
The other perspective on exponence

Sacha Beniamine

University of Surrey,
Surrey Morphology Group

Mae Carroll

Australian National University,
School of Culture, History & Language

1 Introduction

Stump (2001, 11) defines exponence as “the only association between inflectional markings and morphosyntactic properties.”. This association is at the very core of morphology. Just as Stump’s, most definitions of exponence do not expressly indicate a directionality in the association. Yet, most morphological theories address it strictly in the direction of production, asking how marking realizes morphosyntactic meaning; or more generally how inflectional forms are produced. Production is indeed one of the core problems inflection poses to speakers. Yet, linguistic communication generally involves not only a producer, but also a receiver. Hence, a complementary task to production is that of comprehension: when hearing an inflected form, how can we recognize its morphosyntactic meaning? Which parts of words support these inferences? Surprisingly, this perspective has received much less attention. In this abstract, we describe the problems raised by a comprehension-based theory of exponence; propose a simple implemented algorithm for segmentation of fine grained formatives from inflected paradigms and a theory of exponential meaning grounded in set theory. This lets us study patterns in the discriminative power of formatives, and set the grounds for large scale typology of exponence.

2 The missing half of the story

Most morphological theories are formulated as accounts of production. In Hockett’s influential typology of morphological theories, all three models, item-and-arrangement, item-and-process and word-and-paradigm (Hockett, 1954), are concerned with how to produce morphological forms. Similarly, Stump’s (2001) categories along the lexical-inferential and incremental-realizational axes all apply to models of how language speakers produce inflected forms. Blevins (2016) further distinguishes constructive grammars (which focus on building words from component pieces, bottom-up) from abstractive ones (which focus on describing observed relations, top-down). Although the latter could be more suited to the comprehension perspective, work in this area has mostly focused on a productive question, called the Paradigm Cell Filling Problem (or PCFP, Ackerman et al., 2009): *How do speakers draw inferences to produce unseen forms in morphological paradigms?* An exception in this landscape is the discriminative learners of Baayen et al. (2015, 2019), which predict meaning (distributional vectors) from forms (orthographic, phonemic or acoustic) directly, without identifying any intermediate analytic units. The current work adopts the DISCRIMINATIVE perspective, and introduces the Paradigm Cell Recognition problem (or PCR, parallel to the PCFP): *Given an inflected wordform, what, in its shape, allows speakers to infer its morpho-syntactic properties?* We find that this question, which pertains to the sub-word structure, has not been addressed systematically by morphological theories.

In this perspective, we define exponence as the phenomena by which parts of words provide information about inflectional meaning. Although it may seem that the same analyses could account for both production and comprehension, we find that the pressures in favor of production and discrimination are often at odds, and can lead to different generalizations. Take for

example the small set of Latin verbal forms given in Table 1¹. If we ask how to produce the passive forms from the active forms, a clear generalization emerges: add either of *-ri-*, *-er-* or *-ēr-* to the active forms. Now we ask instead a question relating to comprehension: how can one recognize that a word in this table is a passive ? A sufficient, and fully informative clue here is the presence of the *-r-*. It is not however the only informative sub-sequence: noticing the final short *-i* in *DUCO* narrows down the possible cells to either the present or the passive; and in *AMO* and *VIDEO* to either the future or the passive. Thus, *-i-* and *-r-* have different discriminative power (different exponential value), and in comprehension, should be segmented differently.

Table 1: Latin verbal subparadigms for the second person singular passive and active, present and future of a few exemplar lexemes.

