton and is discussed in (Thornton 2010) for Italian. Canonical overabundance characterises the case where *cell mates* of one given cell compete, without any morphological feature[12] permitting to choose one over the other. Table 14 shows examples thereof for Italian verbs.

	CELL-MATE 1	CELL-MATE 2
'languish' 3PL.PRS.SUBJ	*languano*	*languiscano*
'possess' 3PL.PRS.SUBJ	*possiedano*	*posseggano*
'possess' 3SG.PRS.SUBJ	*possieda*	*possegga*
'possess' 1SG.PRS.SUBJ	*possiedo*	*posseggo*

Table 14: Overabundance in Italian (Thornton, 2010)

In French, an example is given by the verb *ASSEOIR* 'to sit' that has two different forms in most cells as shown in Table 15.[13] All French verbs in *-ayer* also exhibit systematic overabundance (see Table 16). Indeed, for some cells, these verbs may use two competing stems (in *-ay-* and in *-ai-*) and therefore have two different inflected forms, which are morphologically equivalent (although semantic, pragmatic, sociolinguistic and other constraints may interfere).

[12] Or any other type of feature.

[13] See for example (Bonami/Boyé 2010) for a longer discussion thereof.

IND.PRES	SINGULAR	PLURAL
P1	*assois* *assieds*	*assoyons* *asseyons*
P2	*assois* *assieds*	*assoyez* *asseyez*
P3	*assoit* *assied*	*assoient* *asseyent*

Table 15: Overabundance in French *asseoir* 'to sit'

IND.PRES	SINGULAR	PLURAL
P1	*balaye* *balaie*	*balayons*
P2	*balayes* *balaies*	*balayez*
P3	*balaye* *balaie*	*balayent* *balaient*

Table 16: Overabundance in French *balayer* 'to sweep'

4.5.1 Formal definition

A paradigm is considered *overabundant* iff there is at least one morphosyntactic feature structure $\varkappa$ belonging to the set $\mathscr{K}_{C_{\mathfrak{I}}}$ of the morphosyntactic feature structures of the category of an inflecteme $\mathfrak{I}$ which $\mathrm{f}_{\mathfrak{I}}$ associates with more than one inflection zone. In that case, $\mathrm{f}_{\mathfrak{I}}$ is a generic binary relation and not a function.

$$\mathrm{f}_{\mathfrak{I}}(\varkappa) = S, \textit{ where } |S| > 1.$$

Example: Italian overabundant verbs Table 14 shows examples of overabundant Italian verbs.[14]

In this case, the inflecteme LANGUIRE has a inflection pattern $\mathrm{f}_{\text{LANGUIRE}}$ which associates the morphosyntactic feature structure $\varkappa$ ={3PL.PRS.SUBJ} with two inflection zones, each producing a different realisation rule φ_1 and φ_2. These two rules thus give rise to two distinct forms within the paradigm $\mathfrak{P}_{\text{LANGUIRE}}$ expressing $\varkappa$: $\mathfrak{f}_1$=*languano* and $\mathfrak{f}_2$=*languiscano*.

4.6 Canonical Inflection

From the definitions of non-canonical phenomena above, we can deduce the following definition of *Canonical Inflection*.

Canonical inflection corresponds to the case where the inflection pattern $\mathrm{f}_{\mathfrak{I}}$ of an inflecteme $\mathfrak{I}$ associates the morphosyntactic feature structures belonging to the set of morphosyntactic feature structures $\mathscr{K}_{\mathfrak{I}}$ for which $\mathfrak{I}$ is defined with inflection zones that constitute the complete set of elements contained within the partition of one unique inflection class $\mathcal{F}$.

In particular, this entails that for all morphosyntactic feature structures $\varkappa$, the inflection pattern $\mathrm{f}_{\mathfrak{I}}$ associates $\varkappa$ with one unique element of the partition of $\mathcal{F}$.

Moreover, the stem pattern associates every $\varkappa \in \mathscr{K}_{\mathfrak{I}}$ with a stem zone ζ belonging to a stem class Γ containing only this one stem zone ζ and that produces a unique stem formation rule σ, whatever $\varkappa$. In other words, $\mathfrak{I}$ has a unique stem.

[14] The data is borrowed from (Thornton 2010).

