

Semi-separate exponence in cumulative paradigms. Information-theoretic properties exemplified by Ancient Greek verb endings

PAOLO MILIZIA¹

By using the system of Ancient Greek verb endings as a case study, this paper deals with the cross-linguistically recurrent appearance of inflectional paradigms that, though generally characterized by cumulative exponence, contain segmentable “semi-separate” endings in correspondence with low-frequency cells. Such an exponence system has information-theoretic properties which may be relevant from the point of view of morphological theory. In particular, both the phenomena of semi-separate exponence and the instances of syncretism that conform to the Brøndalian Principle of Compensation may be viewed as different manifestations of a same cross-linguistic tendency not to let a paradigm’s exponent set be too distant from the situation of equiprobability.

1 Introduction

The practice of linguistic analysis often shows that a strict and rigid distinction between the fusional type on the one hand, and the agglutinative type on the other hand is not able to account for all the relevant properties that may be exhibited by the formal structure of an inflec-

¹Laboratorio di linguistica e fonetica sperimentale. Università degli Studi di Cassino e del Lazio Meridionale.

tional paradigm. This fact also concerns the distinction between cumulative and separate exponence.² In this paper we shall argue that, indeed, a cumulative paradigm is typically a potentially hybrid paradigm, where specific subsets of cells may be associated with a separate or partially separate kind of exponence and we shall propose that this fact may be ascribed to a general tendency of languages which concerns the relative token frequencies of the different inflectional exponents and which will be defined in more detail below. As a case for our discussion we shall take the system of the personal endings of Ancient Greek verb. As will become clear, this system exhibits a particularly neat example of the kind of phenomenon that we aim to address here.

2 Cumulation and segmentability

The endings of the Ancient Greek thematic imperfect are shown in table 1.

TABLE 1
Ancient Greek thematic imperfect endings

	FULLY CUMULATIVE INTERPRETATION		SEMI-SEPARATE INTERPRETATION	
	active	mediopassive	active	mediopassive
1sg	- <i>n</i>	- <i>mēn</i>	1sg	- <i>n</i>
2sg	- <i>s</i>	- <i>o</i>	2sg	- <i>s</i>
3sg	-	- <i>to</i>	3sg	-
1pl	- <i>men</i>	- <i>metha</i>	1pl	- <i>men</i>
2pl	- <i>te</i>	- <i>sthe</i>	2pl	- <i>t-+-e</i>
3pl	- <i>n</i>	- <i>nto</i>	3pl	- <i>n</i>
2du	- <i>ton</i>	- <i>sthon</i>	2du	- <i>t-+-on</i>
3du	- <i>tēn</i>	- <i>sthēn</i>	3du	- <i>t-+-ēn</i>

This subparadigm is inflected for diathesis, which distinguishes active forms and mediopassive forms, for number, which shows the three values of singular, plural and dual, and for person. The system does not distinguish, however, between the first person plural and the first person dual. Now, it is possible to identify sixteen different cumulative markers associated with the sixteen cells that are included in the paradigm. Nevertheless, such an analysis would completely fail to capture a formal regularity shown by a subset of these cells: if we look at the cells of the dual number and at those of the second person plural, we

²For the use of the expression *separate exponence*, see Coates (2000: p. 618). It should be noted that separate exponence per se does not imply agglutination, i.e., does not exclude the use of strategies different from segmental addition at the level of form.

can observe that the corresponding endings can be analyzed as formed by a sequence of two morphs. Indeed, there is a formal parallelism in accordance with which, in this subparadigm (that includes the cells in grey color in the right part of table 1), the mediopassive endings are identical to the corresponding active endings, the only difference being that, while the active endings begin with a simple voiceless coronal stop, the mediopassive endings begin with a sequence of *s* plus an aspirated voiceless coronal. In other words, we can identify two morphs *-t-* and *-sth-* which are predominantly sensitive to verb voice and the morphs *-e-*, *-on* and *-ēn*, which are completely insensitive to verb voice and appear in the second plural, second dual and third dual respectively.³ If we think that it is not desirable for a theory of morphology to simply overlook such instances of segmentability, we must ask ourselves what is the most appropriate way to describe the phenomenon. It is true that in a formal framework it would be certainly possible to devise a system of rules that can incorporate these decomposed or semi-separate morphs and account for their distribution across the cells of the paradigm by means, for instance, of general or language-specific rule-ordering principles. But our basic point is that the crucial datum for understanding this kind of phenomena is the frequency of occurrence of the paradigm cells that are involved.

3 Exponence types and token-frequency of exponents

In order to pursue our line of argument, we can take advantage of the availability of the *Ancient Greek Dependency Treebank* (henceforth *AGDT*), which contains a morphosyntactically annotated corpus of Ancient Greek texts (Bamman et al. 2009). The figures in table 2 represent the relative frequencies of the possible combinations of morphosyntactic properties related to person, number and diathesis, calculated on the basis of the *AGDT*. The values result from the morphosyntactic annotation included in the XML file.⁴ However, some minor adjustments in the classifications have been made, by means of a computer script, because the original annotation of the treebank is more syntax-oriented

³The formal relationship between *-t-* and *-th-* can also be described, without consequences for our considerations, as *s*-insertion accompanied by mutation of *t* into the corresponding aspirated stop. Of course, such an interpretation implies the adoption of a description framework involving rules of exponence not describable in terms of segmental addition.

⁴The values have been extracted from the whole set of the verbal occurrences belonging to finite moods (36711 tokens). A more fine-grained analysis of the Ancient Greek verb (which goes beyond the purpose of the present paper) should take into consideration how belonging to a specific tense/mood subsystem affects the frequency values of the person-number-voice combinations.

than morphology-oriented.⁵

TABLE 2
Relative frequencies of person-number-voice combinations in the *AGDT*

	active	mediopassive	<i>sum</i>
1sg	0.08087	0.02844	<i>0.10931</i>
2sg	0.08142	0.03143	<i>0.11285</i>
3sg	0.40568	0.14775	<i>0.55343</i>
1pl	0.01994	0.01125	<i>0.03119</i>
2pl	0.01525	0.00714	<i>0.02239</i>
3pl	0.114	0.04821	<i>0.16221</i>
2du	0.00136	0.00054	<i>0.0019</i>
3du	0.00414	0.00256	<i>0.0067</i>
<i>sum (all)</i>	<i>0.72267</i>	<i>0.27733</i>	<i>1</i>
<i>sum (gray)</i>	<i>0.02076</i>	<i>0.01024</i>	<i>0.031</i>

Our point is that these data clearly show that the subparadigm including all the dual cells plus the second plural cells is characterized by a significantly low value of relative frequency. The six cells involved, if we sum their values, have a total relative frequency of 0.03. Now, it is true that this datum is perhaps a bit underestimated because the corpus consists of literary texts, and therefore overestimates the frequency of the third person. Nevertheless, this does not seem to be a significant bias for our analysis, since even if we consider only the dramatic plays included in the corpus the total value of the subparadigm is still very low (to be precise it is equal to 0.04).⁶

The appearance of semi-separate exponence in correspondence with paradigm cells with low relative frequency has a major theoretical consequence because it allows us to posit a relationship between this and an apparently unrelated phenomenon, that is, the cross-linguistically

⁵Thus, verbs with medial morphology but active function (including the “medial futures”) are classified as active in the *AGDT*; a special case is, moreover, that of the aorist passive, which shows active morphology in its endings.

⁶In the remainder of this paper, we shall refer to the data of the whole corpus in that they are statistically more significant as regards the distribution of the diathesis values within the dual number. Moreover, we shall not take into consideration, in our calculations, the syncretism between the 1st sg. active ending *-n* and the 3rd pl. active ending *-n*, and shall consider the two forms as distinct morphs. This is done for the sake of simplicity, since, even if we posit a single syncretic morph, the values of the quantities considered below are not affected in a significant way with respect to our argument. On the other hand, the syncretism between 1st sg. and 3rd pl. was eliminated in the Greek varieties where the *-n* of the 3rd pl. was replaced by *-san*.

recurrent appearance of those particular instances of synchronic syncretism that have been interpreted in structuralist terms as related to the so-called “principle of compensation” (see Brøndal 1940; see also Baerman et al. 2005: pp. 22–23; Milizia 2013).

Brøndal’s principle of compensation predicts that the marked value of a superordinate category will allow a smaller number of subdistinctions in a subordinate category than its unmarked counterpart.