Lexeme	AMO	VIDEO	DUCO	AUDIO
IND.PRS.ACT.2SG	amās	vidēs	dūcis	audīs
IND.PRS.PASS.2SG	amāris	vidēris	dūceris	audīris
IND.FUT.ACT.2SG	amābis	vidēbis	dūcēs	audiēs
IND.FUT.PASS.2SG	amāberis	vidēberis	dūcēris	audiēris

A discriminative theory of exponence departs from some of the familiar principles of economy which are relevant in production. First, in production, it is often useful to normalize un-natural distributions of exponents to more natural and compact "meanings", using mechanisms such as rules of referral Zwicky (1985) or morpheme generalization (Trommer, 2013). For the purpose of comprehension, however, the presence of a formative in a set of cells, be it natural or not, is always informative. Thus, there no need for mechanisms to fix imperfect distributions, as what may seem irregular or incoherent from the angle of production is in fact well discriminated in comprehension (Blevins et al., 2017). Second, in production, the presence of distinct formatives expressing the same morpho-syntactic properties appears as a problem, solved by postulating a set of allomorphs, associated with mechanisms (morpho-phonological constraints, inflection class indexes, etc) allowing to choose the appropriate formative for each word. In comprehension, alternants are all informative, and can be maintained as separate discriminative cues.

3 Discriminative exponence

In order to provide a theory of exponence, we are faced with two methodological problems: that of segmentation and of meaning assignment (Manning, 1998; Spencer, 2012; Trommer, 2013):

- **Segmentation problem:** Given a set of inflected word-forms belonging to the same paradigm, what are the minimally informative sub-sequences (or formatives) ?
- **Meaning Assignment problem:** Given the structured morphological paradigm of a lexeme, and a segmentation of its word forms into formatives, what grammatical information can each formative provide ?

We provide a solution to both, starting from fully inflected (but unsegmented) paradigm forms, organized in paradigm tables, with each cell associated with a set of morpho-syntactic

¹We thank Olivier Bonami for suggesting this particular example.

properties ².

We start with segmentation, isolating sequences which systematically co-vary in the paradigm. We call these *FORMATIVES* (after Pike, 1963) defined as the longest contiguous substrings which always recur together in a paradigm. Formatives are identified on the basis of the set of cells in which they recur, observed in matrices of automatically aligned forms (our method for alignment refines that of Beniamine & Guzmán Naranjo, 2021). The operation is first done separately on each phonological tier, in order to allow for supra-segmental exponence. We call *distribution* the set of cells in which a formative occurs. Any formative present in all cells are identified as constant, inert stems (following Bonami & Beniamine, 2021). We focus on formatives which occur in proper subsets of the paradigm cells. Table 2 shows the alignments for the Latin forms of AMO in Table 1, and the distributions of each formative segment. Note that this segmentation does not coincide with traditional segmentations in stems and exponents. For example, in the case of suppletion, only the common substring to all stems (if any) will be considered inert. Any substring which occurs in a proper sub-set of the cells is informative in discrimination, and thus has exponential power (This is not to say that may not also contribute lexical information).

Table 2: Alignments and distributions of formatives for a small Latin sub-paradigm of AMO

IND.PRS.ACT.2SG	ama:	-	-	-	-	s		
IND.PRS.PASS.2SG	ama:	-	-	r	i	s		
IND.FUT.ACT.2SG	ama:	b	-	-	i	s		
IND.FUT.PASS.2SG	ama:	b	e	r	i	s		
	slots	0	1	2	3	4	5	

		ACT	PASS			ACT	PASS
PRS							
FUT	b	b					e

		ACT	PASS			ACT	PASS
PRS			r				i
FUT			r			i	i

This segmentation procedure produces pairs of formatives and their distribution, eg. $\langle /r/, \{ \{ \text{IND}, \text{PRS}, \text{PASS}, \text{2SG} \}, \{ \text{IND}, \text{FUT}, \text{PASS}, \text{2SG} \} \rangle$. These distributions, stated as sets of feature value sets, form part of the knowledge that a language user can be said to have about a given formative. We provide set-theoretic definitions for formatives and distributions, and show how to derive generalized and accurate definitions of their meaning. This relies on defining the set of all minimal and informative descriptions of the distribution. The association between a formative and this morphosyntactic meaning forms the exponence relationship. In the small sub-paradigm of AMO from Table 1, the information carried by $/r/$, $\text{exp}(/r/)$ is $\{ \text{PASS} \}$ (as it is present exactly in the set of cells described by the feature passive, nothing less, nothing more), whereas $\text{exp}(/i/) = \{ \text{FUT}, \text{PASS} \}$ (as it is present both in all the future cells, and all the passive cells, nothing less, nothing more).

