

---

# General Extenders: From Interaction to Model

PETER GRZYBEK (Graz)

DARINKA VERDONIK (Maribor)

## Abstract

Based on the assumption that general extenders represent a separate category in the linguistic and phraseological system of a given language, the present study attempts to show that their frequency occurrence is regular and law-like, as the result of a diversification process. Based on an empirical analysis of the Slovene GOS corpus, a specific discrete probability distribution and its continuous counterpart, the Zipf-Alkseev model, is presented as an adequate and interpretable model.

## 1 INTRODUCTION

In this contribution, we will represent our research and some new ideas with regard to a group of pragmatic markers in spoken interaction. These linguistic units have recently been summarized under the term *general extenders* (GEs); other terms for these set of expressions are, among others: *set marking tags*, *utterance final tags*, *clause terminal tags*, *extension particles*, *generalized list completers*, *generalizers*, *final coordination tags*. In English, this group of expressions includes examples such as *and stuff*, *and everything*, *or something*, *or anything*, and others; in German, we have corresponding expressions such as *und so*, *und so weiter*, *und so weiter und so fort*, *und solche Sachen*, *und alles Mögliche*, *oder so*, *oder so was*, and many others. In speech, such expressions are used in verbal contexts such as:

- (1) *Ich hab' nun jetzt erstmal meine ganzen Pflanzen da in die Erde gebracht und – und – sehr viel Tulpen und Krokusse und so was gesteckt.*
- (2) *I've just got all my plants there in the ground and – and – lots of tulips and crocuses and so was put in.*

(Overstreet 2005: 1849)

Thus far, GEs have not been dealt with systematically in the field of phraseology, and they have been dealt with much more broadly in various branches of linguistics (see below). Yet, from a phraseological point of view, GEs can

be considered to be a special kind of pragmatic phrases<sup>1</sup> (Burger et al. 1982: 110ff., Fleischer 1982: 133f., Jakop 2005) since they are multi-word expressions functioning as one unit being stored and retrieved from memory at the time of use, on the one hand, and because they have no “ideational” meaning on their own, thus being of a pragmatic rather than semantic kind, not contributing to the “propositional content” of the utterance, on the other. As a consequence, the term *pragmatic phrase* seems to be preferable as compared to *pragmatic idiom*, an argument fully in line with Fleischer’s (1982: 132) reasoning that such units lack (full) idiomaticity, and also in accordance with more recent distinctions between phrasemes in a narrow sense (including idiomaticity) and phrasemes in a broader sense (Burger 1998: 32). In fact, the attribution of GEs to pragmatic phrasemes can, from a phraseological perspective, be traced back, among others, to Burger et al. (1982), where the authors list examples such as *und so weiter* (ibid., 126) as speech-specific pragmatic phrases. Likewise, Prodromou (2008: 117) makes a comparable categorization for two-word pragmatic markers such as *you know* and *I mean*, which he calls *pragmatic phrases*, in order to capture the formulaic nature of these discourse markers.

Given the relatively sparse treatment of GEs in the field of phraseology, this contribution will take the following course: Subsequent to a discussion of general extenders from a linguistic point of view (2.1), we will focus on hitherto almost ignored material from Slovene (2.2.) and argue in favor of the notion that GEs represent a specific linguistic and phraseological category in its own right (2.3.) which, as a consequence, follows well-known regularities as to their frequency behavior; in this respect, we will analyze corpus material from the Slovene GOS corpus, attempting to show that the frequencies of GEs are not arbitrarily or chaotically organized, but follow a well-defined frequency distribution model, as a result of what we consider to be a diversification process.

## 2 GENERAL EXTENDERS

### 2.1 General extenders in linguistics

In the field of linguistics, Dines (1980) has been one of the first to pay special attention to GEs, using a variationist framework to study their use in Australian English; in her terminology, expressions of this kind are called *clause terminal tags* and *set-marking tags*. Later, Dubois (1992) used term *extension particles*: she analyzed the socio-demographic factors of their use, carried out

---

<sup>1</sup> German: *pragmatische Phraseologismen*, Slovene: *pragmatični frazemi*.

a distributional analysis of certain expressions which are component parts of extension particles, and conducted an analysis of the conditioning of distinct categories of extension particles by linguistic and social factors. Finally, the term *general extender* was suggested by Overstreet (1999); it is the most-widely term used still today (cf., e.g.: Cheshire 2007; Fernandez/Yuldashev 2011; Martinez 2011; Tagliamonte/Denis 2010).