Finally, the transfer rule $\mathscr{T}_{\chi_{\mathfrak{I}}}$ is the identity function and the set of morphosyntactic feature structures $\mathscr{K}_{\mathfrak{I}}$ defined for $\mathfrak{I}$ equals the set of morphosyntactic feature structures $\mathscr{K}_{C_{\mathfrak{I}}}$ defined for $\mathfrak{I}$'s morphosyntactic category $C_{\mathfrak{I}} \in \mathscr{C}$.

The same holds for the transfer rule $\mathscr{T}_{\psi_{\mathfrak{I}}}$.

5 CONCLUSION

We have presented PARSLI, a formal model of inflectional morphology. PARSLI being completely formalised, it can be implemented. Such an implementation would allow for the comparison of complete morphological descriptions with regard to their complexity. Indeed, previous experiments on complexity evaluation with PARSLI and its implementation within the Alexina lexical framework (Sagot 2010) have already been conducted (Sagot/Walther 2011). The usefulness of PARSLI to build morphological descriptions with reduced descriptive complexity has also been shown in (Walther/Sagot 2011).

But most importantly, in the domain of *Canonical Typology*, PARSLI contains original measures that allow for quantitatively assessing the canonicity of paradigms in the sense of the qualitative caracterisation proposed by the approaches developed within *Canonical Typology* (Corbett 2003).

A SUMMARY OF THE NOTATIONS IN PARSLI

We use the folowing notations in the formal definitions:

- A morphosyntactic feature structure will be noted $\varkappa$,
- an inflection rule f,
- an inflection class $\digamma$,
- an inflection zone ξ,
- a stem selection rule s,
- a stem class Γ,
- a stem zone ζ,
- a transfer rule $\mathscr{T}$
    - a stem transfer rule $\mathscr{T}_{\psi_{\mathfrak{I}}}$,
    - an inflection transfer rule $\mathscr{T}_{\chi_{\mathfrak{I}}}$,
- and a pattern *P*.

B A FORMAL DEFINITIONS OF THE PARSLI MODEL

The next two appendixes provide the actual formalisation underlying the PARSLI model and a formalised representation of paradigm building within the model.

B.1 Phonological material

An elementary sequence of phonological material e is a segmental or suprasegmental combination of sounds. The set of all elementary sequences of phonological material is noted E.

B.2 Morphosyntactic features

B.2.1 Features and feature structures

In this document, we define a *morphosyntactic feature structure* $\varkappa$ as a set of attribute-value pairs. PARSLI makes no strong assumptions about how the feature structures are organised with regard to one another.

B.2.2 Morphosyntactic categories

The set of all feature structures used in a given complete morphological description of an inflecteme is noted $\mathscr{K}$.

An *inflecteme* $\mathfrak{I}$ will be assigned a category depending on the feature structure it has information about: an inflecteme $\mathfrak{I}$ from a *category* C will cover a subset $\mathscr{K}_{\mathfrak{I}}$ of the morphosyntactic feature structure set $\mathscr{K}_C \subset \mathscr{K}$ specific to that category.

The set of all categories is noted $\mathscr{C}$.

B.3 Stems

B.3.1 Definition

A *stem* r is an elementary sequence of phonological material. The set of all stems is noted R.

$$\mathrm{r} \in \mathrm{R} \subset \mathrm{E}$$

B.3.2 Stem formation rule

Stem formation is expressed through stem formation rules. A stem formation rule σ is a function from $\mathscr{K}$ to E which takes a specific morphosyntactic feature structure $\varkappa \in \mathscr{K}$ as an input so as to produce a phonological material e' expressing that feature.

$$\sigma : \mathscr{K} \rightarrow \mathrm{E}$$
$$\sigma\,(\varkappa) = \mathrm{e}'$$

The set of all stem formation rules is noted Σ.

B.3.3 Stem class

A stem class $\mathit{\Gamma}$ is a function from $\mathscr{K}_{\mathit{\Gamma}} \subset \mathscr{K}$ to Σ.

B.3.4 Stem zones

Let $\mathit{\Gamma}$ be a stem class defined over a set $\mathscr{K}_{\mathit{\Gamma}}$ of morphosyntactic feature structures. For each $\mathit{\Gamma}$ a unique partition of $\mathscr{K}_{\mathit{\Gamma}}$ is defined, whose members are noted $\mathscr{K}_{\mathit{\Gamma},k}$, such that:[15]

$$\mathscr{K}_{\mathit{\Gamma}} = \bigsqcup_{k} \mathscr{K}_{\mathit{\Gamma},k}$$

[15] "$\sqcup$" denotes the union of disjoint sets.