A classical example of the influence of the principle of compensation with respect to syncretism is the massive presence of case syncretism in the dual of Sanskrit nominals (see the declension of Sanskrit *-a-* adjectives shown in table 3). Here, we have only three distinct case forms

TABLE 3
Endings of the Sanskrit *-a-* adjectives

	singular			plural			dual		
	m.	n.	f.	n.	m.	f.	m.	n.	f.
nom.	<i>-as</i>	<i>-am</i>	<i>-ā</i>	<i>-āni</i>	<i>-ās</i>		<i>-au</i>	<i>-e</i>	
voc.	<i>-a</i>		<i>-e</i>						
acc.			<i>-ām</i>						
instr.	<i>-ena</i>	<i>-ayā</i>	<i>-ais</i>	<i>-ābhis</i>	<i>-ābhyām</i>				
dat.	<i>-āya</i>	<i>-āyai</i>	<i>-ebhyas</i>	<i>-ābhyas</i>					
abl.	<i>-āt</i>	<i>-āyās</i>	<i>-ānām</i>						
gen.	<i>-asya</i>		<i>-āyos</i>		<i>-ayos</i>				
loc.	<i>-e</i>	<i>-āyām</i>	<i>-eṣu</i>	<i>-āsu</i>					

for eight case values. In Brøndalian terms, the marked value dual of the superordinate category of number determines the fact that, in compliance with the principle of compensation, the number of distinctions in the subordinate category of case is very small. But a different interpretation of the same facts is possible from a frequency-based point of view (see Milizia 2013; Milizia to appear; see also Greenberg 1966: p. 27ff.; Haspelmath 2006; Hawkins 2004: p. 64ff.). Along this line, the fundamental datum is that the cells involved in compensative syncretism have low values of usage frequency. Indeed, my proposal is to assume the following hypothesis, which postulates the existence of a sort of equilibrium principle in morphological encoding.

HYPOTHESIS:

One of the principles conditioning the organization of inflectional paradigms is a tendency not to let the relative frequencies of the exponents excessively differ from each other, i.e., to keep the system as close as possible to the equiprobability situation, that is, the state in which all the exponents have the same relative frequency and their

probability values constitute, therefore, a uniform distribution.

The crucial fact, setting aside the issue of why morphological systems should be organized in such a way, is that this hypothesis is independently corroborated by the instances of both compensative syncretism and semi-separate exponence. Indeed, both phenomena are typically associated with low frequency cells within paradigm showing cumulative exponence and both phenomena have the effect of allowing the system to avoid the presence of cumulative exponents associated with low frequency cells in a one to one relationship. In both cases, indeed, what we find are exponents that appear in more than one low frequency cell and have, therefore, a relative frequency value greater than those of the single cells with which they are associated.

4 Information-theoretic quantities potentially relevant to compensative syncretism

In order to assess to what extent a set of exponents, or better said, the probability distribution of such a set, complies with the tendency that we have hypothesized, we may, at least in a tentative way, resort to some information-theoretic quantities.⁷

We propose, tentatively, to use quantities based on Shannon entropy (expression 4.1) which represents the expected or averaged value of information carried by a variable, and reaches its maximum when all probabilities are equal (on this and other quantities mentioned in the following discussion, see Shannon 1948; Cover and Thomas 2006: p. 13ff.; Rezā 1961: p. 110; MacKay 2003: pp. 32-34). The maximum possible value of entropy depends on the number of possible outcomes N (in our case, the number of different exponents): more precisely, it is equal to the logarithm of N (see expression 4.2).

$$H(X) = - \sum_{i=1}^N p(x_i) \log_2 p(x_i) \quad (4.1)$$

$$0 \leq H(X) \leq \log_2 N \quad (4.2)$$

In particular, redundancy, R , (expression 4.3) is the absolute value of the fractional difference between the observed entropy value and the

⁷Similar, or partially similar, applications of information theory to inflectional morphology are found in Baayen et al. (2011), Milin et al. (2009), Moscoso del Prado Martín et al. (2004), and particularly in Kostić and Božić (2007). More generally, on the adoption of information-theoretic quantities in the description of paradigms, see also Blevins 2013; Ackerman and Malouf (2013); Ackerman et al. (2009).

corresponding maximum entropy value. Significantly, it can be considered as a normalized measure indicating how much divergence exists between a probability distribution and the corresponding (i.e., related to a set of the same cardinality N) equiprobable distribution.⁸ An equivalent quantity with opposite meaning is the relative entropy Hr (expression 4.4).⁹ If a distribution is uniform, then redundancy is equal to zero and relative entropy is one (expression 4.5).

$$R = \frac{\log_2 N - H(X)}{\log_2 N} = 1 - \frac{H(X)}{\log_2 N} = 1 - Hr(X) \quad (4.3)$$

$$Hr(X) = \frac{H(X)}{\log_2 N} \quad (4.4)$$

$$0 \leq R \leq 1 \quad (4.5)$$

Thus, if we want to assess the effect of a syncretism with respect to the redundancy value, we can calculate the difference between the redundancy value of the exponent set of the paradigm in question and that of a hypothetical second paradigm identical to the first except for not exhibiting that syncretism. Such a difference, which we symbolize by ΔR , can be expressed as in (4.6):

$$\Delta R = \frac{(1 - R_0)[\log_2(N - M + 1) - \log_2 N] + P_{sub}H_{sub}}{\log_2(N - M + 1)} \quad (4.6)$$

Here, the relevant parameters are defined as follows:

- P_{sub} is the probability of the subparadigm involved in syncretism, i.e., the sum of the probability values of the cells with homophonous forms.
- H_{sub} is the entropy of the subparadigm involved in syncretism. This quantity is given, according to the formula reported above in 4.1, by the probabilities of the involved cells normalized so as to sum to 1.¹⁰

⁸On the issue of the possible use of other quantities as indicators of “morphological (dis)equilibrium”, see Milizia (2013: p. 54ff.).

⁹It should be noted that the denomination “relative entropy” (here used as in Shannon 1948) is often employed in information theory to designate a different quantity, i.e., the so-called “Kullback-Leibler divergence” (on which, see Cover and Thomas 2006: p. 19).

¹⁰The mathematical relationships treated here ultimately depend on the decomposability property of entropy and on the corresponding formula (see MacKay 2003: pp. 33–34). In particular, the entropy of the non-syncretic cumulative paradigm can be decomposed as follows:

$$H(X) = H(p_1 + p_2 + \dots + p_M, p_{M+1} + p_{M+2} + \dots + p_N) \\ + (p_1 + \dots + p_M)H\left(\frac{p_1}{p_1 + \dots + p_M}, \dots, \frac{p_M}{p_1 + \dots + p_M}\right)$$

- N is the number of cumulative exponents in the paradigm without syncretism.
- M is the number of cells involved in syncretism.¹¹
- R_0 is the value of R calculated for the exponent set of the paradigm without syncretism.¹²

It is possible to show that the redundancy of an exponent set decreases, i.e., that ΔR has a negative value, for sufficiently small values of the quantity that we have symbolized by P_{sub} . This means that if the cells involved have a sufficiently low frequency then the appearance of syncretism makes the distribution of the exponent set get closer to equiprobability.

5 Information-theoretic properties of semi-separate exponence

Now, we can go back to the phenomenon of semi-separation and compare a paradigm showing semi-separate exponence with a corresponding paradigm with totally cumulative exponence.

It is possible to calculate the entropy value of the exponent set of the two paradigms. In the fully cumulative paradigm each cell corresponds to one and only one exponent so that the collection of the relative frequencies of the cells is identical to the probability vector of the cumulative exponents. In the paradigm with semi-separate exponence, however, the frequency of each exponent depends on the frequencies of all the cells where it appears, which can be more than one. For each

$$+(p_{M+1} + \dots + p_N)H\left(\frac{p_{M+1}}{p_{M+1} + \dots + p_N}, \dots, \frac{p_N}{p_{M+1} + \dots + p_N}\right)$$

In the expression above, p_i is the probability (within the considered paradigm) of a cell i and of the exponent univocally associated with it. If we posit that p_1, \dots, p_M , are the probability values of the cells involved in the syncretism, then P_{sub} corresponds to $(p_1 + \dots + p_M)$ and H_{sub} corresponds to $H\left(\frac{p_1}{p_1 + \dots + p_M}, \dots, \frac{p_M}{p_1 + \dots + p_M}\right)$.