General extenders (GEs) have been discussed in different theoretical frameworks, and from different perspectives, and it seems worth to initially mention some major studies, in order to demonstrate the broad spectrum of approaches within the field of linguistics:

- (a) Overstreet has discussed them within the functional grammar framework of ideational and interpersonal function; for the analytic approach she follows discourse analysis, emphasizing the socio-cultural perspective, and claiming that “within their actual context of occurrence, GEs appear to have function that is primarily interpersonal and tied to the nature of the social relationship of the participants” (1999: 145).
- (b) Subsequent to Overstreet’s work, Aijmer (2002) has treated GEs as particles with vague reference; her study is empirical, using corpus methods, in combination with a mixed approach in its analytic parts, including a description with regard to form and function (claiming they have textual and interpersonal functions), prosody and situations in which they are used.
- (c) Yucker et al. (2003) have included GEs in their study as one of the elements of vagueness in conversation.
- (d) Overstreet (2005) has compared the use of GEs in English and German conversations between adults.
- (e) Cheshire (2007) has analyzed GEs in the speech of adolescents from three English towns, comparing the use of different forms, grammaticalization processes and the multifunctionality of GEs.
- (f) Cucchi (2007) has compared the use of the GE *and so on* by native and non-native speakers of English, based on the corpus of EU parliamentary debates.
- (g) Tagliamonte and Denis (2010) have examined GEs in the English spoken in Toronto, using quantitative techniques and investigating their phonetic reduction, de-categorization, semantic change and pragmatic shift.
- (h) Palacios Martinez (2011) has analyzed the use of GEs in the speech of British teenagers and adults, comparing the frequency of use, grammaticalization and pragmatic functions.

- (i) Terraschke (2010) has explored the use of the English GE *or so* by native speakers in New Zealand and by German non-native speakers of English, comparing frequency and functions of use.
- (j) Pichler and Levey (2010) have presented a quantitative analysis of the co-occurrence of GEs (e. g., *and stuff, or something like that*) with other discourse features (e. g., *like, you know*) in a corpus collected in North-East England, concentrating on three age groups.
- (k) Fernandez and Yuldashev (2011) have analyzed the use of GEs in instant messages, comparing their use (forms, frequency, set of reference, functions) by native and non-native English speakers.

As can be seen from the list above, most research on GEs has been done in English language; yet, this category is of course present in many other (if not most) languages. However, for languages other than English, GEs have been dealt with to a lesser degree, and their representation and discussion is far more limited for languages other than English; nevertheless, we find discussions concerning Japanese (Wiezbicka 1991, Honda, 1996), Montreal French (Dubois 1992), Swedish (Winter/Norby 2000), German (Overstreet 2005), Spanish (Cortés/Rodríguez 2006), Persian (Parvaresh et al. 2010), or Lithuanian (Ruzaitė 2010), to mention but a few, and there seems reason to assume that we are concerned with a broadly (if not or even universally) distributed category.

As to the status of this category in the field of linguistics and phraseology, it seems worthwhile noting that, although GEs, on the one hand, seem to represent a specific class of linguistic elements in its own right, with their own specific communicative and pragmatic functions in discourse, and that this class is represented by different kinds of items, or even subcategories, on the other.

Concentrating on GEs as a specific subcategory of the linguistic or phraseological system, we will not deal with details of such further distinctions or sub-categorizations here; let it therefore suffice to mention merely the most important and most general sub-categorization of GEs, with regard to their structure and function, i. e., the distinction of two major groups, which are termed (a) adjunctive and (b) disjunctive: those GEs beginning with *and* in English, or with *und* in German, fall into one category which is called adjunctive, and those beginning with *or* (English) or *oder* (German), belong to the second class of expressions, called disjunctive.<sup>2</sup> Both groups differ not only in structure, but also in (textual) function: according to Aijmer (2002), adjunctive GEs have expanding and illustrative function, whereas disjunctive GEs have

---

<sup>2</sup> In Slovene – which will be discussed in more detail further below in Section 3 – we have analogical expressions, such as, for example, *in* or *pa*, on the one hand, and *ali*.

an approximation function; similar differences are reported by Overstreet (2005: 1855), for example, who summarizes that “the primary function of adjunctive general extenders is to indicate ‘there is more’”, while the primary function of disjunctive GEs is “tied to indicating potential alternatives, and hence hedging on what has been said”. Besides the textual functions mentioned above, there is common acknowledgment that all GEs (adjunctive as well as disjunctive) perform interpersonal functions as well:<sup>3</sup> based on the assumption that GEs are used to indicate that there is, in fact, more to be said than is said explicitly, GEs are seen to express the assumption of shared knowledge and experience between speaker and listener(s), to represent an appeal for solidarity and understanding, to indicate a lack of certainty, etc.

