PRODUCTIVITY OF THE ROMANIAN *MIHI EST* PATTERN

MIHAELA ILIOAIA[1]

Abstract. The present article explores diachronic changes in the productivity of the Romanian construction [dative + *fi* 'be' + N], instantiated by *mi-e foame* lit. me is hunger – 'I'm hungry', which traces back to the Latin *mihi est* pattern. It presents a corpus-based study, which aims to apply a selection of quantitative and qualitative measurements to a Romanian dataset, in order to measure the productivity of the *mihi est* construction. A holistic, combined approach shows that the Romanian *mihi est* pattern has a very dynamic productivity throughout the centuries and has potential to achieve a higher productivity degree due to the increasing number of *hapax legomena* entering this construction with every historical period.

Keywords: dative, *mihi est*, experiencer, Romanian, syntactic productivity.

1. INTRODUCTION

The Romanian *mihi est* pattern illustrated in (1), in which the verb 'to be' is preceded by a dative clitic and followed by a bare noun, has not received much attention in the literature. The noun in this construction denotes a physiological or psychological state, while the dative argument refers to the experiencer of this state and is the semantic subject, *i.e.* the logical subject of the structure.

(1) Mi- e foame/ sete/ frică/ teamă/ ruşine/ greaţă/ silă/ ciudă/ somn/ …
me.DAT is hunger/ thirst/ fear/ fear/ shame/ nausea/ disgust/ envy/ sleep
'I am hungry/ thirsty/ scared/ scared/ ashamed/ nauseous/ disgusted/ envious/ sleepy'

The aim of this paper is to measure the productivity of this construction in contemporary Romanian and the evolution of its productivity throughout the preceding centuries. More specifically, the following research questions will be addressed:

i. How has the degree of productivity of the *mihi est* pattern evolved since the first attested Romanian texts?
ii. Is the *mihi est* pattern expanding or retracting in productivity in present-day Romanian?

The present study is based on a corpus containing data from both contemporary and old Romanian. If the dataset for contemporary Romanian is quite comprehensive, containing a substantial number of examples of the construction under study, the dataset for old Romanian is more restrained, consisting of a reduced number of examples due to the

[1] Ghent University & FWO-Flanders, mihaela.ilioaia@ugent.be.

very limited number of texts available for this period. In the past years, scholars have shown that, when dealing with small corpora, a multidimensional approach containing quantitative as well as qualitative methods is necessary in order to measure the degree of productivity of a pattern (Baayen 2009, Zeldes 2012, Gyselinck 2018). It will be shown that such an approach is suitable for the Romanian data. The quantitative analysis will be organized considering the three aspects of productivity outlined by Baayen (2009) in his insightful study. The three different aspects of productivity he introduces – realized, expanding, and potential productivity – are defined according to the measuring method used to estimate each of them. As for the qualitative analysis, Barðdal's (2008) approach on syntactic productivity proves to be the most appropriate for the available data.

Section 2 gives an overview of the concept of productivity in linguistics, its theoretical background and a selection of the attempts to measure productivity proposed in the literature. Section 3 describes the collection of the Romanian data and the problems caused by the limitations of the dataset. Finally, in Section 4 the measurements proposed in Section 2 are applied on the Romanian dataset. The last section concludes this study and formulates an answer for the two research questions.

2. PRODUCTIVITY

2.1. Theoretical background

The notion of productivity has a long history in the field of morphology, and it has been applied to syntax as well in more recent studies (Goldberg 1995, Barðdal 2008, Zeldes 2012). Yet, there is no consensus concerning the definition of the concept of productivity, its prerequisites, and the ways it can be measured. Beside different definitions of productivity, various concepts of productivity have been discussed in the literature. However, the boundary between these concepts is not easy to identify.

Morphological productivity has most frequently been defined as the property of a word formation process used by speakers to coin "new/ potential words" (Aronoff 1976: 38), or as "the possibility that a pattern will apply to new forms" (Bybee 1995: 430, Bybee and Thompson 1997: 384). As Barðdal (2008: 25) points out in her monograph on productivity, the definition in terms of "new/ potential words" by Aronoff, referring to new types of a pattern, is clearer than the one proposed by Bybee (1995), and Bybee and Thompson (1997). The latter use the term "new forms" without specifying whether they refer to new types of a specific pattern or to new tokens, understood as new actualizations of the same types. Bauer (2001), in his turn, insists on the spreading of the newly coined "lexical items". For instance, the suffix *-th*, as in *length*, *health*, and *growth*, is considered as being non-productive because it cannot be used in modern English to form new nominalizations, whereas the suffix *-ness* is easily available to speakers for deriving a noun from an adjective (Plag 2003: 44–45). This is the case in *nouniness*, a noun that depicts the extent to which a word behaves as a noun, which, in turn, is derived from the adjective *nouny*, itself productively derived from the noun *noun*.

It is only in recent years that linguists have started applying the notion of productivity to the domain of syntax. Certain linguists have attempted to clarify this process

by proposing clear definitions for the specific concept of syntactic productivity, seen either as type frequency, as variability, or as a combination of the two. From a usage-based perspective, syntactic productivity is favored by high type frequency, *i.e.*, by a high number of different items attested in the relevant slot of a construction (Goldberg 1995; Bybee and Thompson 1997). Furthermore, Goldberg (2006) points out that the variability of items matters at least as much as their absolute number. In other words, a pattern is only productive to the extent that it can be instantiated by a high number of non-similar items. This brings us to another view proposed by Barðdal (2008), where two factors, *i.e.* type frequency and semantic variability, are combined. In her view, productivity is a function of type frequency, semantic coherence (which is the opposite of variability), and an inverse correlation between the two. However, the focus of this study on productivity is not on how it should be defined, but rather how it should be measured and operationalized with respect to the available Romanian data.

2.2. Measuring productivity

Although (Bolozky 1999: 3) states that the exact measurement of a productivity process cannot be seen as a sensible objective, several corpus-based measures have been proposed for measuring different aspects of productivity. Baayen (2009) gives a good overview of these different attempts. In his paper, he identifies three different aspects of productivity based on the measuring method, *i.e.* realized productivity, expanding productivity and potential productivity.

Realized productivity has been defined as the extent of use of a specific pattern (Baayen 1993) and can be gauged by counting the number of types in different historical periods, as suggested by Anshen and Aronoff (1989) or through the structural type distribution measurement proposed by Baayen (2001).