¹¹Note that the term $N - M + 1$ corresponds to the number of the exponents of the paradigm showing syncretism, since the M cells involved in syncretism are replaced by a single cell.

¹²The expression of ΔR given above is obtained from the following equalities, where the index 1 refers to the paradigm with syncretism and the index 0 to the one without syncretism :

$$\Delta R = R_1 - R_0$$

$$R_1 = 1 - \frac{H(X)_1}{\log_2(N-M+1)}$$

$$H(X)_0 = \log_2 N(1 - R_0)$$

$$H(X)_1 = H(X)_0 - P_{sub} \times H_{sub}$$

The first of these expressions defines ΔR , the second and the third follow from the general definition of R ($R = 1 - \frac{H(X)}{\log_2 N}$), while the fourth depends on the decomposability of entropy (see above in footnote 10).

exponent of the semi-separate paradigm, we can take the sum of the frequencies of all the cells that are associated with it.

However, the collection of values obtained in such a way does not represent a probabilistic vector since the overall sum will be greater than one. This is due, obviously, to the fact that the frequency values of the cells with two morphs (i.e., the cells colored in gray in tables 1 and 2) will be counted twice, one time for each of the two exponents. Thus, in order to obtain the relevant probabilistic vector of the exponents, the values must be normalized so as to sum to 1. This step consists in dividing values by $1 + P_{sub}$, where P_{sub} is the probability of the subparadigm showing semi-separate exponence, i.e., the sum of the probability values of the cells with two exponents (see table 4).¹³

TABLE 4
Frequencies of the semi-separate exponents in table 1

	non-normalized	normalized to sum 1
<i>-n(1sg)</i>	0.08087	0.07844
<i>-s</i>	0.08142	0.07897
<i>-</i>	0.40568	0.39348
<i>-men</i>	0.01994	0.01934
<i>-n(3pl)</i>	0.114	0.11057
<i>-mēn</i>	0.02844	0.02758
<i>-o</i>	0.03143	0.03049
<i>-to</i>	0.14775	0.1433
<i>-metha</i>	0.01125	0.0109
<i>-nto</i>	0.04821	0.0468
<i>-t-</i>	0.01525+0.00136+0.00414	0.02013
<i>-sth-</i>	0.00714+0.00054+0.00256	0.00993
<i>-e-</i>	0.01525+0.00714	0.02172
<i>-o-</i>	0.00136+0.00054	0.00185
<i>-ēn-</i>	0.00414+0.00256	0.0065
sum	1.031 (1 + P_{sub})	1

¹³The rationale behind this normalization may be clarified as follows. If we have a corpus containing N occurrences of inflected forms belonging to the paradigm at issue, it will be expected to contain $N \times (1 + P_{sub})$ occurrences of inflectional exponents related to that paradigm (since $N \times (1 - P_{sub})$ verb occurrences are expected to contain one exponent and $N \times P_{sub}$ verb occurrences are expected to contain two of them). At the same time, in that corpus, an exponent x that appears in a cell a having probability p_a and in a cell b having probability p_b will be expected to have $N \times (p_a + p_b)$ occurrences. Now, if the set of the occurrences of inflectional exponents is considered in isolation, the expected relative frequency of the exponent x within that set will be $[N \times (p_a + p_b)]/[N \times (1 + P_{sub})]$, i.e., $(p_a + p_b)/(1 + P_{sub})$.

The crucial fact is that a fixed relationship exists between the entropy value of a paradigm with semi-separate exponence and the one of the corresponding paradigm with fully cumulative exponence. This relationship can be defined by the following expression:

$$H(X)_{ssep} = \frac{H(X)_{cum} + P_{sub}I(F\alpha; F\beta)_{sub} - P_{sub} \log_2 P_{sub}}{1 + P_{sub}} + \log_2(1 + P_{sub}) \quad (4.7)$$

Here, $I(F\alpha; F\beta)_{sub}$ is the mutual information between the two variables represented by the two sets of grammatical category values that are expounded (semi-)separately within the semi-separate subparadigm. In our example $F\alpha = \{\text{active, mediopassive}\}$ and $F\beta = \{\text{2nd person plural, 2nd person dual, 3rd person dual}\}$.

Generally speaking, the mutual information between two variables represents how much on average our uncertainty about one variable decreases when we learn the value of the other variable. Mutual information can be calculated by means of the expression (4.8).¹⁴

$$I(X; Y) = H(X) + H(Y) - H(X, Y) \quad (4.8)$$

As concerns our example, the relevant probability distributions are reported in table 5.

It should be noticed, at this point, that if the two sets of category values involved in semi-separate exponence are mutually independent, then their mutual information is close to zero (in fact, it is ideally equal to zero). As is known, the fact that the values of two categories are cumulatively expressed by the same set of exponents does not imply an interdependency of those categories at a syntactic or functional-semantic level. On the contrary, what frequently, if not mostly, occurs is that the cumulatively expounded categories are scarcely or not at all correlated to each other. This is what happens in our case, in which the mutual information value is 0.0025 bit. Under this circumstance the whole term containing mutual information can be overlooked, and expression (4.7), can be rewritten as follows:

¹⁴On mutual information, see Cover and Thomas (2006: p. 21). It should be noticed that the fact that this mutual information value appears in expression (4.7) is directly related to the definition of $I(X; Y)$ given in (4.8). Indeed, if a paradigm shifts from cumulative to semi-separate exponence, then a subset of cumulative endings that is associated with the joint distribution of the two involved categories (i.e., *-te, -ton, -tēn, -sthe, -sthon, -sthēn*) is replaced by two sets of semi-separate endings that are associated each with one of the corresponding marginal distributions (i.e., *-t-, -sth-*, and *-e, -on, -ēn*).

TABLE 5
Entropy and Mutual Information of the feature values included in the
semi-separate subparadigm

NON-NORMALIZED VALUES

	active	mediopassive	sum
2pl	0.01525	0.00714	0.02239
2du	0.00136	0.00054	0.0019
3du	0.00414	0.00256	0.0067
sum	0.02076	0.01024	0.031 (P_{sub})

NORMALIZED VALUES

	active	mediopassive	sum
2pl	0.49209	0.23022	0.72231
2du	0.04394	0.01757	0.06511
3du	0.13356	0.08260	0.21616
sum	0.66959	0.33040	1

ENTROPY VALUES

joint $H = 1.97688$	H of $F\beta = 1.06412$
H of $F\alpha = 0.91533$	

mutual information = 0.00257

$$H(X)_{ssep} = \frac{H(X)_{cum} - P_{sub} \log_2 P_{sub}}{1 + P_{sub}} + \log_2(1 + P_{sub}) \quad (4.9)$$

Significantly, from this expression we can infer that the entropy of the exponent set of a semi-separate paradigm will be greater than the one of the corresponding fully cumulative paradigm if the inequality in (4.10) is satisfied.

$$H(X)_{cum} < \frac{(1 + P_{sub}) \log_2(1 + P_{sub})}{P_{sub}} - \log_2 P_{sub} \quad (4.10)$$

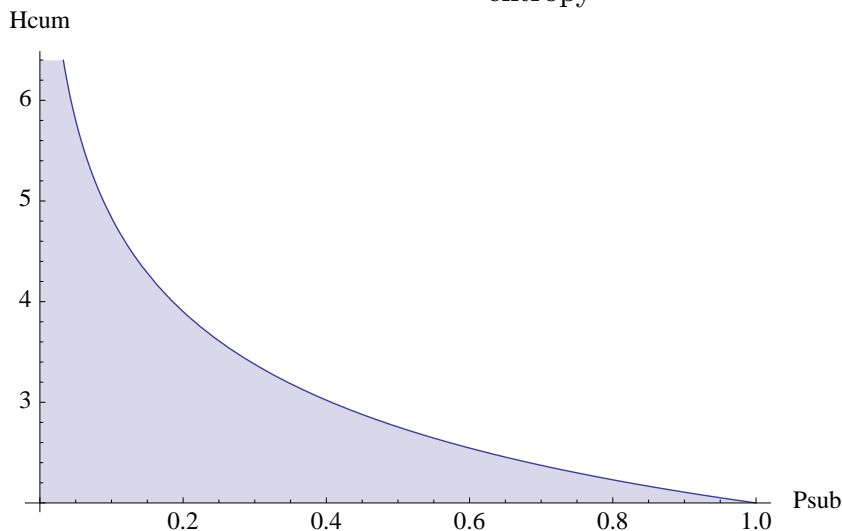
The right term tends to infinity — thereby ensuring that the inequality holds — as P_{sub} tends to zero. In other words, if the probability of the semi-separate subparadigm is sufficiently small, then the shift from cumulative to semi-separate exponence involves an increase in entropy. As concerns our example, the entropy of the exponent set goes from 2.79 to 2.9.