As has been said above, in this contribution we do not intend to go into further details of describing the GEs discourse functions: accepting that differences in usage and subdivisions exist, the category of GEs thus consisting of heterogeneous elements, we focus instead on GEs as a distinct class of linguistic elements as a whole, the members of which we consider to be the result of some diversification process (see below).<sup>4</sup> In order to attempt to provide plausible arguments in favor of this assumption, we will base our analyses on Slovenian material, which shall be presented in the next subsection.

### 3 GENERAL EXTENDERS IN SLOVENE: DATA

As has been mentioned above, GEs have received only sporadic attention in Slovene (Verdonik/Kačič 2012). As a consequence, it seems reasonable, in a first step, to provide some relevant information about this kind of expressions

---

<sup>3</sup> Overstreet (1999) even claims that such interpersonal functions are primary to general extenders; as compared to this, Cheshire (2007) postulates one should consider their functions within the local context in which they are used, that they are multifunctional, and that we cannot define a principal function – even more, it would be counterproductive to do so.

<sup>4</sup> The idea to see GEs as a distinct class of linguistic and/or phraseological elements is of course not new; Overstreet (1999: 6), for example, starts her monograph on GEs with the words: “They represent a distinct set of linguistic elements /.../”. However, in the further course of her book, she mentions the possibility that “it might be possible to describe general extenders as types of ‘discourse markers’” (ibid.: 12). In fact, GEs and discourse markers have some common characteristics, but there are important differences, too; as a consequence, it has been rather a matter of perspective to either group GEs into one common class with discourse markers, or to see both as members of some larger class of forms, e. g., ‘pragmatic operators’. For details on this discussion see, among others: Dubois (1992), Aijmer (2002), or Martinez (2011).

which, in Slovene, is represented by items such as, e. g., *in tako naprej* ‘and so on’, *pa to* ‘and so’, *in podobno* ‘and similar’, *pa tako* ‘and so’, *ali pa kaj takega* ‘or something like that’ – they all fall into the GE category, cf. the following examples:

- (1) *se boste o otrocih pogovarjale **pa tko***  
‘you will talk about kids **and so**’
- (2) *mama je umrla in mislim z bratom sva drugač zmenena **in tko naprej a ne on** y’know’*  
‘my mother died and I mean me and my brother have different arrangements **and so on** y’know’
- (3) *recimo ko gremo v savno **al pa kaj takšnega***  
‘for example when we go to sauna **or something like that**’

The data used for our analysis are taken from the GOS (*GOvorjena Slovenščina*) corpus, the reference corpus of Slovene speech (Verdonik/Zwitter Vitez 2011; available also at <[www.korpus-gos.net](http://www.korpus-gos.net)>). This corpus consists of 1,032,775 words, or 120 hours of recordings. The GOS corpus contains speech events from five different discourse types with different channels, as shown in Table 1. An important characteristic of the corpus, which also ensures a good comparability of different discourse types in the corpus, is that the majority of the recordings include spontaneous speech (as opposed to read speech).

Discourse type	Channel	Number of tokens	Totals	Percentage
Classes		162,750	162,750	15.76 %
Media – informative	Radio	94,536	196,799	19.06 %
	TV	102,263		
Media – entertainment	Radio	123,152	228,765	22.15 %
	TV	105,613		
Official	Phone	33,484	153,471	14.86 %
	Personal communication	119,987		
Private	Phone	68,083	290,990	28.18 %
	Personal communication	222,907		

Table 1: Discourse types in the GOS corpus.

The data in the GOS corpus are available in two transcription formats: a pronunciation-based transcription and a standardized transcription. Pronun-

ciation-based transcription is an orthographic transcription which represents a more or less faithful account of acoustic forms of words; the standardized transcription follows the Slovene written standard and offers a common form for different pronunciation realizations of the same lexeme. The web-interface enables also listening the audio recording for each concordance.