As for expanding productivity, it has been defined as the rate at which a pattern expands by attracting new members. Known as the *hapax-conditioned degree of productivity* (Baayen 1993), it can be estimated by calculating the ratio between the total number of *hapax legomena* occurring in the construction and the total number of *hapax legomena* in the corpus. Gaeta and Ricca (2005) propose an alternative method also based on the count of *hapax legomena*. The measurement realized through the count of *hapax legomena* complements very well the estimation given by type frequency with regard to the degree of productivity of a pattern. However, there is a third aspect that can be estimated neither by type frequency, nor by the number of *hapax legomena*, *i.e.* potential productivity.

Potential productivity is defined as the rate at which the vocabulary increases and is estimated by means of two methods: *category-conditioned degree of productivity* and global productivity (Baayen and Lieber 1991, Baayen 1993).

Another approach to measuring productivity paying special attention to the specifics of syntactic productivity is suggested by Barðdal (2008). In her view, the productivity of a syntactic construction is a function of its type frequency, its semantic coherence, and an inverse correlation between the two. Type frequency and semantic coherence (the opposite of Goldberg's variability) have both been previously mentioned in the literature as affecting

productivity[2]. However, Barðdal (2008) is the first scholar who suggested a systematic link between the two as a predictor for productivity.

3. THE ROMANIAN DATA

In what follows it will be presented how the Romanian data has been gathered and organized in order to respond to the requirements of the present study. Since the aim of this paper is to study the productivity of this construction from a diachronic perspective, a corpus containing examples from all historical periods of Romanian have been gathered, from the first available texts, dating back to the 16th century, until the latest writings attesting present-day language.

3.1. Periodization

Traditionally, six historical periods of Romanian are distinguished:[3] (1) early old Romanian (1521–1640); (2) late old Romanian (1640–1780); (3) pre-modern Romanian (1780–1830); (4) modern Romanian (1830–1899); (5) contemporary Romanian (20th century) and (6) present-day Romanian (from 1989 until today). However, for the purpose of this study only four historical periods were distinguished:

i. Old Romanian, corresponding roughly to the 16th–18th centuries (ORom)
ii. Modern Romanian, covering the 19th century (MRom)
iii. Contemporary Romanian, corresponding to the 20th century (CRom)
iv. Present-day Romanian of the 21st century (PDRom)

3.2. Gathering the corpus

When the data for the present study were gathered, no 'official' corpus of Romanian was freely accessible[4] to the research community, neither for present-day Romanian, nor for modern or old Romanian. Therefore, a corpus has been compiled for the purpose of this study. For present-day Romanian the Romanian Web Corpus 2016 (roTenTen16) provided by the electronic platform Sketch Engine was used. Gathered in 2016, this corpus contains more than two milliard words. As for the older periods of Romanian, a considerable number of old and modern texts have been collected from different sources: some of them from the World Wide Web, but most of them provided by fellow scholars of Romanian.[5]

[2] *Cf.* Bybee (1995: 430), Goldberg (1995), Bybee and Thompson (1997), Clausner and Croft (1997), Barðdal (2001, 2006), Clausner (2002), Croft and Cruse (2004), amongst others.

[3] *Cf.* Gheție (1997: 52–53), adopted also in the Syntax of old Romanian by Pană Dindelegan and Maiden (2016).

[4] A noteworthy project, started by the Romanian Academy, is CoRoLa, – a corpus of contemporary and present-day Romanian containing written and oral texts from 1945 until now. This project started in 2012 and the corpus was launched in December 2017.

[5] I am profoundly grateful to Dana Niculescu, University of Amsterdam, and to Camelia Stan, University of Bucharest, who contributed with several texts to the corpus.

After converting the relevant pdf documents to txt format, these texts have been parsed using the tools provided by Sketch Engine. In this way, it has been possible to compile a corpus for old, modern, and contemporary Romanian, containing in total over eight million words.

3.3. Collecting the data

From these corpora, the relevant data have been extracted and separate datasets have been compiled. The workload was organized in two phases. First, the set of nouns occurring in the *mihi est* structure in each period was established. Next, all potential patterns of experiencer constructions in which these nouns occur were collected for each period. Several patterns were identified. Beside the *mihi est* pattern, instantiated by constructions of the type [dative + *fi* 'be' + noun], patterns with a nominative, a dative or an accusative experiencer have been as well collected: [nominative + *avea* 'have' + noun], [dative + verb + noun] or [accusative + verb + noun].

For the purpose of the present study, two datasets will be considered: (i) the full dataset, containing the examples instantiating all the mentioned patterns with an experiencer, and (ii) the *mihi est* dataset, containing examples instantiating only the *mihi est* pattern. Table 1 gives the number of examples gathered for each of the two datasets in the four historical periods. Note that in both datasets, the total number of examples varies from one period to another, due to the size of the available corpora for each period.

Table 1

The number of examples gathered for the mihi est pattern and for all experiencer constructions

Period		**All experiencer patterns**	***Mihi est* pattern**
ORom	16th – 18th	514	160
MRom	19th	613	335
CRom	20th	1 090	691
PDRom	21st	6 279	2 781
Total		**8 496**	**3 967**

Figure (1a) gives an overview of the list of nouns occurring in the *mihi est* construction over the entire documented period. Figure (1b) provides English translations of the nouns in (1a). The most frequently used nouns in the *mihi est* construction are also the most constant ones throughout the centuries. In order to make the graph more readable, only the nouns with a token frequency higher than or equal to 3 were plotted.

It must be noted that the set of nouns occurring in the *mihi est* pattern is rather stable during the older periods of Romanian, but allows new occurrences in the more recent periods. Out of the total number of nouns occurring in the *mihi est* pattern in the older periods of Romanian, a small number have disappeared during the 19th century (*nevoie* 'need') or during the 21st century (*jind* 'desire'), even though these nouns are still used in the present-day language in other constructions. New nouns occur in this construction in every period (*teamă* 'fear' in the 19th century, *jenă* 'uneasiness' in the 20th century and several nouns in the 21st century).