A stem zone ζ for Γ is then defined as a pair

$$\zeta = (\mathscr{K}_{\Gamma,k}, \Gamma)$$

where $\mathscr{K}_{\Gamma,k}$ is one element of the partition.
Let $\zeta = (\mathscr{K}_{\Gamma,k}, \Gamma)$ be a zone for Γ. We define the operators $\tilde{}$ and $\hat{}$ as follows: $\tilde{\zeta}$ is the second element of ζ, i.e., its stem class Γ, and $\hat{\zeta}$ is the first element of ζ, i.e., the corresponding element of the partition of $\mathscr{K}_{\Gamma}$.

The set of zones for a stem class Γ is noted $Z(\Gamma)$. The set of all stem zones for all stem classes is noted $\mathscr{Z}$.

$$\mathscr{Z} = \bigcup_{\Gamma} Z\,(\Gamma)$$

B.3.5 Stem pattern

A *stem pattern* is a binary relation s associating an element from a given $\mathscr{K}_s \subset \mathscr{K}$ with one or more stem zones. A given morphosyntactic feature structure $\varkappa \in \mathscr{K}_f$ will be associated through s with stem zones of the form $\zeta = (\mathscr{K}_{\Gamma,k}, \Gamma)$. From there we can retrieve the stem formation rule $\sigma \in \sum$ corresponding to a given $\varkappa \in \mathscr{K}_s$:

$$\text{If } (\varkappa, \zeta) \in s \wedge \zeta = (\mathscr{K}_{\Gamma,k}, \Gamma),$$

then, provided we are given a certain $\varkappa' \in \mathscr{K}_{\Gamma,k}$ (be it equal to $\varkappa$ or not), one of the corresponding stem formation rule σ verifies

$$\sigma = \tilde{\zeta}\,(\varkappa') = \Gamma\,(\varkappa')$$

B.4 Inflection

B.4.1 Realisation rule

Inflection is expressed through realisation rules. A realisation rule φ is a function from E x $\mathscr{K}$ to E which takes specific phonological material e[16] as an input so as to produce a modified phonological material e' in order to express a specific morphosyntactic feature structure $\varkappa \in \mathscr{K}$.

$$\varphi : E \times \mathscr{K} \rightarrow E$$
$$\varphi\,(e,\varkappa) = e'$$

The set of all realisation rules is noted Φ.

B.4.2 Inflection class

An inflection class $\digamma$ is a function from $\mathscr{K}_{\digamma} \subset \mathscr{K}$ to Φ.

[16] Namely the stem produced by the corresponding stem formation rule.

B.4.3 Inflection zones

Let F be an inflection class defined over a set $\mathscr{K}_{\mathsf{F}}$ of morphosyntactic feature structures.

For each F is defined a unique partition of $\mathscr{K}_{\mathsf{F}}$, whose members are noted $\mathscr{K}_{\mathsf{F},k}$, such that:

$$\mathscr{K}_{\mathsf{F}} = \bigsqcup_{k} \mathscr{K}_{\mathsf{F},k}$$

An inflection zone ξ for F is then defined as a pair

$$\xi = (\mathscr{K}_{\mathsf{F},k}, \mathsf{F})$$

where $\mathscr{K}_{\mathsf{F},k}$ is one element of the partition.

Let $\xi = (\mathscr{K}_{\mathsf{F},k}, \mathsf{F})$ be a zone for F. We define the operators ˜ and ˆ as follows: $\tilde{\xi}$ is the second element of ξ, i.e., its inflection class F, and $\hat{\xi}$ s the first element of ξ, i.e., the corresponding element of the partition of $\mathscr{K}_{\mathsf{F}}$.

The set of zones for an inflection class F is noted $\mathsf{X}(\mathsf{F})$. The set of all inflection zones for all inflection classes is noted $\mathscr{X}$.

$$\mathscr{X} = \bigcup_{\mathsf{F}} \mathsf{X}(\mathsf{F})$$

B.4.4 Inflection pattern

An *inflection pattern* is a binary relation f associating an element from a given $\mathscr{K}_{\mathrm{f}} \subset \mathscr{K}$ with one or more inflection zones. A given morphosyntactic feature structure $\varkappa \in \mathscr{K}_{\mathrm{f}}$ will be associated through f with inflection zones of the form $\zeta = (\mathscr{K}_{\mathsf{F},k}, \mathsf{F})$. From there we can retrieve the inflectional function $\varphi \in \Phi$ corresponding to a given $\varkappa \in \mathscr{K}_{\mathrm{f}}$:

$$\text{If } (\varkappa, \xi) \in \mathrm{f} \wedge \xi = (\mathscr{K}_{\mathsf{F},k}, \mathsf{F})$$

then, provided we are given a certain $\varkappa' \in \mathscr{K}_{\mathsf{F},k}$ (be it equal to $\varkappa$ or not), one of the corresponding inflectional function φ verifies

$$\varphi = \tilde{\xi}\,(\varkappa') = \mathsf{F}(\varkappa')$$

B.5 Transfer rules

We define a transfer rule $\mathscr{T}$ as a function from its domain $\mathscr{K}_{\mathscr{T}} \in \mathscr{K}$ to $\mathscr{K}$.

Given an inflecteme $\mathfrak{J}$, there are two types of transfer rules. One $\mathscr{T}_{\psi_{\mathfrak{J}}}$ for stem formation and one $\mathscr{T}_{\chi_{\mathfrak{J}}}$ for inflection.

B.6 Pattern

B.6.1 Subpattern

A subpattern is defined for a given inflecteme $\mathfrak{I}$. It is a 4-tuple consisting of a stem zone ζ, an inflection zone ξ and two transfer rules, $\mathscr{T}_{\psi_{\mathfrak{I}}}$ and $\mathscr{T}_{\chi_{\mathfrak{I}}}$. To be valid, a subpattern requires that the set of morphosyntactic feature structures $\hat{\zeta} \in \mathscr{K}$ and $\hat{\xi} \in \mathscr{K}$ associated respectively with ζ and ξ have a non-empty intersection.

$$\hat{\zeta} \cap \hat{\xi} \neq \emptyset$$

B.6.2 Pattern

A pattern *P* is the set of all valid subpatterns defined for a given inflecteme $\mathfrak{I}$.

B.7 Inflectemes

B.7.1 Definition

Formal definition of an inflecteme:

An inflecteme $\mathfrak{I}$ is a 7-tuple $(\mathscr{K}_{\mathfrak{I}}, \mathcal{C}_{\mathfrak{I}}, s_{\mathfrak{I}}, f_{\mathfrak{I}}, \mathscr{T}_{\psi_{\mathfrak{I}}}, \mathscr{T}_{\chi_{\mathfrak{I}}}, P_{\mathfrak{I}})$, where

- $\mathscr{K}_{\mathfrak{I}}$ is the set of morphosyntactic features $\varkappa$ expressable by $\mathfrak{I}$,

- $\mathcal{C}_{\mathfrak{I}}$ is $\mathfrak{I}$ morphosyntactic category, and $\mathcal{C}_{\mathfrak{I}} \in \mathscr{C}$, where $\mathscr{C}$ is the set of morphosyntactic categories that exist in a morphological description for a given language,

- $s_{\mathfrak{I}}$ is a *stem pattern*, a binary relation from $\mathscr{K}_{\mathfrak{I}}$ to $\mathcal{Z}_{s_{\mathfrak{I}}}$, the set of *stem zones* compatible with $\mathfrak{I}$; $\mathcal{Z}_{s_{\mathfrak{I}}} \subset \mathscr{Z}$, where $\mathscr{Z}$ is the set of all stem zones in a morphological description of a given language,

- $f_{\mathfrak{I}}$ is a *inflection pattern*, binary relation from $\mathscr{K}_{\mathfrak{I}}$ to $\chi_{f_{\mathfrak{I}}}$, the set of *inflection zones* according to which a given inflecteme is inflected; $\chi_{f_{\mathfrak{I}}} \subset \mathscr{X}$, where $\mathscr{X}$ is the set of all inflection zones in a morphological description of a given language,

- $\mathscr{T}_{\psi_{\mathfrak{I}}}$ is a *transfer rule*, i.e., a function defined over at least all morphosyntactic feature structures $\varkappa \in \mathscr{K}_{\mathfrak{I}}$, such that $\mathscr{T}_{\psi_{\mathfrak{I}}}(\varkappa)$ belongs to the set of morphosyntactic features realised through the stem zones defined for $\mathscr{K}_{\mathfrak{I}}$.