In general terms, the gray area in the graph in figure 1 represents the combination of values of P_{sub} and of the entropy of the corresponding

fully cumulative paradigm for which the appearance of semi-separate exponence involves a greater entropy value.

FIGURE 1

Conditions under which semi-separate exponence involves an increase in entropy



By definition, an increase in entropy means getting closer to the equiprobable distribution. This is relevant for our issue independently of the question whether the shift between a fully cumulative paradigm and a corresponding paradigm with semi-separate exponence involves a change in the number of exponents.

Moreover, if we choose R as an indicator of the deviation from equiprobability, the possible change in the number of exponents is also relevant. Indeed, if the semi-separate subsystem involves more than four cells, then the appearance of semi-separate exponence also causes a decrease in the number of the exponents N . This fact, in turn, reduces the maximum possible entropy value ($\log_2 N$) and thereby contributes, according to (4.3), to reducing the value of R .

A last relevant quantity is represented by the expected length of the inflectional part of a word belonging to a given paradigm. In a fully cumulative paradigm the inflected part of each word contains one and only one morph, in a semi-separate paradigm a word will contain 1 morph with probability $1 - P_{sub}$ and 2 morphs with probability P_{sub} . As a consequence, the average ending length — or, in probabilistic terms, the “expected” ending length — measured in number of morphs will have different values in the two types. These values, which we shall call L_{cum} for the fully cumulative paradigm and L_{sssep} for the paradigm with semi-separate exponence, can be expressed by the following equalities:

$$L_{cum} = 1 \quad (4.11)$$

$$L_{sssep} = 1 + P_{sub} \quad (4.12)$$

By definition, L_{sssep} is greater than L_{cum} , i.e., the semi-separate paradigm has, on the average, longer endings (in terms of number of morphs) than the fully cumulative paradigm.

To sum up, when applied to low frequency cells, both syncretism and semi-separate exponence make the distribution of exponents get closer to the equiprobable distribution. But while syncretism achieves this effect at the cost of losing a certain amount of information, semi-separate exponence does so at the cost of increasing the expected length expressed in morphs of the inflectional part of the words.

6 A few cross-linguistic examples

In order to find instances of the phenomena that we have described, we may look at paradigms that show particularly rare morphosyntactic properties. In this regard, the presence of the dual number is a good hint to follow. It is not difficult to find, across languages, paradigm cells associated with the dual that exhibit some kind of syncretism or some instance of semi-separate exponence or both. Here we shall limit ourselves to three examples, just in order to show in which way (and how easily) cross-linguistic evidence supporting our general hypotheses can be gathered (see also Milizia 2013: p. 89ff.).¹⁵ Firstly, when we look at the non-singular numbers of the Ancient Greek verbal inflection we can see that, in addition to semi-separate exponence by means of the morphs *-t-* and *-sth-*, Ancient Greek verb endings show the following phenomena (see table 6):

- homophony between 2nd pl. primary and secondary ending;¹⁶
- homophony between 2nd du. primary ending, 3rd du. primary ending, and 2nd du. secondary ending;
- homophony/morphosyntactic neutralization between 1st plural and 1st dual.

¹⁵It should be noted that, in the absence of arguments to the contrary, we must necessarily attribute a universal character to our hypothesis, i.e., assume that the tendency to morphological equilibrium is not a language-specific property, but a force active (even when not apparent) in the diachronic development of all linguistic systems.

¹⁶According to our approach, which focuses on the frequency of exponents, instances of systematic homophony concerning parts of inflected words are as relevant as cases of whole-word syncretism (on the distinction between whole-word syncretism and non-whole-word — or, according to Stump's terminology, "block" — syncretism, see also Stump 2001: p. 217 and Milizia 2013: p. 49).

TABLE 6
Ancient Greek non-singular verb endings

		plural		dual	
		primary	secondary	primary	secondary
active	1st	<i>-men</i>			
	2nd	<i>-t-+-e</i>		<i>-t-+-on</i>	
	3rd	voc. mora+ <i>si</i>	<i>-n/-san</i>	<i>-t-+-ēn</i>	
mediop.	1st	<i>-metha</i>			
	2nd	<i>-sth-+-e</i>		<i>-sth-+-on</i>	
	3rd	<i>-ntai</i>	<i>-nto</i>	<i>-sth-+-ēn</i>	

A second example can be taken from the personal pronoun of classical Arabic (table 7, see Fischer 2006: pp. 124–125), which is inflected for gender (masculine and feminine) and number (singular, plural, and dual). Here, the dual number shows:

- homophony between masculine and feminine in the 2nd person and in the 3rd person;
- homophony/morphosyntactic neutralization between 1st plural and 1st dual;
- semi-separate exponence of the dual number in the 2nd person and in the 3rd person: indeed, an element *-ā* is added to the corresponding masculine plural form in order to have the dual form.

TABLE 7
Classical Arabic personal pronoun

	singular	plural	dual
1st m.-f.	^ʔ <i>anā</i>	<i>nahnu</i>	
2nd m.	^ʔ <i>anta</i>	^ʔ <i>antum</i>	^ʔ <i>antum-ā</i>
2nd f.	^ʔ <i>anti</i>	^ʔ <i>antunna</i>	
3rd m.	<i>huwa</i>	<i>hum</i>	<i>hum-ā</i>
3rd f.	<i>hiya</i>	<i>hunna</i>	

As a third example, in Nganasan (Samoyedic group) we can find an ideal example of coexistence of syncretism and semi-separate exponence. Here, the conjugation of transitive verbs shows number agreement for both subject and object, and in both cases number is structured according to the triple distinction between singular, plural and dual. Not surprisingly, precisely in correspondence with this double presence of the dual number, we find the phenomena relevant to our

discussion. If we look at table 8 (see Helimski 1998: pp. 502–504), we can see:

- syncretism between 2nd person subject and 3rd person subject in both plural-object and dual-object conjugations;
- semi-separate exponence of the dual-object agreement: i.e., all dual-object forms are identical to the corresponding plural-object forms except for the presence of a suffix *(-a)kəi-* before the subject-agreement ending.

TABLE 8

A part of the paradigm of the Nganasan verb *kotu-* ‘to kill’

S ↓ : O →	indefinite	singular	plural	dual
1sg	<i>kodaʔam</i>	<i>kodaʔamə</i>	<i>kodaʔinʔə</i>	<i>kodaʔakəinʔə</i>
2sg	<i>kodaʔaŋ</i>	<i>kodaʔarə</i>	<i>kodaʔitʔə</i>	<i>kodaʔakəitʔə</i>
3sg	<i>kodaʔa</i>	<i>kodaʔatu</i>	<i>kodaʔitʔü</i>	<i>kodaʔakəitʔü</i>
1pl	<i>kodaʔamuʔ</i>		<i>kodaʔinʔüʔ</i>	<i>kodaʔakəinʔüʔ</i>
2pl	<i>kodaʔaruʔ</i>		<i>kodaʔitʔüʔ</i>	<i>kodaʔakəitʔüʔ</i>
3pl	<i>kodaʔaʔ</i>	<i>kodaʔaδuŋ</i>	<i>kodaʔitʔüŋ</i>	<i>kodaʔakəitʔüŋ</i>
1du	<i>kodaʔami</i>		<i>kodaʔinʔi</i>	<i>kodaʔakəinʔi</i>
2du	<i>kodaʔari</i>		<i>kodaʔitʔi</i>	<i>kodaʔakəitʔi</i>
3du	<i>kodaʔagəj</i>	<i>kodaʔadi</i>		

7 Brief remarks on the theoretical status of a principle of equilibrium in morphological encoding

It should be stated explicitly that the cross-linguistic tendency hypothesized above cannot be considered as the only factor determining the organization of morphological exponence within paradigms. In particular, as regards syncretism, it should be noted that if the equilibrium-hypothesis can be applied to “compensative syncretism”, i.e., to instances of syncretism that, in Brøndalian terms, might be viewed as effects of the “principle of compensation”, it is also true that phenomena of syncretism that involve cells of relatively high frequency — and are, therefore, in absolute disaccordance with the equilibrium tendency — are all but infrequent.