In the GOS corpus, we found more than 50 different Slovene expressions functioning as general extenders; at closer sight it becomes evident, however, that many of them are merely different variations of one basic expression. On the whole, we identified four basic variants:

- (1) PRONUNCIATION VARIATION: Since the GOS corpus includes pronunciation-based transcriptions, many vocal reductions and other phoneme-based variations were evident, for example *tako* ‘so’ (occurring in the general extenders *in tako naprej* ‘and so on’, *in tako* ‘and so’, *ali pa tako* ‘or so’, etc.) has variations like: *tko*, *tk*, *tak*, *teku*, etc.
- (2) GRAMMATICAL VARIATION: Slovene is a highly inflectional language. A common variation in general extenders includes the opposition of nominative form (e. g., *in podobno* ‘and similar’, *in vse te stvari* ‘and all that things’) vs. genitive form (e. g., *in podobnega* ‘and similar’, *in vseh teh stvari* ‘and all that things’).
- (3) SYNONYM VARIATION: Slovene general extenders vary considerably by including different synonyms; common variations are: synonym conjunctions *in* ‘and’, which is standard (*in to* ‘and that’), vs. *pa* ‘and’, which is colloquial (*pa to* ‘and that’); synonym or semantically close pronouns *ta* ‘this’ (*in te stvari* ‘and these things’) vs. *tak* ‘such’ (*in take stvari* ‘and such things’) vs. *takšen* ‘such’ (*in takšne stvari* ‘and such things’); synonym nouns *stvar* ‘thing’ (*in te stvari* ‘and these things’) vs. *zadeva* ‘thing’ (*in take zadeve* ‘and such things’) vs. *reč* ‘thing’ (*pa take reči* ‘and such things’); synonym adverbs *naprej* ‘on’ (*in tako naprej* ‘and so on’) vs. *dalje* ‘forth’ (*in tako dalje* ‘and so forth’).
- (4) OPTIONAL ITEM ADDITION: Most disjunctive general extenders in Slovene have the optional particle *pa* after the beginning conjunction *ali* ‘or’, e. g., *ali karkoli* ‘or whatever’ vs. *ali pa karkoli* ‘or whatever’ (there is no equivalent in English). Some general extenders also vary in length, i. e., they have a basic form (e. g., *pa vse* ‘and all’) which can be prolonged by one item (e. g., *pa vse to* ‘and all that’) or even more items (e. g., *pa vse to skupaj* ‘and all that together’); other examples would be *pa vse* ‘and all’ vs. *pa vse skupaj* ‘and all together’ vs. *pa vse to skupaj* ‘and all that together’; *ali karkoli* ‘or whatever’ vs. *ali karkoli drugega* ‘or whatever else’; *ali pa kaj* ‘or something’ vs. *ali pa kaj takega* ‘or something like that’.

We considered that keeping all different expressions – which sum up to more than 50 – would blur the common picture of general extenders frequencies, as there would be a substantially bigger gap among some expressions (e. g., between *in tako naprej* ‘and so on’ vs. *pa to* ‘and that’) than among others (e. g., *in tako naprej* ‘and so on’ vs. *pa tako naprej* ‘and so on’). Therefore we decided to group expressions into basic groups of general extenders, disregarding the variation described above, as a result obtaining 14 groups of general extenders. These are represented in the Table 2, along with their frequencies<sup>5</sup> of usage in the GOS corpus.

GE group	English translation	Variations	Frequency
<i>in tako naprej</i>	and so on	in/pa tako naprej/dalje	286
<i>in podobno</i>	and similar	in podobne zadeve/reči/dalje	27
<i>pa to</i>	and that	pa/in to/tega (vse)	342
<i>in te stvari</i>	and such things	in/pa te/teh/take/takšne stvari/zadeve/reči	35
<i>in vse stvari</i>	and all things	in/pa vse/vseh (te) stvari/reči	4
<i>pa vse (to skupaj)</i>	and all (that together)	pa/in vse/vsega (to/tega) (skupaj)	96
<i>pa tako</i>	and so	pa/in tako	155
<i>ali pa kaj (takega)</i>	or something like that	ali (pa) kaj (takega/takšnega) (v tem smislu)	94
<i>ali pa kaj jaz vem</i>	or I don't know	–	4
<i>ali nekaj takega</i>	or something like that	ali (pa) nekaj takega/takšnega	39
<i>ali kaj podobnega</i>	or similar	ali (pa) kaj/česa podobnega	19
<i>ali karkoli (že/takega)</i>	or whatever (similar)	ali (pa) karkoli/česarkoli (že/ drugEGA/takega/pač)	32
<i>ali pa to</i>	or that	–	5
<i>ali pa tako</i>	or so	–	10
<b>TOTAL</b>			<b>1148</b>

Table 2: Slovene general extenders in the GOS corpus.