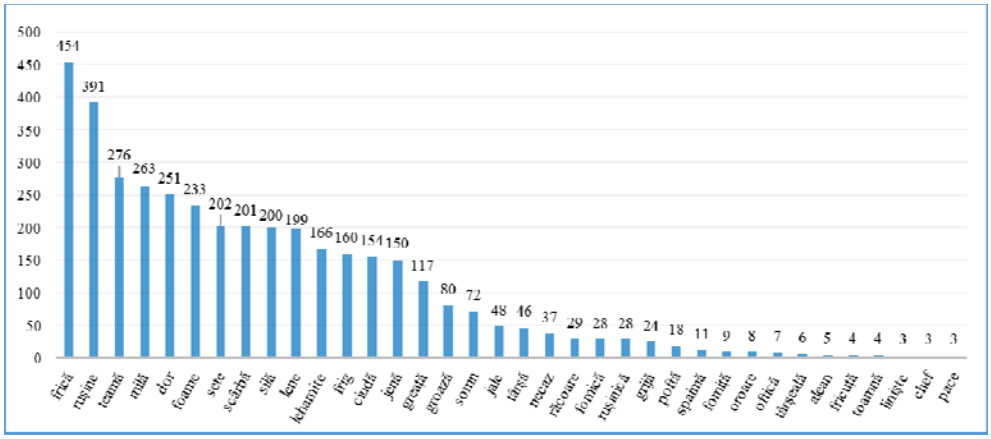

Figure 1a. Nouns in the mihi est construction: 16th – 21st centuries (without hapax legomena).

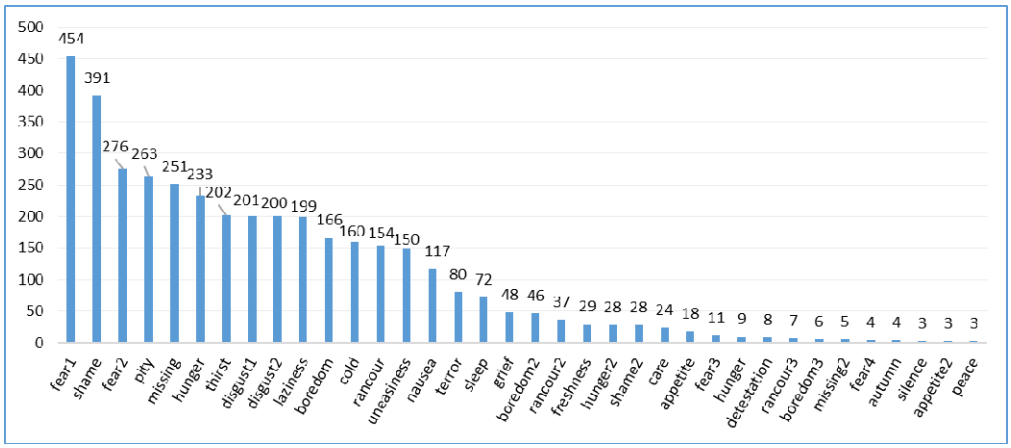

Figure 1b. Nouns in the mihi est construction: 16th – 21st centuries (without hapax legomena) (English).

Summarizing, the use of the present datasets is not unproblematic, first because the number of tokens and the corpus size for each period are very different, and second because the corpora for the different periods are composed differently with respect to text genre. This heterogeneity of the data raises important methodological difficulties, which will be discussed below.

4. OPERATIONALIZING PRODUCTIVITY MEASUREMENTS

Several attempts have been made to tackle productivity from a statistical perspective. When used to measure morphological processes, many of these attempts have produced interesting results. More recently, the same measurements have been applied successfully to the field of syntax as well (Baroðdal 2008, Zeldes 2012, cited by Gyselinck

and Colleman 2016: 76). However, as mentioned above, several methodological difficulties deserve attention.

A first problem concerns working with corpora of different sizes, since, as generally accepted, the productivity index of a process drops as a function of increasing sample size (Bauer 2001: 149). Although this seems to be an important issue, especially for diachronic corpus-based studies, the use of normalized values of frequencies, calculated per million words, should solve this problem. Representativity is affected by the size of the corpus as well. When a corpus is too small to contain enough tokens of a specific pattern, there is a high risk of drawing inaccurate conclusions.

Another important issue is the representativity of the corpus. Representativity implies a balanced combination of texts instantiating different genres. However, for the older periods the available number of texts is so limited that it is impossible to select texts from different genres. Therefore, most of the examples of old Romanian come from literary or religious sources, with a high degree of subjectivity. As for present-day Romanian, for which a web corpus has been used, a high number of examples from news-related sources, hence with a low degree of subjectivity, compete with an almost equal number of examples from social media, which can be considered as having a high degree of subjectivity.

Many of the above-mentioned difficulties also arise in the study of the Romanian data. The heterogeneity of the data requires the use of several quantitative measurements in combination with a thorough qualitative analysis in order to accurately estimate the productivity of the pattern under study. The remainder of this section presents the results obtained when applying the measurements described in Section 2.2 to the present Romanian data.

4.1. Realized productivity

One of the three different aspects of productivity discussed by Baayen (2009) is realized productivity. This aspect of productivity, also called *extent of use*, is restricted to *past achievement* (Baayen 1993). The realized productivity of a morphological category C is obtained by the type frequency V(C,N) of its members in a corpus with N tokens. Baayen points out that, in order to generate comparable figures, the corpora used should have similar sizes. Since the data are very limited for the older periods of Romanian, normalizing the figures could allow applying this measurement to the dataset.

When presenting his first attempt to measure productivity, Aronoff (1976) rejected type frequency as an indication of productivity, claiming that it "isn't fair" because of the morphological restrictions on the base (1976: 36). Bauer (2001: 145) argues against Aronoff's claim showing that type frequency is related to the perceived productivity of a particular process. As an example, he mentions the suffix *-ter*, as in *laugh-ter*, which is not viewed as very productive since it does not occur in many words (Bauer 2001: 145). Aronoff himself revises his point of view about type frequency later on, proposing it as one of the methods of gauging productivity across centuries (Anshen and Aronoff 1999).

In order to apply the measurement proposed by Anshen and Aronoff (1999) to the Romanian data, the different types of state nouns occurring in the *mihi est* pattern were counted for each documented historical period. Table 2 shows the absolute and the normalized type frequency, *i.e.* the number of types normalized per million words for each of the four periods of Romanian.

Table 2
Absolute and normalized type frequency of the Romanian mihi est pattern.