But the fact that not all syncretism is compensative does not imply that “compensative syncretisms” do not exist at all. Consider, for example, the paradigm of the Sanskrit *-a-* adjectives (shown in table 3 above) on the one hand, and the present indicative of the French verb *chanter* ‘sing’ on the other (see table 9).

TABLE 9
Present indicative of French *chanter*

	Singular	Plural
1st	fāt	fāt̃
2nd		fāte
3rd		

There can be no doubt that the paradigm of *chanter*, where syncretism affects all cells but the two that are least frequent, severely contradicts the tendency to equilibrium in morphological encoding. But do this and similar cases suffice to make us reject the hypothesis that such a tendency exists? A positive answer to this question would lead us to the idea that the blatant correlation between frequency and syncretism shown by Sanskrit nominals is no more than a matter of chance. Thus, if we are not willing to accept this unattractive conclusion, we are compelled to suppose that more than one type of possible structural/diachronic grounds for syncretism insurgence should be identified, with the tendency to equilibrium in morphological encoding representing only one of them.¹⁷

Generally, two observations can be made:

1. Both functional/morphosyntactic and purely morphological (“morphomic”)¹⁸ principles of organization can underlie the appearance of syncretism instances that are completely extraneous to compensation phenomena and independent of the tendency to equilibrium in morphological encoding.
2. A cross-linguistic tendency towards the “phonological optimization” of paradigms is often responsible for the appearance of syncretism phenomena that, in complete contrast with “compen-

¹⁷It would be difficult not to agree with a reviewer of this paper when he observes that, all other things being equal, one would prefer a single explanation for the appearance of syncretism, whether it affects common or rare cells, and notes that, e.g., a general preference for a smaller set of exponents might be posited as a cause of syncretism independently of frequency data. But in the opinion of the present author, it would not be correct to say that all other things remain equal once the additional assumption is made that there is a link between certain phenomena of syncretism and a tendency towards uniformity in the probability distribution of exponent sets. Indeed, as we have noted, such an assumption is capable of accounting for the specific distribution of syncretism within paradigms like that of the Sanskrit *-a-* adjectives, while the hypothesis of a general preference for a smaller set of exponents is not.

¹⁸In particular, phenomena of inflectional homophony accidentally arisen by virtue of sound changes may secondarily determine the creation of a purely morphological solidarity between paradigm cells so as to make them share the same exponents (see the notion of morphomic index in Baerman et al. 2005: p. 173ff.).

sative syncretism”, typically involve high-frequency cells.

The latter statement, which deserves some explanation, ultimately depends on the fact that, at a cross-linguistic level, inflectional paradigms tend to have phonologically more complex markers in correspondence with less frequent paradigm cells (and, consequently, phonologically less complex markers with more frequent cells) rather than vice versa.¹⁹ Indeed, this tendency, which represents a means of optimizing phonological encoding in that it implies a minimization of the expected phonological length/complexity of an ending,²⁰ has the side-effect of favoring the appearance of syncretism in high-frequency cells. This straightforwardly follows from the fact that, given two endings of equal length/complexity, the shorter (or less complex) they are, the more likely it is that they are identical.²¹

Thus, very frequent patterns of syncretism involving cells of relatively high-frequency (such as, e.g., a syncretism affecting two or more persons in the singular of verbal paradigms accompanied by absence of syncretism in the corresponding plural forms) must be regarded as related to structural tendencies that are completely independent of the phenomenon of compensative syncretism.

In conclusion, cases of “non-compensative” syncretism, i.e., cases of syncretism that do not comply with the “compensative tendency” do not necessarily constitute counter-evidence for the hypothesis proposed above.

A central question is: if our approach is correct, how can it be that morphological grammars tend to be organized in such a way?

The position that we want to hold here is that the locus of the tendency towards exponent equiprobability is to be identified with language change. That is, the synchronic hypothesis formulated above implies a general hypothesis about diachrony that might be stated as follows:

DIACHRONIC HYPOTHESIS: All other things being equal, given two possible language changes, the one that is less in accordance with the ‘equilibrium principle hypothesis’ (see above) is less likely to occur.

It is evident that our diachronic hypothesis cannot be directly corroborated, in that it is practically impossible to observe more historical

¹⁹On the question of how to demonstrate and how to explain such a tendency, see Milizia (2013: p. 155ff.).

²⁰In this case, what is meant is a length measured in terms of phonological/prosodic features or units.

²¹Complementarily, from a diachronic point of view, the shorter two endings are, the more probable it is that they will become identical because of some sound change.

developments starting from exactly the same system conditions. In our view, the main support to the existence of such a preferential direction in language change is indirectly provided by synchronic cross-linguistic evidence, which can be interpreted in the light of the fact that the morphological grammar of a language is necessarily the product of a series of historical developments occurred to the systems (or “grammars”) characterizing its previous stages.²²

When historical or historical-comparative evidence allows us to follow the evolution of a language from a diachronic stage to another, it is sometimes possible to get a concrete picture of how linguistic change can comply with the preference postulated here.

In this regard, the development between the Proto-Indo-European verb ending system and the Ancient Greek one is particularly revealing.

8 From Proto-Indo-European to Ancient Greek

In Proto-Indo-European (see table 10) the 2nd singular, 3rd singular, and 3rd plural endings (of both primary and secondary set) showed a morph **-o* specifically associated with the mediopassive diathesis. For instance, the second singular had a secondary active ending **-s* and a corresponding mediopassive ending **-s-o*, a primary active ending **-s-i* and a corresponding mediopassive ending **-s-o-i* (see Rix 1992: pp. 239–266).

TABLE 10
A subset of Proto-Indo-European verb endings

	primary		secondary	
	active	mediopassive	active	mediopassive
2sg	<i>*-si</i>	<i>*-soi</i>	<i>*-s</i>	<i>*-so</i>
3sg	<i>*-ti</i>	<i>*-toi</i>	<i>*-t</i>	<i>*-to</i>
3pl	<i>*-nti</i>	<i>*-ntoi</i>	<i>*-nt</i>	<i>*-nto</i>

This kind of semi-separate exponence of the morphosyntactic property [mediopassive] was lost in Ancient Greek because of a combination of sound changes and inflection marker replacements. The most relevant

²²It should be emphasized that the distinction between “compensative” and “non-compensative” syncretism is not to be conceived as a classification relevant to the synchronic structure of languages. It is rather a heuristic tool for assessing the value of cross-linguistic evidence with respect to the equilibrium-tendency hypothesis. In other words, the qualification of “non-compensative” simply means that a syncretism does not make the distribution of the exponent set get closer to equiprobability and cannot, therefore, be included among the evidence supporting our claims.

facts are the following (see Schwyzer 1939: pp. 657–672; Chantraine 2005: pp. 290–308):

- I) In the primary series, the morph *-o-* was replaced²³ by a morph *-a-* on the model of a newly created 1st person singular mediopassive ending *-mai*.²⁴ From this point forward, we have, therefore, *-sai*, *-tai*, *-ntai* in the primary set, but *-so*, *-to*, *-nto* in the secondary one:
- i. *-ti* : *-toi*, *-t* : *-to* → *-ti* : *-tai*, *-t* : *-to*
- II) The primary 2nd sg. ending pair **-si* (active), **-sai* (middle) lost its parallelism since **-si* was replaced by *-s* or *-is*, while *-sai* underwent, in thematic paradigms, a phonological change due to the loss of intervocalic *s*. Such a deletion also affected the secondary 2nd sg. endings *-s*, *-so*:
- i. **légesi* : **légesai* → *légeis* : *légēi*
2nd sg. present of *légō* ‘say’ (thematic)
- ii. **éleges* : **elégeso* → *éleges* : *elégou*²⁵
2nd sg. imperfect (thematic)
- III) As for the primary 3rd sg. pair *-ti*, *-tai*, the active ending *-ti* was replaced by *-i* in the thematic conjugation:
- i. **légeti* : **légetai* → **légei* : **légetai*
- IV) In the athematic conjugation *-ti* was preserved, but, in a group of dialects including Attic, it underwent the *ti* > *si* assibilation. In both thematic and athematic conjugations the *ti* > *si* change also broke the formal parallelism in the 3rd plural primary ending pair *-nti*, *-ntai*:
- i. **deíknūti* : **deíknutai* → *deíknūsi* : *deíknutai*
3rd singular present of *deíknūmi* ‘show’ (athematic)
- ii. **légonti* : **légontai* → *légousi* : *légontai*
3rd plural present (thematic)
- V) The secondary 3rd sg. and pl. pairs *-t*, *-to* and *-nt*, *-nto* were affected by the phenomenon of final stop deletion:
- i. **éleget* : **elégeto* → *élege* : *elégeto*
3rd singular imperfect (thematic)
- *élegont* : **elégonto* → *élegon* : *elégonto*
3rd plural imperfect (thematic)

²³As Mycenaean, Arcadic, and Cypriot show, this change did not affect all Greek dialects.