<sup>5</sup> It should be noted that most of the phrases had to be disambiguated because they can function either as a general extender (e. g., *vrtna opravila pa to* ‘garden work and so’) or not (e. g., ordinary conjunction + pronoun: *potem bo pa to pomenilo* ‘than this will mean’).



Yet, in this contribution, we intend to go one step further, not restricting our objective to the mere description of Slovene GEs, their functions, forms and usage. Rather, we will extend our interest to theoretical issues, regarding GEs to represent a specific linguistic sub-system in its own right, characterized by a common (pragmatic) function. In this context, locating them on the phraseological level of language, we assume GEs to underlie a process of diversification; as a result, the frequency with which each member of this class (i. e., each individual GE) occurs, should not be arbitrary or chaotically organized, but rather in a law-like manner, each individual GE as well as its frequency of occurrence thus being the result of a specific diversification process. If this is true, we will thus, with our contribution, not only enrich hitherto research by adding new Slovene material, but substantially expand their theoretical treatment from a methodological point of view, seeing GEs from a synergetic perspective.

## **4 MODELING GENERAL EXTENDERS IN TERMS OF A DIVERSIFICATION PROCESS**

### **4.1 Diversification**

Diversification is a well-known process characterizing all living and dynamic systems: in biology, for example, there would be no variation in organic nature without diversification. In linguistics, diversification processes are well-known, too, mainly in the field of grammar and semantics. Here they are assumed to take place when the attribute space of an entity expands in one or more dimensions, e. g., when a morpheme gets enriched by new allomorphs, when a word gets enriched by new meanings, etc. Since over-diversification in language would not be economic for the communicative system as a whole, satisfying equally well producers' and receivers' interests, diversification processes must necessarily be counter-balanced by processes of unification, the whole system thus turning out to be in a state of dynamic equilibrium.

### **4.2 Diversification in linguistics**

In trying to model the observed frequencies, one of our basic assumptions is, as has been said above, that general extenders “represent a distinct class of linguistic elements” (Overstreet 1999: 3). Along with this assumption goes the hypothesis, well-known from the field of quantitative linguistics, that frequencies with which different forms of a linguistic category occur, are regularly

distributed. In other words: not only the lexical inventory (of a given text or corpus), but also specific sub-categories are expected to follow regularities as to their frequency behavior. This can be seen in context of, or as a result of, diversification processes: just like in biology (or bionics) the rise of new species is a result of diversification, or as the introduction of new products into the market can be seen as diversification process in economics, diversification is a crucial process in linguistics (Altmann 1991, 2005). Here, diversification can be seen as a process of enlarging the number of forms or meanings of a given linguistic entity. In this sense, diversification processes have repeatedly been described with regard to various linguistic levels:

- (a) *Paradigmatic*: e. g. the rise of cases, numbers, tenses, etc.,
- (b) *Phono-morphemic*: e. g. the rise of allophones, allomorphs etc.,
- (c) *Geographical*: e. g. the increase in the number of different expressions of a concept,
- (d) *Social*: e. g. the rise of different words or meanings of a word or different pronunciations,
- (e) *Idiolectal*: within a community,
- (f) *Semantic*: e. g. the increase in synonymy and polysemy,
- (g) *Contextual*: e. g. the increase in the usage of a unit in different contexts.

Given these observations, and based on the assumption that every linguistic entity diversifies, that is, it generates variants and secondary forms, we set up the hypothesis that the frequency distribution of general extenders as a specific subcategory of language is the result of a diversification process, and that, as a result of this, general extenders are regularly distributed.