	Type frequency absolute	**Type frequency normalized/ million words**
ORom	14	4.10
MRom	22	9.11
CRom	25	10.57
PDRom	95	0.04

As illustrated in Table 2, the absolute number of types increases considerably in present-day Romanian, compared with previous periods. However, the normalized figures reveal that there is an increase in type frequency in modern and contemporary Romanian, which does not continue in present-day Romanian, where a drastic drop in type frequency can be observed. This drop is surprising. Nevertheless, as pointed out by Baayen (2009), only one measurement of productivity does not reveal every aspect about the productivity of a specific pattern. The different aspects of productivity must be seen as complementary. Therefore, it is necessary to calculate also expanding and potential productivity.

4.2. Expanding productivity

The second aspect of productivity in Baayen's (2009) view, expanding productivity, is reminiscent of Corbin's (1987) *profitability*, but is oriented towards what is expected for the near future. Expanding productivity is estimated by means of the number of *hapax legomena* V(1,C,N) occurring in a morphological category C, in a corpus of N tokens. Let V(1,N) denote the total number of *hapax legomena* in the corpus. The ratio P* = V(1,C,N)/V(1,N) is an estimate of the contribution of a morphological category C to the growth rate of the total vocabulary. This measure is referred to as the *hapax-conditioned degree of productivity* in Baayen (1993), where, unfortunately, the same label P* is used as in Baayen and Lieber (1991) for global productivity.

In order to apply this method to the Romanian data, the ratio between the number of *hapax legomena* occurring in the *mihi est* pattern and the number of *hapax legomena* in the whole corpus was calculated for each historical period. Table 3 presents the degree of expanding productivity of the *mihi est* pattern during the four different documented periods. It shows that the general tendency is towards an increase in expanding productivity throughout the centuries. However, the P* values in the last column of Table 3 indicate that there is a lower degree of expanding productivity in modern Romanian than in the previous or the following periods. Nevertheless, since P* is especially suited for ranking productive processes in Baayen (1993: 194) view, its interpretation is possible only in relation to the results of potential and global productivity measurements, discussed in the next section, whose primary use is to distinguish between unproductive and productive processes.

Table 3

Expanding productivity of the Romanian mihi est pattern throughout the centuries.

	Hapax in *mihi est* V(1,C,N)	**Hapax in corpus V(1,N)**	**Ratio P* V(1,C,N)/V(1,C)**
ORom	6	9	0.67
MRom	2	4	0.5
CRom	7	10	0.7
PDRom	49	32	1.53

4.3. Potential productivity

Finally, potential productivity, also called *category-conditioned degree of productivity* in Baayen (1993), gauges the extent to which the market for a category is saturated. A rule or a process with a low risk of saturation has greater potential for expansion, and hence a greater potential productivity. This type of productivity is calculated by dividing the number of the *hapax legomena* in the corpus V(1,C,N) to the total number of tokens N(C) in that corpus: P = V(1,C,N)/N(C) (Baayen and Lieber 1991). The obtained ratio estimates the growth rate of the vocabulary of the morphological category itself.

The problem is that unequal sample sizes raise crucial complications for this measurement (Bauer 2001: 153). As mentioned *supra*, possible solutions are either to calculate the normalized frequencies per million words or to take equal random samples of the corpus for each category. In the specific context of the Romanian data, the test was applied several times on different sample sizes in order to verify its reliability. First, the application of the formula was done on the original dataset for each historical period (*cf.* Table 4). Then, samples were extracted from each documented period, equal to the smallest dataset, which is the one for old Romanian, counting 160 occurrences of the *mihi est* pattern. The last step was repeated in order to verify whether the samples are representative for the whole corpus. Since the figures for the two equally sized samples were very different from each other, their mean was calculated.

For the readability of the tables, the notation suggested by Baayen (2009) was adopted, adding an index for each sample: the number of the *hapax legomena* is noted with $V_0(1,C,N)$ for the original dataset, and with $V_1(1,C,N_1)$, $V_2(1,C,N_2)$ and $V_m(1,C,N_m)$ for the two samples and their mean, respectively. The total number of tokens in the sample is noted with $N_0(C)$ for the original dataset, $N_1(C)$, $N_2(C)$ and respectively, $N_m(C)$ for the two samples and for their mean. The original dataset and the samples are noted accordingly with N_0, N_1, N_2 and respectively N_m. Similarly, the ratio P is labelled accordingly: P_0 for the original dataset, P_1, P_2 and respectively P_m, for the two samples and their mean.

As mentioned above, the potential productivity of the original dataset was first calculated, and then the calculations for the two samples and their mean were done. As shown in Table 4, for the earliest period of Romanian, this formula estimates a potential productivity index of 0.038. Modern Romanian shows a much lower potential productivity index of 0.006. Note that, although the number of *hapax legomena* decreases to half, the

token frequency doubles between the two periods. This explains the lower potential productivity index for modern Romanian. For the next two periods, the potential productivity index increases, but remains much lower than the productivity index for old Romanian. Interestingly, the number of *hapax legomena* for this period is eight times higher, compared to old Romanian, and the token frequency increases about seventeen times compared to the same period.

Table 4

Potential productivity (P) for the original dataset.

	$V_0(1,C,N)$	$N_0(C)$	P_0
ORom	6	160	0.038
MRom	2	335	0.006
CRom	7	691	0.010
PDRom	49	2781	0.018

Note that normalizing the figures, which is a first step in neutralizing the differences in sample size, does not change the index of potential productivity at all, the same value being obtained when the formula is applied to the absolute figures or to the normalized ones. The second strategy mentioned above is expected to document the validity of this measurement. Table 5 presents the data for each of the two equally sized samples and for their mean, as well as the estimated potential productivity index for each of them per period.

Table 5

Potential productivity (P) for sample N_1 and sample N_2 and the mean values of the two samples.

	$V_1(1,C,N_1)$	$N_1(C)$	P_1	$V_2(1,C,N_1)$	$N_2(C)$	P_2	$V_m(1,C,N_m)$	$N_m(C)$	P_m
ORom	6	160	0.038	6	160	0.038	6	160	0.038
MRom	4	160	0.025	3	160	0.019	3.5	160	0.022
CRom	5	160	0.031	6	160	0.038	5.5	160	0.034
PDRom	10	160	0.063	15	160	0.031	7.5	160	0.047

Comparing the potential productivity values obtained for the original dataset with the ones obtained as a mean of the two equally sized samples, the same tendency towards an increase in productivity can be observed. The high potential productivity index for old Romanian decreases remarkably in the samples, and drastically in the original dataset in modern Romanian. Starting with the latter period, the potential productivity increases steadily until the present-day Romanian period. However, note that in the original dataset, the increase is less significant than in the case of the samples. The potential productivity index for present-day Romanian, estimated at 0.018, represents still less than half of the index for old Romanian (0.038). Figure 2 visualizes the variations between the original dataset and the mean of the two equal samples.