²⁴In its turn, the ending *-mai* represents a sort of crossing between the inherited **-ai* (*-H₂ei* according to the laryngealistic reconstruction) and the active 1st sg. *-mi* (see Rix 1992: p. 253).

²⁵The digraph *jouj* represents a long high (originally medium-high) back vowel.

It should be noted that in the pre-history of Ancient Greek, sound changes such as *-ti-* > *-si-* or intervocalic *-s-* deletion were blocked (or their outcomes were replaced) in some cases on morphological grounds: see *légonti* (dat. sg. of the present participle); *lúsō* (future of *lúō* ‘solve’). Nevertheless no analogical change restored the semi-separate exponence of diathesis. Significantly, even if in the 2nd sg. of the medial athematic conjugation the intervocalic ending-initial *-s-* was restored,²⁶ such a restoration did not result in the reestablishing of the original parallelism between active and medial endings: in the present tense of *dídōmi* ‘give’, the active form corresponding to the mediopassive 2nd sg. *dídōsai* is *dídōs* and not **dídōsi*; in the imperfect, the 2nd sg. pair act. *edeíknūs*, mediop. *edeíknuso* preserved the original *C : Co* relationship, but such a relationship is synchronically unparalleled within that paradigm.

These facts suggest that in proto-historical Greek the inherited formal parallelism between active and mediopassive endings was not anymore a segmentation cue.²⁷

Thus, the inherited system of semi-separate exponence of diathesis, which was associated with high frequency paradigm cells (2sg, 3sg, 3pl), was abandoned. At the same time, a new system of semi-separate exponence of the same category, which was associated with low frequency paradigm cells (2pl, 2du, 3du), was created.

The *-t-* / *-sth-* alternation is, indeed, a Greek innovation (see Schwyzler 1939: pp. 670f.). If we look at the Old Indic ending system (table 11; see Macdonell 1910: p. 314; Thumb and Hauschild 1939: par. 417ff.), we can see that there is nothing comparable to the *-t-* / *-sth-* parallelism. The starting point of the innovative Greek pattern may have been the 2nd plural, since the segment *-th-* of the Greek mediopassive endings can be compared with the segment *-dh-* that is contained in the Old Indic 2nd plural medial endings *-dhve* (primary) and *-dhvam* (secondary) (see Rix 1992: p. 248).²⁸

²⁶It should be reminded that, in the 2nd sg. of the athematic mediopassive perfect of consonant-final roots, the *-s-* of the ending *-sai* is in postconsonantal position and, therefore, does not undergo deletion. Thus, the phonological alternation between postvocalic *-ai* (< *-sai*) and postconsonantal *-sai* gave rise to a purely morphological opposition between an ending *-ai* specific to the thematic conjugation and an ending *-sai* specific to the athematic one (see Rix 1992: pp. 253f.).

²⁷It should be emphasized that, even if in several cases the loss of segmentability is directly or indirectly favored or caused by the loss of phonological segments or feature specifications due to sound changes, the focus of our argument is represented by the shift from analysable to non-analysable endings and not by the entities of the phonological-prosodic level. Indeed, a loss of segmentability can occur even in absence of phonological deletions.

²⁸The comparison between Greek *-sthe* and Old Indic *-dhvam* / *-dhve* is not completely straightforward from the point of view of historical phonology, since the

TABLE 11
A subset of Sanskrit verb endings

	primary		secondary	
	active	mediopassive	active	mediopassive
1sg	<i>-mi</i>	<i>-e</i>	<i>-m</i>	<i>-i</i>
2sg	<i>-si</i>	<i>-se</i>	<i>-s</i>	<i>-thās</i>
3sg	<i>-ti</i>	<i>-te</i>	<i>-t</i>	<i>-ta</i>
1pl	<i>-mas(i)</i>	<i>-mahe</i>	<i>-ma</i>	<i>-mahi</i>
2pl	<i>-tha</i>	<i>-dhve</i>	<i>-ta</i>	<i>-dhvam</i>
3pl	<i>-nti/-anti</i>	<i>-ante</i>	<i>-n/-an</i>	<i>-anta</i>
1du	<i>-vas</i>	<i>-vahe</i>	<i>-va</i>	<i>-vahi</i>
2du	<i>-thas</i>	<i>-āthe</i>	<i>-tam</i>	<i>-āthām</i>
3du	<i>-tas</i>	<i>-āte</i>	<i>-tām</i>	<i>-ātām</i>

Significantly, the *-t-* / *-sth-* pattern was extended not only to the dual number but also to the third person of the singular (cf. 3sg imperative active *legétō* : mediopassive *legésthō*) and —with some dialectal variation (see Chantraine 2005: p. 271; Schwyzer 1939: p. 801f.; Milizia 2013: p. 89ff.) — plural of the imperative mood. Crucially, in the imperative subparadigm, the third person is relatively infrequent, so that also this development is in line with the tendential correlation between appearance of semi-separate exponence and low usage-frequency of paradigm cells.

In sum, the overall diachronic evolution from Proto-Indo-European to Ancient Greek seems to reflect precisely a preference for having non-analyzable fully cumulative endings in high frequency cells and analyzable semi-separate endings in low frequency cells. This is consistent with our proposals.

9 Concluding remark

In this paper we have tried to show that two phenomena frequently exhibited by cumulative inflectional paradigm in correspondence with cells with low usage-frequency — syncretism and presence of segmentable “(semi-)separate” exponents — may be traced back to the same general tendency concerning the probability distribution of a paradigm’s exponent set, i.e., a dispreference for exponents with a frequency significantly rarer than the average value for the involved

segment *-s-* of the Greek morph finds no counterpart in the Old Indic endings, so that either its presence in Greek or its absence in Old Indic must be ascribed to some phenomenon of reshaping (e.g., reanalysis affecting the morphological boundary between a stem and the following ending). On the other hand, the loss of **-v-* in Greek is expected (see Rix 1992: p. 93).

paradigm. In this respect, we have also proposed, tentatively, that information-theoretic quantities may be used for describing and measuring such a tendency in a scientifically verifiable way. At the same time, according to our approach, the hypothesized preference for keeping a sort of equilibrium in morphological encoding is to be viewed as a product of diachrony.

Appendix: alternative models of semi-separate exponence

In section 5 of this paper, in order to obtain a probabilistic vector related to the exponent set of a paradigm with semi-separate exponence we took the probabilistic vector of the cells of the paradigms, then calculated a (non-probabilistic) vector containing, for each exponent, the sum of the probabilities of the cells where it appears, and finally normalized the so obtained vector to sum to one, which is done by dividing each value by $1 + P_{sub}$.

Now, it might be wondered if it is possible to define the probability distribution of the exponent set in a way that dispenses us from the necessity of a normalization step. In this appendix, we shall discuss two possible modes of modeling the phenomenon of semi-separate exponence that seem to the present author to be worth mentioning and do not involve normalization procedures.

The first method to be discussed consists in dealing with two different exponent sets defined according to the features that are expounded separately in the semi-separate subparadigm.²⁹ In our case, we have to distinguish exponence of person/number and exponence of diathesis, with cumulative exponents belonging to both sets but separate exponents belonging to only one.³⁰

As is clear from table 12, if we compare one or another of the two semi-separate sets with the exponent set related to the corresponding fully cumulative paradigm, we observe a relationship which is perfectly analogous to that existing between a cumulative paradigm with syncretism and a corresponding cumulative paradigm without syncretism. Indeed, as is expected given the low frequency of the involved cells, both the probability distribution related to diathesis and the one related to person/number have a value of R that is lower than that associated

²⁹The possibility of such a treatment was suggested to me by a reviewer of this paper.