#### **4.3 Linguistic law of diversification**

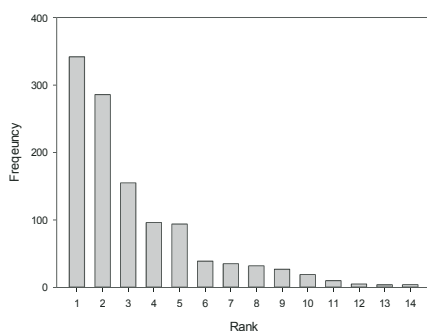
Generally speaking, a frequency distribution is based on the individual class members' frequency of occurrence. With regard to diversification processes, we are concerned with the more specific hypothesis that the diversifying linguistic entity under study abides by a ranking law, resulting in a specific rank-frequency distribution: if the members of the diversified entity are ordered according to their frequency, then the frequencies are "lawfully" connected. It goes without saying that rank-frequency distributions as functions expressing the decrease of frequencies ranked according to their magnitude, there are, *eo ipso*, no bell-shaped frequency distributions, which are rather of a left-skewed form.

Table 3 represents the data presented above (cf. Table 2), transformed in a rank-frequency distribution: in the first column, the rank ( $x$ ) is given, in the second column, the frequency  $f(x)$  of the corresponding entities.

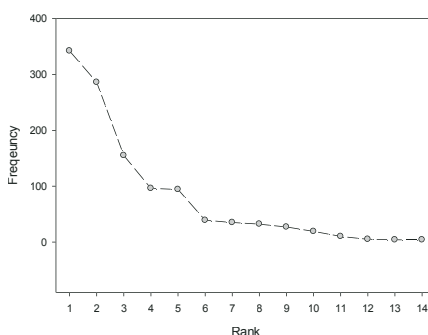
$x$	$f(x)$
1	342
2	286
3	155
4	96
5	94
6	39
7	35
8	32
9	27
10	19
11	10
12	5
13	4
14	4

Table 3: Rank frequencies of Slovene general extenders in the GOS corpus.

The rank-frequency distribution is graphically illustrated in Figures 1a and 1b: whereas Figure 1a presents them in the form of a bar chart, usual for the representation of discrete frequency distributions, Figure 1b presents them in the form of a line plot, usually preferred for the representation of continuous data and functions.



1a: Discrete bar chart



1b: Continuous line plot

Figure 1: Frequencies of general extenders in the Slovene GOS Corpus.

Given the observed frequencies, the next step will include their theoretical modeling.

#### 4.3 Modeling the frequencies of general extenders in a diversification framework

An attempt to model the frequencies illustrated in Figures 1a and 1b implies the search for an adequate mathematical function, i. e., a discrete probability distribution and/or a continuous function. As to a theoretical derivation of such a model, it seems first important, albeit trivial, to take account of the fact that, if the above hypothesis holds, the frequencies of elements of the given linguistic class are not distributed uniformly. Because the entities are ranked, and because of the corollary, it is true that for the probabilities of classes it holds that  $P_x \leq P_{x-1}$  – here,  $P_x$  is the frequency of the elements of a given class, and  $P_{x-1}$  is the frequency of the elements of its preceding class. Moreover, since  $P_x$  and  $P_{x-1}$  ( $x = 2, 3, \dots$ ) are related in a law-like manner, we can understand  $P_x$  to be a function of  $P_{x-1}$ . In mathematical terms, we can thus write

$$P_x = g(x)P_{x-1} \quad (1)$$

with  $g(x) \leq 1$ , since we have a monotonously decreasing rank order. Furthermore, attempting to find such a model, it seems reasonable to take into consideration the fact that every diversification process evokes a unification process operating on the same entity and working against the total decay of the phenomenon. This idea goes back to George K. Zipf's (1935, 1949) ideas about the antagonistic economy of producer and recipient in communication processes, and they are well-known today by the name of Zipfian forces in quantitative linguistics. As a result, assuming  $g(x)$  to represent the Zipfian forces of unification and diversification,  $g(x)$  can be set up as

$$g(x) = \frac{f(x)}{h(x)}, \quad (2)$$

where  $f(x)$  represents the diversification process, and  $h(x)$  the unification component, i. e., the controlling, regulating effect of the communication community. Moreover,  $f(x)$  can be understood to be a function composed of a language constant (e. g.,  $a$ ), on the one hand, plus the diversifying force of the producer (e. g.,  $b \cdot x$ ), on the other. With  $h(x) = c \cdot x$ , we would thus obtain

$$g(x) = \frac{a + bx}{cx}, \quad (3)$$

so that, according to equation (1)