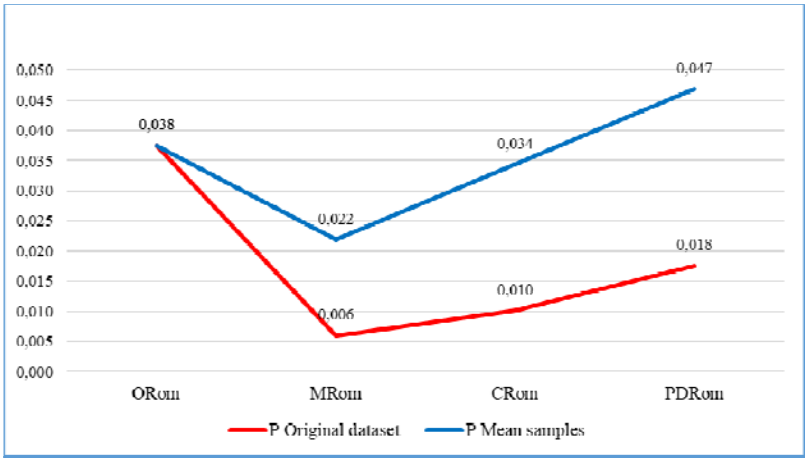

Figure 2. Potential productivity for original dataset and mean of the two samples.

Complementary to potential productivity, which takes into account the number of *hapax legomena* and the token frequency of a specific construction, is global productivity, introduced by Baayen and Lieber (1991). Global productivity brings a more complete perspective to productivity, showing the added value of type frequency, which represents the extent of use of a construction, labeled V, when combined with the potential productivity ratio, labeled P. The authors propose to represent the global productivity of a process on a P-V plane, with the degree of potential productivity (P) on the horizontal axis and the extent of use, *i.e.* the type frequency (V), on the vertical axis.

The application of global productivity to Romanian data consists in adding the type frequency (V) to the already calculated potential productivity index (P), in order to plot it on the P-V plane. The same labels are used as in the case of potential productivity, as well as the labels suggested by Baayen and Lieber (1991) and Baayen (1992). Hence, the type frequency of the original dataset and, respectively, of the two samples and their mean have been labelled $V_0(C,N)$, $V_1(C,N_1)$, $V_2(C,N_2)$, and $V_m(C,N_m)$. The symbols P_0, P_1, P_2, and P_m label the potential productivity index for the original dataset, the two samples, and their mean. The global productivity for the original dataset, the two samples and, respectively, their mean has been labelled P^*_0, P^*_1, P^*_2, and P^*_m.[6] An overview of the fluctuations in global productivity throughout the centuries is given in Table 6 for the complete dataset, and in Table 7 for the two samples and their mean.

Table 6

Global productivity (P^*_0) for original dataset.

	$V_0(C,N)$	**P_0**	**$P^*_0(x;y)$**
ORom	14	0.038	(0.038;14)
MRom	22	0.006	(0.006;22)
CRom	25	0.010	(0.010;25)
PDRom	95	0.018	(0.018;95)

[6] The term *global productivity* was initially labeled *P** by Baayen himself. Later, he confusingly used the same label for the *hapax-conditioned degree of productivity* (Baayen 1993)

Table 7

Global productivity for samples N_1 (P^*_1), N_2 (P^*_2), and for their mean N_m(P^*_m).

	$V_1(C,N)$	P_1	$P^*_1(x,y)$	$V_2(C,N_2)$	P_2	$P^*_2(x,y)$	$V_m(C,N_m)$	P_m	$P^*_m(x;y)$
ORom	14	0.038	(0.038;14)	14	0.038	(0.038;14)	14	0.038	(0.038;14)
MRom	20	0.025	(0.025;20)	19	0.019	(0.019;19)	19.5	0.022	(0.022;19.5)
CRom	17	0.031	(0.031;17)	19	0.038	(0.038;19)	18	0.034	(0.034;18)
PDRom	31	0.063	(0.063;31)	28	0.031	(0.031;28)	29.5	0.047	(0.047;29.5)

Figures (3a) and (3b) represent the visualization of the global productivity of the *mihi est* pattern over the four historical periods of Romanian. The position of each historical period is plotted on the P-V plane, the degree of potential productivity (P) being situated on the horizontal axis and the extent of use, *i.e.*, the type frequency (V), on the vertical axis.

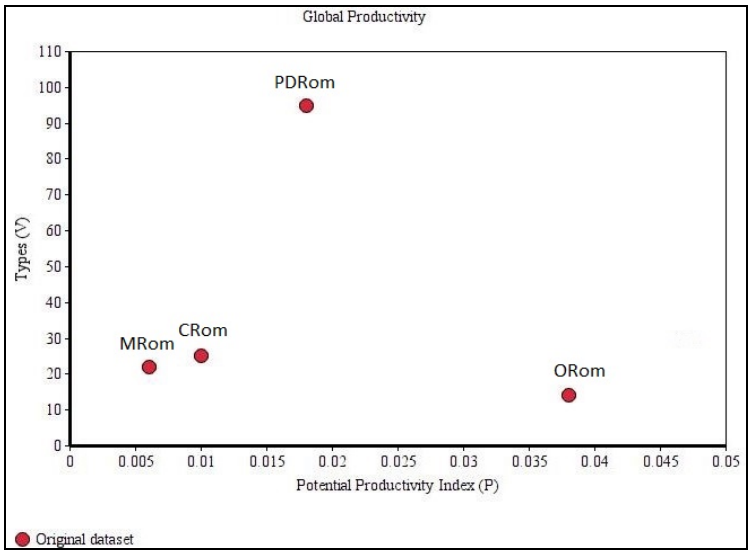

Figure 3a. Global productivity (P*) Original dataset.