³⁰Of course, since both sets include the cumulative exponents, the probability distribution relative to person/number will also contain information related to diathesis and vice versa, a fact that reveals that this model, while doing without normalization steps, has, all the same, a certain amount of intricacy.

TABLE 12
 Entities and probabilities for the two-variables model of semi-separate
 exponence

SEMI-SEPARATE A			SEMI-SEPARATE B		
	active	mediop.		active	mediop.
1sg	<i>-n</i>	<i>-mēn</i>	1sg	<i>-n</i>	<i>-mēn</i>
2sg	<i>-s</i>	<i>-o</i>	2sg	<i>-s</i>	<i>-o</i>
3sg	<i>-</i>	<i>-to</i>	3sg	<i>-</i>	<i>-to</i>
1pl	<i>-men</i>	<i>-metha</i>	1pl	<i>-men</i>	<i>-metha</i>
3pl	<i>-n</i>	<i>-nto</i>	3pl	<i>-n</i>	<i>-nto</i>
2pl			2pl	<i>-e</i>	
2du	<i>-t-</i>	<i>-sth-</i>	2du	<i>-on</i>	
3du			3du	<i>-ēn</i>	

SEMI-SEPARATE A			SEMI-SEPARATE B		
	active	mediop.		active	mediop.
1sg	0.08087	0.02844	1sg	0.08087	0.02844
2sg	0.08142	0.03143	2sg	0.08142	0.03143
3sg	0.40568	0.14775	3sg	0.40568	0.14775
1pl	0.01994	0.01125	1pl	0.01994	0.01125
3pl	0.114	0.04821	3pl	0.114	0.04821
2pl			2pl	0.02239	
2du	0.02076	0.01024	2du	0.0019	
3du			3du	0.0067	

with the probability distribution of the corresponding fully cumulative paradigm, with values of ΔR being equal, respectively, to -0.07177 and -0.04896 .

On the other hand, this way of formalizing the problem has the significant downside of not being able to provide a synthetic indicator of the effect of a shift from cumulative to semi-separate exponence, since it compels us to deal with a different ΔR for each feature.³¹

The second theoretical possibility to be mentioned involves the adoption of a chain model. If we focus on the linear succession of morphs within a corpus, we may view the text as a sequence of morphs and imagine to “extract” the chain of the verb ending morphs from such a sequence.

³¹It should be emphasized that a synthetic indicator cannot be provided on the basis of the joint distribution of the two variables involved by this model. Indeed, such a joint distribution would practically coincide with the probability distribution of the exponent set of the corresponding fully cumulative paradigm, and would, therefore, be of no use for the purpose of assessing the consequences of a shift from a fully cumulative to a semi-separate organization.

If, for instance, at a certain point in a Greek corpus, there is an occurrence of the word *e-legé-tēn* ‘they two said’, then a series of non-verbal forms, and then an occurrence of *é-lege-s* ‘you (sg.) said’, the corresponding “extracted” morph sequence will be:

..., -t-, -ēn, -s, ...

Such a chain will contain three types of symbols:

A fully cumulative morphs: -n (1st sg. act.), -s, -, -men, -n (3rd pl. act.), -mēn, -o, -to, -metha, -nto;

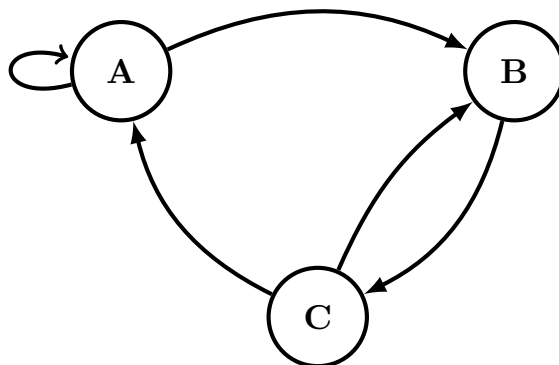
B morphs corresponding to the first part of a segmentable ending: -t-, -sth-;

C morphs corresponding to the second part of a segmentable ending: -e, -on, -ēn;

The morphs of the set A and the ones of the set B are always preceded by morphs of the sets A or C, while the morphs of the set C are always preceded by morphs of the set B. As for these three sets, the chain can be described by the following graph (figure 2):

FIGURE 2

A possible chain model of semi-separate exponence



In particular, in our example, on the basis of the data listed above, we can posit that, if a certain item of the chain is -t-, the subsequent morph will be -e with probability 0.7349, -on with probability 0.06562, and -ēn with probability 0.19948. Analogously, on the basis of the same data, given that an item of the chain is -sth-, the subsequent morph will be -e, -on, or -ēn, with probability values of 0.6968, 0.05319, and 0.25 respectively. Moreover, as for $\rightarrow A$ and $\rightarrow B$ transitions, we may assume, for the sake of the simplicity of the model, that the fact that a verb form belongs to a certain cell of a paradigm does not influence the probability that the next verb form in the text belongs to one or

another paradigm cell.³²

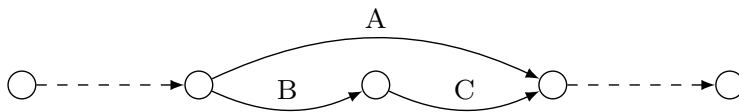
The chain of the verb ending morphs can be modeled as a Markov chain of the first order, i.e., as a sequence of symbols where the probability of occurrence of a symbol depends only on the immediately preceding element or, perhaps more precisely, cannot depend on preceding elements lying further than the immediately preceding one. Indeed, on the basis of our presuppositions, if we take two (identical or different) symbols i and j from the total set of symbols (i.e., morphs), it is possible to determine P_{ij} , i.e., the probability $P(x_t = j | x_{t-1} = i)$ that the t -th element of the chain is j , given the knowledge that the immediately preceding element is i . The set of all the P_{ij} values constitutes a transition matrix (see table 13; the matrix is abbreviated since, given our stipulations, all i -s belonging to A or C determine the same P_{ij} values).³³

As for the initial distribution (i.e., the distribution at $t = 1$, if we label the first item as x_1), we can posit that it is identical to the distribution of an element preceded by an A or C symbol (i.e., we rule out C morphs as possible initial elements since these constitute the second part of an ending).

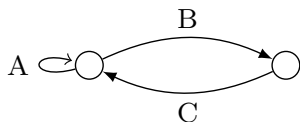
Thus, we obtain a chain that, after a sufficient number of steps,

³²In fact it may be observed that this assumption does not exactly reflect the behavior of texts, since the appearance of a verb form belonging to a paradigm cell may cause an increase of the probability that the next verb form will belong to the same cell. It does not seem, however, that there are reasons to think that this inaccuracy of the model can affect the core of our argument.

³³As a reviewer of this paper points out, another possible way of describing semi-separate morphs (which is perhaps the most appropriate one for the task of formalizing a morphological grammar) is to treat them as transitions between states rather than as states. Such an interpretation corresponds to the following graph:



In our argumentation, the only morphs to be taken into account are those involved in the inflectional paradigm in question. Therefore, the relevant transition network may be represented as follows:



It should be noted, on the other hand, that, for the purpose of our argumentation, we are exclusively interested in the probability values of exponents. In particular, the transition probability matrix relevant to our main problem can be most easily obtained in the framework of the model proposed above, where morphs are interpreted as states: indeed, they are dealt with as if they were words in a bigram model of word sequences (see Manning and Schütze 1999: p. 192ff.).