- $\mathscr{T}_{\chi_{\mathfrak{I}}}$ is a *transfer rule*, i.e., a function defined over at least all morphosyntactic feature structures $\varkappa \in \mathscr{K}_{\mathfrak{I}}$, such that $\mathscr{T}_{\chi_{\mathfrak{I}}}(\varkappa)$ belongs to the set of morphosyntactic features realised through the inflection zones defined for $\mathscr{K}_{\mathfrak{I}}$;

- $P_{\mathfrak{I}}$ is a pattern, i.e., a set of subpatterns defined as a 4-tuple of the form $(\zeta_{\varkappa}, \xi_{\varkappa}, \mathscr{T}_{\psi_{\mathfrak{I}}}, \mathscr{T}_{\chi_{\mathfrak{I}}})$, where $\zeta_{\varkappa}$ is a stem zone associated with a given morphosyntactic feature structure $\varkappa$ through $s_{\mathfrak{I}}$ and $\xi_{\varkappa}$ an inflection zone associated to $\varkappa$ through $f_{\mathfrak{I}}$.

B.8 Paradigms

Let $(\mathcal{K}_{\mathfrak{I}}, \mathcal{C}_{\mathfrak{I}}, s_{\mathfrak{I}}, f_{\mathfrak{I}}, \mathcal{T}_{\psi_{\mathfrak{I}}}, \mathcal{T}_{\chi_{\mathfrak{I}}}, P_{\mathfrak{I}})$ be an inflecteme.

B.8.1 Forms

A form $\mathfrak{f}$ is a combination of elementary sequences of phonological material. It expresses a set of morphosyntactic features $\varkappa$ for the inflecteme $\mathfrak{I}$ =$(\mathcal{K}_{\mathfrak{I}}, \mathcal{C}_{\mathfrak{I}}, s_{\mathfrak{I}}, f_{\mathfrak{I}}, \mathcal{T}_{\psi_{\mathfrak{I}}}, \mathcal{T}_{\chi_{\mathfrak{I}}}, P_{\mathfrak{I}})$ and is obtained from a stem r of $\mathfrak{I}$ by the realisation rule φ corresponding to one of the appropriate inflection zones ξ, obtained through the inflection pattern $f_{\mathfrak{I}}$. φ is then equal to $\tilde{\xi}$.

$$\mathfrak{f} = \varphi(r_{\mathfrak{I}},\varkappa')$$

where $\varkappa$' is the output of $\mathcal{T}_{\chi_{\mathfrak{I}}}(\varkappa)$.
From there, we can also express $\mathfrak{f}$ in the following way:

$$\mathfrak{f} = \tilde{\xi}\ (r_{\mathfrak{I}},\mathcal{T}_{\chi_{\mathfrak{I}}}(\varkappa)) = \tilde{\xi}\ (\tilde{\zeta}\,(\mathcal{T}_{\psi_{\mathfrak{I}}}(\varkappa),\mathcal{T}_{\chi_{\mathfrak{I}}}(\varkappa))$$

B.8.2 Definition

A *paradigm* $\mathfrak{P}_{\mathfrak{I}}$ of a given inflecteme $\mathfrak{I}$ is the set of all form-morphosyntactic feature structure pairs $(\mathfrak{f},\varkappa)$ such that $\varkappa \in \mathcal{K}_{\mathfrak{I}}$ and

$$\mathfrak{f} = \tilde{\xi}(\tilde{\zeta}\,(\mathcal{T}_{\psi_{\mathfrak{I}}}(\varkappa),\mathcal{T}_{\chi_{\mathfrak{I}}}(\varkappa))$$

B.8.3 Formal definition of canonical inflection

From the definitions of non-canonical phenomena above, we can deduce the following definition of *Canonical Inflection*.

Definition of Canonical Inflection

$\exists F$; such that $\forall \varkappa \in \mathcal{K}_{\mathfrak{I}}$, $\tilde{f}_{\mathfrak{I}}(\varkappa, F)$ which means that $|\{\tilde{\xi}\,|\,\xi \in \mathrm{X}_{f_{\mathfrak{I}}}\}| - 1 = 0$

$\exists \Gamma$, such that $\forall \varkappa \in \mathcal{K}_{\mathfrak{I}}$, $\tilde{s}_{\mathfrak{I}}(\varkappa, \Gamma)$; where Γ is a function independant from $\varkappa$

and $\mathcal{T}_{\chi_{\mathfrak{I}}}$ = id
and $\mathcal{T}_{\psi_{\mathfrak{I}}}$ = id
and $\mathcal{K}_{\mathfrak{I}} = \mathcal{K}_{\mathcal{C}\mathfrak{I}}$.

C BUILDING A PARADIGM WITH PARSLI

In this section we give a short example of how PARSLI can be used to model the building of a given inflecteme's paradigm. As an illustration we shall use the simple case of an Italian adjectival paradigm, the paradigm of the inflecteme CARO *dear*.