$$P_x = \frac{a + bx}{cx} P_{x-1}, \quad (4)$$

which would, after re-parametrization, result in the well-known negative binomial distribution, in its zero-truncated (positive) form. This model has repeatedly been applied to model diversification processes in linguistics; there is no need to go into further details here, the more since in our case, i. e., with regard to general extenders, it seems more adequate to set up  $f(x)$  differently, with  $f(x) = a + b \cdot \ln(x)$ . This results in the Zipf-Alekseev distribution

$$P_x = C \cdot x^{-(a+b \ln x)} P_{x-1} \quad x = 1, 2, 3, \dots \quad (5)$$

with  $C^{-1} = \sum_{j=1}^{\infty} j^{-(a+b \ln j)}$  as the normalizing constant.

Analogically, one obtains the continuous Zipf-Alekseev function

$$y = f(x) = C \cdot x^{-(a+b \ln x)}, \quad (6)$$

for which the requirement that the sum of all probabilities  $\sum P = 1$  needs not be fulfilled, as is the case in the discrete model (5).

This Zipf-Alekseev model, too, has been proven to be adequate in the modeling of linguistic diversification (cf. Altmann 2005). In this context, it has also been theoretically derived, with reference to either the psychophysical Weber-Fechner law, based on the perceptibility of minimal differences between classes (cf. Hammerl 1991), or to the well-known Menzerath-Altmann law from linguistics, concerning the construct-constituent relation of linguistic units (cf. Hřebíček 1996, Altmann/Hřebíček 1996).

Applying this model to our data, in its discrete form, it seems reasonable to use the right-truncated version

$$P_x = C \cdot x^{-(a+b \ln x)} P_{x-1} \quad x = 1, 2, 3, \dots, n \quad (7)$$

with  $C^{-1} = \sum_{j=1}^n j^{-(a+b \ln j)}$  as the normalizing constant.

With parameter values  $a = -0.03$  and  $b = 0.61$ , the fit turns out to be good<sup>6</sup>, as indicated by the discrepancy coefficient  $C = 0.019$ . Figure 2 illustrates the result in graphical form.

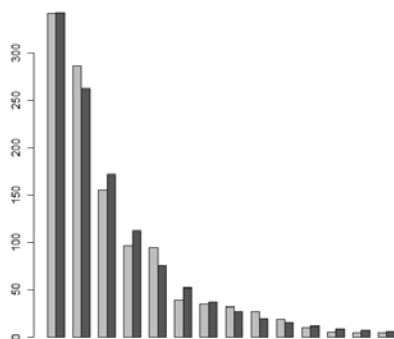


Figure 2: Observed and theoretical frequencies of general extenders in the Slovene GOS Corpus – right-truncated Zipf-Alekseev distribution.

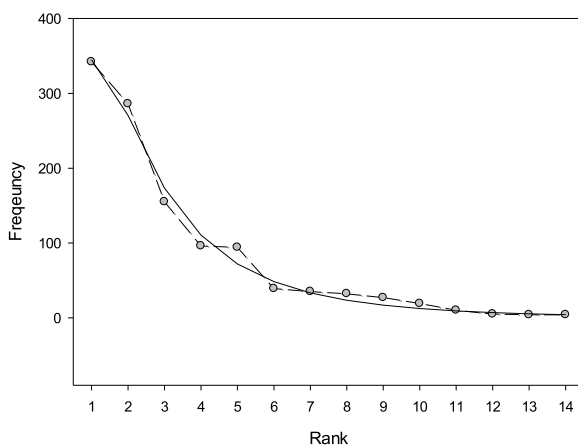


Figure 3: Observed and theoretical frequencies of general extenders in the Slovene GOS Corpus – continuous Zipf-Alekseev function (6).

<sup>6</sup> The usual goodness of fit test would be the well-known chi square test. Since the  $\chi^2$  value increases linearly with an increase of sample size, it tends to yield significant result the sooner, the larger the sample is. Since this is the standard case in linguistics, the discrepancy coefficient is preferred for larger samples, with  $C < 0.02$  being interpreted as a good,  $C < 0.01$  as a very good fit.

A similar good fit is obtained with the continuous function (6): here, the goodness of fit uses to be evaluated with reference to the determination coefficient  $R^2$ , which in our case, with parameter values  $C = 345$ ,  $a = 0.12$ , and  $b = -0.68$  is excellent ( $R^2 = 0.99$ ).<sup>7</sup> The result is graphically represented in Figure 3.