TABLE 13
Transition matrix for the proposed chain model

$i \downarrow: j \rightarrow$	<i>-n</i> (1sg)	<i>-s</i>	null	<i>-men</i>	<i>-n</i> (3pl)
any A or C	.08087	.08142	.40568	.01994	.11400
<i>-t-</i>	0	0	0	0	0
<i>-sth-</i>	0	0	0	0	0

(continued)

$i \downarrow: j \rightarrow$	<i>-mēn</i>	<i>-o</i>	<i>-to</i>	<i>-metha</i>	<i>-nto</i>
any A or C	.02844	.03143	.14775	.01125	.04821
<i>-t-</i>	0	0	0	0	0
<i>-sth-</i>	0	0	0	0	0

(continued)

$i \downarrow: j \rightarrow$	<i>-t-</i>	<i>-sth-</i>	<i>-e</i>	<i>-on</i>	<i>-ēn</i>
any A or C	.02076	.01024	0	0	0
<i>-t-</i>	0	0	.7349	.06562	.19948
<i>-sth-</i>	0	0	.6968	.05319	.25

will converge to a stationary distribution, i.e., to a steady-state where $P(x_t = i)$ no longer depends on t .³⁴

A first significant fact to be noticed is that such a distribution is identical to the one obtained above by sum normalization; i.e., $P(x_t = i)$ is identical to the probability of the morph i shown in the right-column of table 4: e.g., $P(x_t = -s)$ is 0.07897. In other words, this model gives us the same results as the procedure described above in section 5.

A second fact concerning the chain model is also worth noting. Indeed, for a Markov process it is possible to calculate the entropy rate $H(\mathcal{X})$, a quantity that represents the average information content per symbol (see Cover and Thomas 2006: p. 77). More precisely, given the P_{ij} values contained in the transition matrix and the probability values (symbolized here by μ_i) of the stationary distribution, the entropy rate

³⁴On stationary distributions and how to obtain them from transition matrices, see Cover and Thomas (2006: pp. 72–73,77). It should be noted here that not every first-order Markov chain converges to a steady state. In order for convergence to occur, it is necessary that the chain is irreducible and aperiodic (see Levin et al. 2009: pp. 8, 52ff.). A chain is irreducible when it is possible to transition from any i to any j in a finite number of steps. This property is clearly satisfied for the chain in our example. A chain is said to be aperiodic if, for every state i , the greatest common divisor of the set of the lengths of all paths from i to itself is 1. In particular, if a chain is irreducible and has a state to which it can return in a single step (as is the case for A morphs in our example), then it is also aperiodic (see also Mitrani 1998: p. 164).

is defined as follows:

$$H(\mathcal{X}) = \sum_{ij} \mu_i P_{ij} \log_2 P_{ij} \quad (4.13)$$

Importantly, it can be shown that the entropy rate of our chain can be expressed by the following equality (where the terms P_{sub} and $I(F\alpha; F\beta)$ have the same meaning as above; cf. expression 4.7):

$$H(\mathcal{X})_{ssep} = \frac{H(X)_{cum} + P_{sub}I(F\alpha; F\beta)_{sub}}{1 + P_{sub}} \quad (4.14)$$

If the term $I(F\alpha; F\beta)$ is expected to be near to zero (see above), and if we take into account expression (4.12), then (4.14) can be rewritten as follows:

$$H(\mathcal{X})_{ssep} = \frac{H(X)_{cum}}{L_{ssep}} \quad (4.15)$$

Now, if we extend this model to the fully cumulative paradigm, we can observe that the chain of the cumulative endings has an entropy rate identical to $H(X)_{cum}$.³⁵ This means that shifting from a cumulative paradigm to a corresponding semi-separate paradigm involves a change in the entropy rate that is correlated to the expected length of the semi-separate endings by a relation of inverse proportionality. Thus, in such a framework, we can say that resorting to a semi-separate exponence system has the disadvantage of involving a decrease in the entropy rate of the inflectional paradigm. Such a statement might perhaps represent a mathematically interesting way of expressing the conclusion drawn above at the end of section 5, when it was observed that semi-separate exponence has the cost of increasing the expected length of the inflectional part of the words.

Acknowledgments

This paper is an enlarged version of a section of Milizia 2013 (in Italian). I am grateful to all the participants in the discussion at the TACMO workshop in Geneva. A particular thank is due to Olivier Bonami for his observations. I also wish to thank Marco Mancini, Luca Lorenzetti, and Giancarlo Schirru for their help and support in my research. The author is, of course, responsible for any mistakes or omissions.

³⁵The quantity $\frac{H(X)_{cum}}{L_{cum}}$ equals $H(X)_{cum}$ since L_{cum} is 1 by definition.

References

- Ackerman, Farrell, James P. Blevins, and Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In J. P. Blevins and J. Blevins, eds., *Analogy in Grammar: Form and Acquisition*, pages 54–82. Oxford: Oxford University Press.
- Ackerman, Farrell and Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language* 89/3:429–464.
- Baayen, R. Harald, Dusica Filipović Đurđević, Peter Hendrix, and Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118/3:438–481.
- Baerman, Matthew, Dunstan Brown, and Greville G. Corbett. 2005. *The Syntax Morphology Interface. A Study of Syncretism*. Cambridge: Cambridge University Press.
- Bamman, D., F. Mambrini, and G. Crane. 2009. An ownership model of annotation: The ancient greek dependency treebank. In *The Eighth International Workshop on Treebanks and Linguistic Theories*. Milan: EDUcatt.
- Blevins, James P. 2013. The information-theoretic turn. *Psihologija* 46/4:355–375.
- Brøndal, Viggo. 1940. Compensation et variation, deux principes de linguistique générale. *Scientia* 68:101–109.
- Chantraine, Pierre. 2005. *Morphologie historique du grec*. Paris: Klincksieck, 2nd edn.
- Coates, Richard. 2000. Exponence. In G. Booij, C. Lehmann, and J. Mugdan, eds., *Morphology. An International Handbook on Inflection and Word-Formation*, pages 616–630. Berlin: de Gruyter.
- Cover, Thomas M. and Joy A. Thomas. 2006. *Elements of Information Theory*. Hoboken: Wiley, 2nd edn.
- Fischer, Wolfdietrich. 2006. *Grammatik des klassischen Arabisch*. Wiesbaden: Harrassowitz, 4th edn.
- Greenberg, Joseph. 1966. *Language Universals, with Special Reference to the Feature Hierarchies*. The Hague: Mouton.
- Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42/1:25–70.
- Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford/New York: Oxford University Press.
- Helimski, Eugene. 1998. Nganasan. In D. Abondolo, ed., *The Uralic Languages*, pages 480–515. London: Routledge.
- Kostić, Alexandar and Milena Božić. 2007. Constraints on probability distributions of grammatical forms. *Psihologija* 40:5–35.
- Levin, David A., Yuval Peres, and Elizabeth L. Wilmer. 2009. *Markov Chain and Mixing Times*. Providence (Rhode Island): American Mathematical Society.

- Macdonell, Arthur Anthony. 1910. *Vedic Grammar*. Strassburg: Trübner.
- MacKay, David J. C. 2003. *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press.
- Manning, Christopher D. and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge (Massachusetts): The MIT Press.
- Milin, Petar, Victor Kuperman, Aleksandar Kostić, and R. Harald Baayen. 2009. Words and paradigms bit by bit. an information-theoretic approach to the processing of inflection and derivation. In J. P. Blevins and J. Blevins, eds., *Analogy in Grammar: Form and Acquisition*, pages 214–252. Oxford: Oxford University Press.
- Milizia, Paolo. 2013. *L'equilibrio nella codifica morfologica*. Roma: Carocci.
- Milizia, Paolo. to appear. Patterns of syncretism and paradigm complexity: the case of Old and Middle Indic declension. In M. Baerman, D. Brown, and G. G. Corbett, eds., *Understanding and Measuring Morphological Complexity*. Oxford: Oxford University Press.
- Mitrani, Isi. 1998. *Probabilistic Modelling*. Cambridge: Cambridge University Press.
- Moscoso del Prado Martín, Fermín, Aleksandar Kostić, and R. Harald Baayen. 2004. Putting the bits together: an information theoretical perspective on morphological processing. *Cognition* 94:1–18.
- Rezā, Fazlollāh. 1961. *An Introduction to Information Theory*. New York: Dover.
- Rix, Helmut. 1992. *Historische Grammatik des Griechischen*. Darmstadt: Wissenschaftliche Buchgesellschaft, 2nd edn.
- Schwyzler, Eduard. 1939. *Griechische Grammatik. Erster Band*. Munich: Beck.
- Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27:379–423.
- Stump, Gregory T. 2001. *Inflectional Morphology. A Theory of Paradigm Structure*. Cambridge: Cambridge University Press.
- Thumb, Albert and Richard Hauschild. 1939. *Handbuch des Sanskrit: Eine Einführung in das sprachwissenschaftliche Studium des Altindischen*. Heidelberg: Winter.