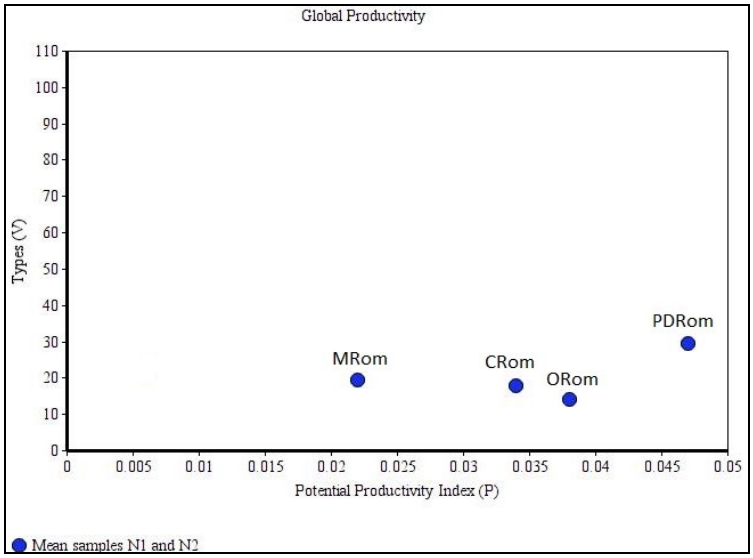

Figure 3b. Global productivity (P*) Mean samples.

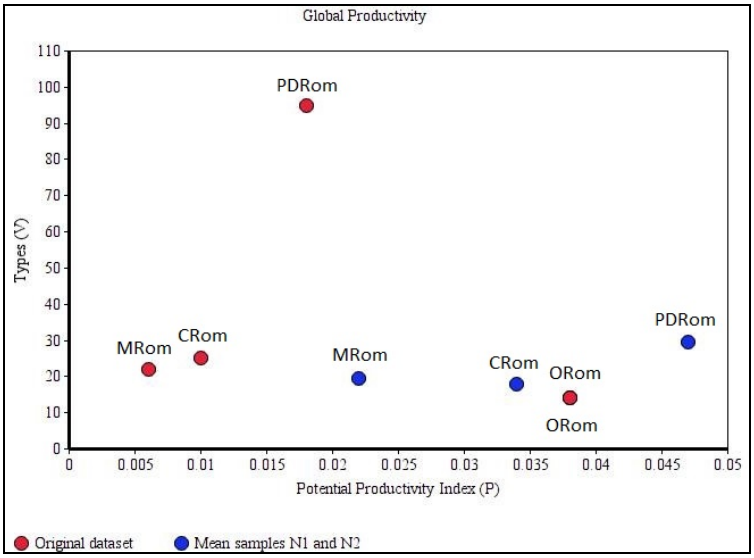

Figure 4. Global productivity (P*) for the original dataset and the mean of the samples.

Baayen and Lieber (1991) expect globally more productive processes to have large values for both V and P, whereas globally unproductive processes would have low values for both type frequency (V), and potential productivity (P). Figure 4 gives a complete image of the plotted global productivity (P*) for both the original dataset and the mean of the samples. As pointed out by the two scholars, the drawback of this measurement is that, when taking into account only V or only P, it is impossible to estimate which process is

more productive. In the original Romanian dataset (the red circles on the graph), the *mihi est* pattern in present-day language has a low potential productivity index on the X-axis, namely 0.018, but a very high number of types (V) on the Y-axis, namely 95. Correspondingly, in the representation of the mean of the two samples of 160 tokens each (the blue circles on the graph), the construction has scored a reasonably high mean potential productivity value, namely 0.047, as shown by its position on the right extremity of the X-axis, but a low number of types on the Y-axis, with a mean of 29.5. Even though one might be tempted to claim that both positions on the graph indicate a high degree of productivity, this conclusion is not fully validated through this measurement.

Returning to the estimated expanding productivity as presented in 4.2, its interpretation clearly depends on the results of potential and global productivity measurement (Baayen 1993: 194). The figures for expanding productivity shown in Table 3 are in line with the figures for potential and global productivity. In the Romanian data, the general tendency is towards an increase in expanding, potential, and global productivity throughout the centuries. In Baayen's terms, a category can be expanding at a higher or lower rate, which determines the degree of expanding productivity of the specific category. Similarly, a process can show a low or high risk of saturation, which determines its potential for expansion, and hence its degree of potential productivity.

To sum up, this section has shown that the integrated analysis of the three aspects of productivity confirms the increasing realized productivity of the Romanian *mihi est* pattern throughout the first four centuries, as well as its ability to continue to be productive by expanding especially in the last century, its already high potential for productivity.

4.4. A different approach to productivity

Baayen's (2009) measures were originally designed to operationalize the productivity of word formation processes and affixes. Nevertheless, these methods have been gradually transferred to syntactic and argument structure innovation processes (Zeldes 2012). Indeed, as well argued by Barðdal (2008: 30), there is a parallelism between morphological constructions, or words instantiated by their morphemes, and syntactic constructions, which, like argument structure constructions, are instantiated by the predicates occurring in them.

The perspective suggested by Barðdal (2008) is that the productivity of a construction is predictable based on its type frequency, its coherence, and an inverse correlation between the two. For a syntactic construction, such as the *mihi est* pattern, all predicates that can instantiate it make up its type frequency. The *coherence* of a construction is defined as the internal morphological, phonological, or semantic consistency found between all the members of each construction or category. In the specific situation of these kind of syntactic constructions, *semantic coherence* is considered the most relevant.

According to Barðdal (2008), productivity can be visualized in a graph with type frequency on the Y-axis and semantic coherence on the X-axis, as in Figure 5. The inverse correlation between these two is illustrated by a productivity cline with, at one end, the highest type frequency, and, at the other end, the highest degree of semantic coherence. The cline illustrates the scalar character of productivity, with the highest degree of productivity at the upper end of the cline, where it intersects with high type frequency, and the lowest degree of productivity at the lower end of the cline, closer to the highest semantic

coherence, where analogical formations are possible. In other words, full productivity and analogical formations are situated at the opposite ends of the productivity cline, representing two different sides of the same coin (Barðdal 2008: 3). The scholar points out that there are no extensions of non-productive processes, but only different levels of schematicity, influenced by differences in type frequency and degree of entrenchment of the schema.

Hence, the higher the type frequency of a construction, the lower the degree of semantic coherence is required for a construction to be productive. Conversely, the lower the type frequency of a construction, the higher degree of semantic coherence is necessary for a construction to be productive.