C.1 Definition of the inflecteme CARO within the lexicon

The inflecteme CARO is defined within the lexicon as the 7-tuple

$$\text{CARO} = (\mathscr{K}_{\text{CARO}}, \mathcal{C}_{\text{CARO}}, \mathrm{s}_{\text{CARO}}, \mathrm{f}_{\text{CARO}}, \mathscr{T}\psi_{\text{CARO}}, \mathscr{T}\chi_{\text{CARO}}, P_{\text{CARO}}),$$

where

$$\begin{aligned}\mathscr{K}_{\text{CARO}} = (&\{\text{GENDER } \mathit{masc}, \text{ NUMBER } \mathit{sg}\}, \{\text{GENDER } \mathit{masc}, \text{ NUMBER } \mathit{pl}\},\\ &\{\text{GENDER } \mathit{fem}, \text{ NUMBER } \mathit{sg}\}, \{\text{GENDER } \mathit{fem}, \text{ NUMBER } \mathit{pl}\}, \})\end{aligned}$$

and the inflecteme's morphosyntactic category is

$$\mathcal{C}_{\text{CARO}} = \text{adjective}$$

Let us note ζ the unique stem zone used for the building of this Italian adjective form. The stem pattern s_{CARO} of CARO associates each possible morphosyntactic feature structure defined for CARO with this unique stem zone ζ.

$$\begin{aligned}\mathrm{s}_{\text{CARO}} = \{&(\{\text{GENDER } \mathit{masc}, \text{ NUMBER } \mathit{sg}\}, \zeta), (\{\text{GENDER } \mathit{masc}, \text{ NUMBER } \mathit{pl}\}, \zeta),\\ &(\{\text{GENDER } \mathit{fem}, \text{ NUMBER } \mathit{sg}\}, \zeta), (\{\text{GENDER } \mathit{fem}, \text{ NUMBER } \mathit{pl}\}, \zeta)\}\end{aligned}$$

Let us note ξ_{masc} and ξ_{fem} the inflection zones used for the building of Italian adjective forms. The inflection pattern f_{CARO} of CARO associates any morphosyntactic feature structure defined for CARO with either of these two inflection zones, depending on the corresponding gender feature.[17]

$$\mathrm{f}_{\text{CARO}} = \{(\{\text{GENDER } \mathit{masc}\}, \xi_{masc}), (\{\text{GENDER } \mathit{fem}\}, \xi_{fem})\}$$

The inflecteme CARO does not display form-function mismatches. Its transfer rules hence equal the identity function.

$$\begin{aligned}\mathscr{T}\psi_{\text{CARO}} &= \mathrm{id}\\ \mathscr{T}\chi_{\text{CARO}} &= \mathrm{id}\end{aligned}$$

Having computed all the necessary elements we can now express the inflecteme's pattern P_{CARO}.

$$P_{\text{CARO}} = \{(\zeta, \xi_{masc}, \mathscr{T}\psi_{\text{CARO}}, \mathscr{T}\chi_{\text{CARO}}), (\zeta, \xi_{fem}, \mathscr{T}\psi_{\text{CARO}}, \mathscr{T}\chi_{\text{CARO}})\}$$

[17] Stating the existence of two inflection zones for Italian adjectives has been decided on the properties of some Italian nouns. It is however clear that this is a descriptional choice made by the author and that other representations would be possible as well.

C.2 Building the paradigm of the infecteme CARO

The stem zone ζ is the unique element of the default stem class for Italian adjectives. It associates each morphosyntactic feature structure within $\mathscr{K}_{\text{CARO}}$ with a unique stem formation rule σ. Hence we can compute the stem formation rule for the inflecteme CARO.

$$\forall \varkappa \in \mathscr{K}_{\text{CARO}}\,,\ \sigma(\varkappa) = \mathrm{r}_{\text{CARO}} \text{ where } \mathrm{r}_{\text{CARO}} = [\text{kar}]$$

The two computed inflection zones ξ_{masc} and ξ_{fem} each produce two form realisation rules. The form realisation rules allow for building the four forms belonging to the inflecteme's paradigm.