4 Conclusion

We describe a formal theory of discriminative exponence, which describes how information for the comprehension task is organized and distributed in inflected words. Both the procedure for segmentation and meaning assignment are entirely formalized and implemented, and can be applied to machine readable lexicons of inflected forms. This methodology lets us produce com-

²we recognize that such tables are themselves the result of analyses. Within this work, they are taken as axioms, although our theory can be applied to different paradigmatic analyses of the same data, with potential for meta-theoretical comparisons.

parable analyses of typologically different inflectional systems. On this basis, observe patterns of exponence in verbal lexicons of Modern Standard Arabic verbs, French, Georgian, Navajo, Ngkolmpu, and Yaitepec Chatino. We show how our analyses capture generalizations about the discriminative nature of exponents, and produce clear, comparable and reproducible accounts of phenomena of interest in the typology of exponence, such as inflection classes, cumulation, syncretism, verbosity, etc. For the purpose of this presentation, we draw analyses from entire lexicons. We leave to future work the study of analytic variation when sampling sub-sets of forms, as well as the treatment of reduplication and metathesis.

References

- Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.), *Analogy in Grammar*, 54–82. Oxford: Oxford University Press.
- Baayen, R. H., C. Shaoul, J. Willits & M. Ramscar. 2015. Comprehension without segmentation: A proof of concept with naive discrimination learning. doi:10.1080/23273798.2015.1065336.
- Baayen, R. Harald, Yu-Ying Chuang, Elnaz Shafaei-Bajestan & James P. Blevins. 2019. The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity* 1–39. doi:10.1155/2019/4895891.
- Beniamine, Sacha & Matías Guzmán Naranjo. 2021. Multiple alignments of inflectional paradigms. *Proceedings of the Society for Computation in Linguistics* 4(21). <https://scholarworks.umass.edu/scil/vol4/iss1/21>.
- Blevins, James P. 2016. *Word and Paradigm Morphology*. Oxford: Oxford University Press.
- Blevins, James P., Petar Milin & Michael Ramscar. 2017. The Zipfian Paradigm Cell Filling Problem. In Ferenc Kiefer, James P. Blevins & Huba Bartos (eds.), *Morphological paradigms and functions*, 141–158. Leiden: Brill.
- Bonami, Olivier & Sacha Beniamine. 2021. Leaving the stem by itself. In Marcia Haag, Sedigheh Moradi, Andrija Petrovic & Janie Rees-Miller (eds.), *All things morphology: Its independence and its interfaces*, vol. 353 Current Issues in Linguistic Theory, 81–98. Amsterdam: John Benjamins Publishing Company. doi:10.1075/cilt.353.05bon.
- Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10. 210–234.
- Manning, Christopher D. 1998. The segmentation problem in morphology learning. In *Proceedings of the joint conferences on new methods in language processing and computational natural language learning* NeMLaP3/CoNLL '98, 299–305. Stroudsburg, PA, USA: Association for Computational Linguistics. doi:10.3115/1603899.1603953.
- Pike, Kenneth L. 1963. Theoretical implications of matrix permutation in fore (new guinea). *Anthropological Linguistics* 5(8). 1–23. <http://www.jstor.org/stable/30022433>.
- Spencer, Andrew. 2012. Identifying stems. *Word Structure* 5(1). 88–108. doi:10.3366/word.2012.0021.
- Stump, Gregory T. 2001. *Inflectional morphology. a theory of paradigm structure*. Cambridge: Cambridge University Press.
- Trommer, Jochen. 2013. Paradigmatic generalization of morphemes. *Morphology* 23(2). 269–289. doi:10.1007/s11525-013-9226-4.
- Zwicky, Arnold M. 1985. How to describe inflection. *Annual Meeting of the Berkeley Linguistics Society* 11. doi:10.3765/bls.v11i0.1897.