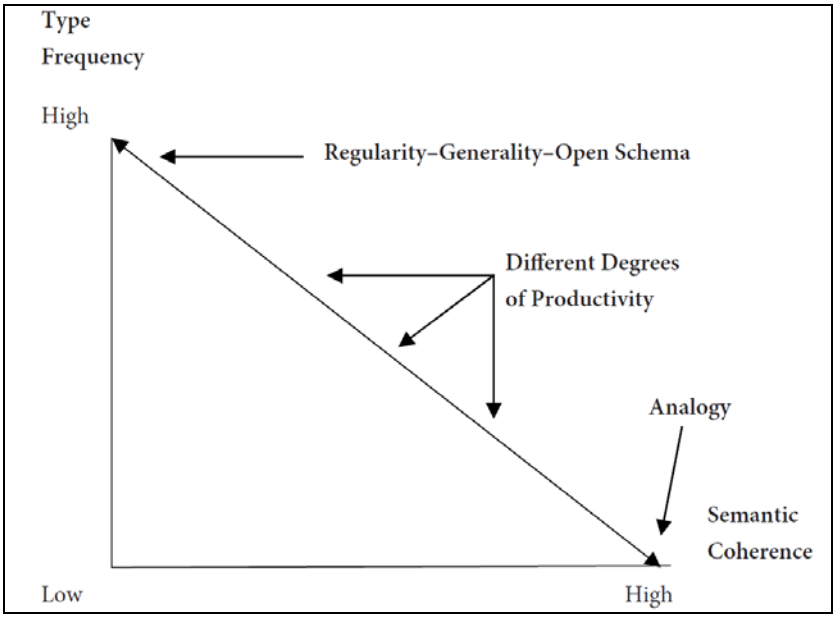

Figure 5. Different aspects of the cline of productivity (Barðdal 2008: 38).

In view of the discussion above, the Romanian *mihi est* pattern is to be situated on the lower extremity of the productivity cline, since its semantic coherence is fairly high. That is, the semantic domain of the predicates occurring in this type of construction is restricted to a limited number of semantic classes such as emotional, ontological, and bodily states (Barðdal 2004, Barðdal et al. 2012). This situation remains stable across the first three historical periods of Romanian. However, in present-day Romanian, the semantic coherence seems to start dissolving due to the high number of types entering this construction from various semantic domains, such as events, acts, time, phenomena, and personality traits, as visualized in Figure 6. A few cases of such new types are illustrated in examples (2–6).

(2) Îmi este *plecare* şi -mi este
me.DAT is departure and me.DAT is
şi *vânt.* (Event/Act)
too wind
'I feel like leaving and it feels like wind too.'
(http://reteaualiterara.ning.com/m/blogpost?id=1971741%3ABlogPost%3A37656)

(3) mi- e *joc* *de* *copii* (Act)
me.DAT is game_ of kids
'I feel like playing as a child.'
(http://wwwalexsmallthings-desprenimic.blogspot.com/2014/05/?m=0)

(4) Mi- e *iarnă* -n sufletu- mi firav şi -mi
me.DAT is winter in soul.the mine frail and me.DAT
ninge blând (Time)
snow.3SG gently
'I feel like winter in my frail soul and I feel like gentle snow.'
(http://alinacristian.blogspot.com/2015/06/mi-e-inima-plina-de-tine.html)

(5) Îmi este *durere* şi -mi este
me.DAT is pain and me.DAT is
furtună (Meteorological phenomena)
storm
'I feel pain and I feel stormy (in my soul).'
(http://reteaualiterara.ning.com/m/blogpost?id=1971741%3ABlogPost%3A37656)

(6) Mi- e *căldură* şi *blândeţe* (Personality traits)
me.DAT is warmth and gentleness
'I feel warmth and gentleness.'
(http://www.catchy.ro/vacanta-in-rai/66720/comment-page-1)

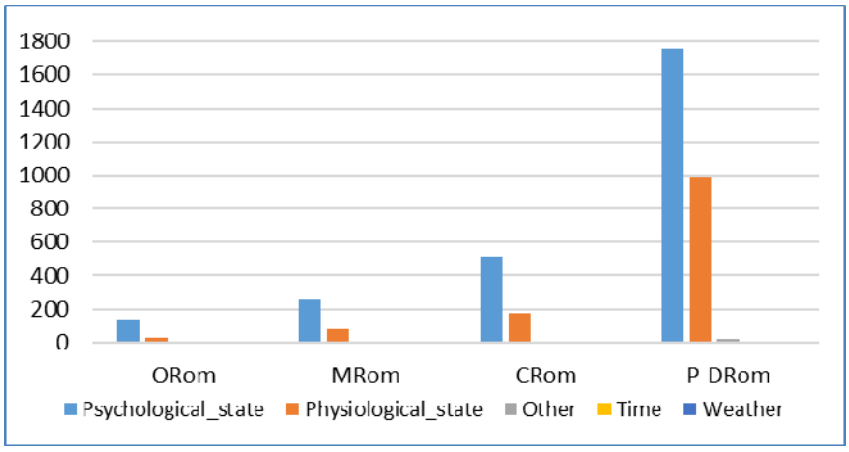

Figure 6. Semantic fields per period.

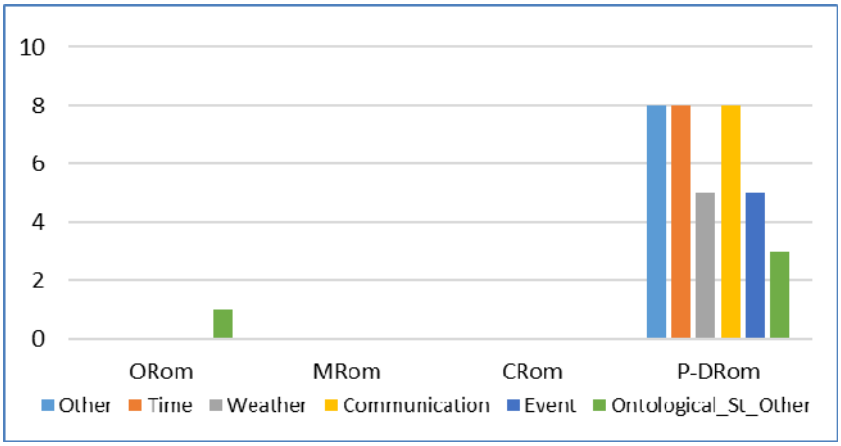

Figure 7. *Zoom in the less frequent semantic classes.*

Figure 7, which zooms in on the less frequent semantic classes, shows the repartition of various semantic fields throughout the four historical periods of Romanian. Note that the semantic class of psychological states remains the predominant one, gaining new types with each historical period. Moreover, the new types are all in metaphorical use and still express emotional states, which shows that the construction has not extended its meaning and, hence, is not to be positioned higher on the productivity cline.