$\varphi_{masc}([\textsc{gender}\ masc,\ \textsc{number}\ sg]) = \mathrm{r}_{\text{CARO}}+\mathrm{m}_{masc—sg}$ where $\mathrm{r}_{\text{CARO}}+\mathrm{m}_{masc—sg} = [\text{karo}]$

$\varphi_{masc}([\textsc{gender}\ masc,\ \textsc{number}\ pl]) = \mathrm{r}_{\text{CARO}}+\mathrm{m}_{masc—pl}$ where $\mathrm{r}_{\text{CARO}}+\mathrm{m}_{masc—pl} = [\text{kari}]$

$\varphi_{masc}([\textsc{gender}\ fem,\ \textsc{number}\ sg]) = \mathrm{r}_{\text{CARO}}+\mathrm{m}_{fem—sg}$ where $\mathrm{r}_{\text{CARO}}+\mathrm{m}_{fem—sg} = [\text{kara}]$

$\varphi_{masc}([\textsc{gender}\ fem,\ \textsc{number}\ pl]) = \mathrm{r}_{\text{CARO}}+\mathrm{m}_{fem—pl}$ where $\mathrm{r}_{\text{CARO}}+\mathrm{m}_{fem—pl} = [\text{kare}]$

Thus, the paradigm $\mathfrak{P}_{\text{CARO}}$ of the Italian adjective CARO is:

$$\begin{aligned}\mathfrak{P}_{\text{CARO}} = \{&([karo],\ \{\textsc{gender}\ masc,\ \textsc{number}\ sg\}),\ ([kari],\ \{\textsc{gender}\ masc,\ \textsc{number}\ pl\}),\\ &\{([kara],\ \{\textsc{gender}\ fem,\ \textsc{number}\ sg\}),\ ([kare],\{\textsc{gender}\ fem,\ \textsc{number}\ pl\})\}\end{aligned}$$

References

ANDERSON, Stephen R. (1992) *A-morphous Morphology*. Cambridge, UK: Cambridge University Press.

ARONOFF, Mark (1994) *Morphology by Itself*. MIT Press.

BAERMAN, Matthew (2006) Deponency in serbo-croatian. Online Database: http://www.smg.surrey.ac.uk/deponency/Examples/Serbo-Croatian.htm. Typological Database on Deponency. Surrey Morphology Group, CMC, University of Surrey.

BAERMAN, Matthew (2007) “Morphological Typology of Deponency.” In: M. Baerman/G. G. Corbett/D. Brown/A. Hippisley (eds.) *Deponency and Morphological Mismatches*, volume 145, p. 1-19. The British Academy Oxford University Press.

BAERMAN, Matthew/CORBETT, Greville G. /BROWN Dunstan (2010) *Defective Paradigms*. Oxford, UK: Oxford University Press. Proceedings of the British Academy 145.

BONAMI, Olivier/BOYÉ Gilles (2002) “Suppletion and dependency in inflectional morphology.” In: F. V. Eynde/L. Hellan/D. Beerman (eds.) *The Proceedings of the HPSG '01 Conference*. Stanford, USA: CSLI Publications.

BONAMI, Olivier/BOYÉ Gilles (2010) “La morphologie flexionnelle est-elle une fonction?” In: I. Choi-Jonin/M. Duval/O. Soutet (eds.) *Typologie et comparatisme. Hommages offerts a Alain Lemarechal*, p. 21-35. Louvain, Belgique: Peeters.

BOYÉ, Gilles (2006) “Suppletion.” In: K. Brown (ed.) *Encyclopedia of Language and Linguistics* (2nd ed.), volume 12, p. 297-299. Oxford, UK: Elsevier.

BRESNAN, Joan (ed.)(1982) *The Mental Representation of Grammatical Relations*. MIT Press.

CORBETT, Greville G. (2003) “Agreement: the range of the phenomenon and the principles of the Surrey database of agreement.” *Transactions of the philological society* 101: 155-202.

CORBETT, Greville G. (2007) “Canonical typology, suppletion and possible words.” *Language* 83: 8-42.

CORBETT, Greville G./FRASER Norman (1993) “Network Morphology: a DATR account of Russian nominal inflection.” *Journal of Linguistics* 29: 113-142.

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

FRADIN, Bernard/KERLEROUX Françoise (2003) “Troubles with lexemes.” In: G. Booij/ A. R. Janet de Cesaris/Sergio Scalise (eds.) *Selected papers from the Third Mediterranean Morphology Meeting, Topics in Morphology*, p. 177-196. Barcelona, Spain: IULA-Universitat Pompeu Fabra.