5. CONCLUSIONS

In this paper, a number of attempts to measure the degree of productivity of the *mihi est* pattern were applied to a Romanian dataset. Two research questions have been addressed: (i) how has the degree of productivity of this construction evolved since the first documented Romanian texts?; (ii) is the *mihi est* pattern expanding or retracting in productivity in present-day Romanian?

The approach used in this paper is based on Baayen's (2009) distinction between three aspects of productivity: realized, expanding, and potential productivity. Realized productivity was estimated by applying a test based on type frequency, as proposed by Anshen and Aronoff (1989). According to this test, the pattern increases in potential productivity during the first three historical periods, but decreases considerably in present-day Romanian. However, expanding productivity, estimated based on the *hapax-conditioned degree of productivity* method (Baayen 1993) shows an increasing tendency throughout the centuries. Similarly, the potential productivity test, estimated by the two methods proposed by Baayen and Lieber (1991), the *category-conditioned degree of productivity* and global productivity, shows an increasing tendency of the productivity degree throughout all the historical periods.

These results are pertinent to the first research question, and show that there is a constant increase in productivity during the first three periods of Romanian, which is, however, not clearly confirmed in present-day Romanian in terms of realized productivity.

These contrasting results confirm the necessity for a multidimensional approach relying on quantitative, but also qualitative tests, especially when the productivity of a pattern is studied from a diachronic perspective.

As for the second research question, *i.e.* whether the *mihi est* pattern is expanding or retracting in productivity in Present-Day Romanian, the results of the quantitative tests show contrastive results. A very interesting picture emerges upon qualitative analysis of the data. After three periods of stability, in which only nouns expressing physiological or psychological states are found in the *mihi est* pattern, the situation changes in present-day Romanian, where the construction allows original combinations with nouns coming from different semantic fields such as events, acts, time, meteorological phenomena, and personality traits. Nevertheless, these nouns are used metaphorically, so that the construction continues to express physiological or psychological states. This finding confirms the productivity of the construction, which can coerce predicates from other fields into a psychological interpretation, due to its high semantic coherence.

REFERENCES

Anshen, F., M. Aronoff, 1989, "Morphological productivity, word frequency and the Oxford English Dictionary", *Language Change and Variation,* Amsterdam, Benjamins, 197–202.

Anshen, F., M. Aronoff, 1999, "Using dictionaries to study the mental lexicon", *Brain and Language,* 68, 16–26.

Aronoff, M., 1976, *Word formation in generative grammar*, Linguistic Inquiry Monographs Cambridge, Mass., MIT Press, 1, 1–134.

Baayen, R. H., 1992, "Quantitative aspects of morphological productivity", in: G. Booij, J. van Marle (eds), *Yearbook of Morphology*, 1991, Dordrecht, Kluwer, 109–149.

Baayen, R. H., 1993, "On frequency, transparency and productivity", *Yearbook of morphology,* Berlin, Springer, 181–208.

Baayen, R. H., R. Lieber, 1991, "Productivity and English derivation: a corpus-based study" *Linguistics*, 29, 5, 801–844.

Baayen, R. H., 2001, *Word frequency distributions*, 18, Springer Science and Business Media.

Baayen, R. H., 2009, "Corpus linguistics in morphology: morphological productivity", 43, *Corpus Linguistics. An International Handbook*, 900–919.

Barðdal, J., 2001, *Case in Icelandic-A synchronic, diachronic and comparative approach*, PhD Thesis, Lund University.

Barðdal, J., 2006, "Predicting the productivity of argument structure constructions", *Annual Meeting of the Berkeley Linguistics Society*, 32, 467–478.

Barðdal, J., 2008, *Productivity: Evidence from case and argument structure in Icelandic*, 8, Amsterdam, John Benjamins Publishing.

Bauer, L., 2001, *Morphological productivity*, 95, Cambridge, Cambridge University Press.

Bolozky, S., 1999, *Measuring productivity in word formation: The case of Israeli Hebrew*, 27, Leiden, Brill.

Bybee, J., 1995, "Regular morphology and the lexicon", *Language and Cognitive Processes,* 10, 5, 425–455.

Bybee, J., S. Thompson, 1997, "Three frequency effects in syntax", *Annual Meeting of the Berkeley Linguistics Society*, 23, 378–388.

Clausner, T. C., 2002, "How conceptual metaphors are productive of spatial-graphical expressions", *Proceedings of the Annual Meeting of the Cognitive Science Society*, 24, 208–213.

Clausner, T. C., W. Croft, 1997, "Productivity and schematicity in metaphors", *Cognitive Science*, 21, 3, 247–282.
Corbin, D., 1987, *Morphologie dérivationnelle et structuration du lexique*, 1, Max Niemeyer.
Croft, W., D. A. Cruse, 2004, *Cognitive linguistics*, Cambridge, Cambridge University Press.
Gaeta, L., D. Ricca, 2005, *Aspetti quantitativi della produttività morfologica*, Aracne.
Gheție, I., 1997, *Istoria limbii române literare. Epoca veche (1532–1780)*, Bucureşti, Editura Academiei.
Goldberg, A. E., 1995, *Constructions: A construction grammar approach to argument structure*, Chicago, University of Chicago Press.
Goldberg, A. E., 2006, *Constructions at work: The nature of generalization in language*. Oxford, Oxford University Press.
Gyselinck, E., 2018, *The constructional network as a dynamic system: the role of expressivity and productivity in (re)shaping the network,* PhD Thesis, Ghent.
Gyselinck, E., T. Colleman, 2016, "Tracking shifts in the literal versus the intensifying fake reflexive resultative construction: The development of intensifying *dood* 'dead' in 19th–20th Century Dutch", *Belgian Journal of Linguistics*, 30, 55–90. https://doi.org/10.1075/bjl.30.04gys
Pană Dindelegan, G., M. Maiden (eds), 2016, *The Syntax of Old Romanian*, 19, Oxford, Oxford University Press.
Plag, I., 2003, *Word-formation in English*, Cambridge, Cambridge University Press.
Zeldes, A., 2012, *Productivity in argument selection: From morphology to syntax*, 260, Berlin, Walter De Gruyter.