KARTTUNEN, Lauri (1989) “Radical lexicalism.” In: M. R. Baltin/A. S. Kroch (eds.) *Alternative Conceptions of Phrase Structure*, p. 43-65. Chicago: University of Chicago Press.

MATTHEWS, Peter H. (1972) *Inflectional Morphology: a Theoretical Study Based on Aspect of Latin Verb Conjugation*. Cambridge, UK: Cambridge University Press.

MATTHEWS, Peter H. (1974) *Morphology*. Cambridge, UK: Cambridge University Press.

ROBINS Robert H. (1959) “In defense of WP.” *Transactions of the Philological Society* 1959, p. 116-144.

SAGOT, Benoît (2010) “The Le*fff*, a freely available, accurate and large-coverage lexicon for French.” In: *Proceedings of the 7th Language Resource and Evaluation Conference*. Valletta, Malta.

SAGOT, Benoît /WALTHER Géraldine (2011) “Non-canonical inflection : data, formalisation and complexity measures.” In: *Proceedings of the workshop Systems and Frameworks in Computational Morphology (SFCM 2)*. Zurich, Switzerland.

STUMP, Gregory T. (2001) *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge, UK: Cambridge University Press.

THORNTON, Anna M. (2010) “Towards a typology of overabundance.” Presented at the D écembrettes 7, Toulouse, France.

WALTHER, Géraldine/SAGOT, Benoît (2011) “Modélisation et implémentation de phénomènes non-canoniques.” *Revue Traitement Automatique des Langues* 52(2).

ZAUNER, Alfonz (1973) *Praktická prírucka slovenského pravopisu*. Martin, Slovakia: Vydavate lstvo Osveta.

ZWICKY, Arnold M. (1985) “How to describe inflection.” In: M. Niepokuj et alii (eds.) *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistic Society*, p. 372-386. Berkeley, USA: Berkeley Linguistic Society.

Abstract
MEASURING MORPHOLOGICAL CANONICITY

The question of regularity within morphological paradigms has been formerly addressed within approaches falling in the scope of *Canonical Typology* (Corbett 2003). The aim of this paper is to provide a means for assessing the notion of morphological canonicity through original measures developed within our new morphological framework PARSLI. In particular, we introduce original measures for non-canonical phenomena such as *heteroclisis*, *deponency*, *defectiveness* and *overabundance*.

We introduce PARSLI a new model for inflectional morphology using an inferential-realisational approach (Matthews 1974; Zwicky 1985; Anderson 1992). Our model precisely provides a formal representation of the lexicon/grammar interface. It relies on a formal definition of a lexical entry and a complete formal apparatus for computing all relevant form realisation rules for each lexeme, including stem formation rules. Realisation rules themselves may be expressed through any suitable realisation-based formalism (e.g. PFM or Network Morphology). We introduce several formal innovations such as *inflection zones*, that constitute partitions of given inflection classes. They are in particular used in modelling heteroclisis.

Povzetek
MERJENJE MORFOLOŠKE KANONIČNOSTI

Vprašanja pravilnosti morfoloških paradigem so se že lotevali pristopi, ki sodijo v okvir *kanonične tipologije* (Corbett 2003). Cilj pričujočega članka je prispevati izvirne načine, ki bodo na podlagi meril, ki smo jih izdelali znotraj našega novega morfološkega modela PARSLI, omogočali ovrednotiti pojem morfološke kanoničnosti. Še posebej pozorno pa smo vpeljali nove načine merjenja nekanoničnih pojavov, kot smo npr. *heterokliza*, *deponentnost*, *nezapolnjnenost*, *prenapolnjenost*.

V članku predstavljamo PARSLI, ki je nov oblikoslovni model, ki se opira na inferenčno-uresničitveni pristop (Matthews 1974; Zwicky 1985; Anderson 1992). Naš model ponuja prav formalno predstavitev slovarsko-slovničnega vmesnika. Temelji na formalni definiciji leksikalne iztočnice in popolnem formalnem aparatu, ki omogoča izpeljavo vseh relevantnih oblikoslovnih uresničitvenih pravil za vsak leksem, kamor sodijo tudi pravila oblikovanja osnove. Uresničitvena pravila lahko oblikujemo znotraj katerega koli ustreznega formalnega modela (na primer, teorija paradigmatskih funkcij ali morfologija mrež /ang. *Network Morphology*/). Vpeljemo vrsto formalnih novosti, na primer *pregibna območja* (ang. *inflection zones*), ki tvorijo dele posameznega pregibnega razreda. Posebej koristni so pri modeliranju heteroklize.