## 5 CONCLUSION

Given the theoretical discussions and empirical findings reported above, a number of conclusions seem to be justified, which shall be pointed out here:

- (1) General extenders are fixed (stereotypic) pragmatic expressions.
- (2) In a pragmatic framework, general extenders can be described as expressions with vague reference, used to indicate that there is more to say, to express assumption of shared knowledge, appealing for solidarity and understanding, or indicating lack of certainty.
- (3) General extenders are a distinct category of linguistic elements, which refer to preceding items, and which are known for their fairly homogeneous structure and fixed clause position.
- (4) As a distinct linguistic category, general extenders as a group lend themselves to processes of diversification (and, as a consequence, unification going along with it).
- (5) As a result of the diversification processes, general extenders occur with varying frequencies.
- (6) The frequencies of general extenders are regulated in a law-like manner, representing a process of self-regulation.
- (7) For Slovenian general extenders, the well-known Zipf-Alekseev model (both as a continuous function and as a discrete probability model) turns out to be adequate to theoretically model the general extenders' frequencies and explain the generating process behind it.
- (8) It will be interesting and important to do analogical analyses with linguistic material from other languages, in order to arrive at more general conclusions.

---

<sup>7</sup> Setting  $C = f_1 = 342$  (cf. Table 3) results in a reduction of the number of parameters to be estimated; in this case, and with a slightly different parameter value for  $a = 0.135$ , the result is practically unchanged ( $R^2 = 0.99$ ).

## REFERENCES

- AIJMER, Karin, 2002: *English discourse particles: Evidence from a Corpus*. Amsterdam/Philadelphia: John Benjamins.
- ALTMANN, Gabriel 2005: Diversification processes. Köhler, Reinhard/Altmann, Gabriel/Piotrowski, Rajmund G. (eds.): *Quantitative Linguistics. An International Handbook. Quantitative Linguistik. Ein Internationales Handbuch*. Berlin/New York: de Gruyter. 646–658.
- ALTMANN, Gabriel/HŘEBÍČEK, Luděk, 1996: Levels of Order. Schmidt, Peter (ed.): *Glottometrika 15. Issues in General Linguistic Theory and the Theory of Word Length*. Trier: wvt. 38–61.
- ARIEL, Mira, 1994: Pragmatic Operators. Asher, Ron E. (ed.): *Encyclopedia of Languages and Linguistics*, vol. 6. Oxford: Pergamon and Aberdeen University Press. 3250–3253.
- BLAKEMORE, Diane, 2002: *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*. Cambridge: Cambridge University Press.
- BURGER, Harald/BUHOFER, Annelies/SIALM, Ambros, 1982: *Handbuch der Phraseologie*. Berlin/New York: de Gruyter.
- BURGER, Harald, 1998: *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Berlin: Erich Schmidt Verlag.
- CHESHIRE, Jenny, 2007: Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics* 11/2, 155–193.
- CORTÉS RODRÍGUEZ, Luis, 2006: Los elementos de final de serie enumerativa del tipo ‘y todo eso, o cosas así, y tal etc. Perspectiva interactiva. *Boletín de Linguística XVIII* (26), 102–129.
- CUCCHI, Constanza A., 2007: *An investigation of general extenders in a corpus of EU parliamentary debates. Proceedings from the Corpus Linguistics Conference Series, University of Birmingham (UK), 27–30 July 2007*. <<http://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2007/242Paper.pdf>>.
- DINES, Elizabeth R., 1980: Variation in discourse – “and stuff like that?”. *Language in Society* 9/1, 13–31.
- DUBOIS, Sylvie, 1992: Extension particles, etc. *Language Variation and Change* 4, 179–203.
- FERNANDEZ, Julieta/YULADASHEV, Aziz, 2011: Variation in the use of general extenders and stuff in instant messaging interactions. *Journal of Pragmatics* 43, 2610–2626.
- FLEISCHER, Wolfgang, 1982: *Phraseologie der deutschen Gegenwartssprache*. Leipzig: VEB Bibliographisches Institut Leipzig.
- FRASER, Bruce, 1999: What are discourse markers? *Journal of Pragmatics* 31/7, 931–952.
- HAMMERL, Rolf, 1990: Untersuchungen zur Verteilung der Wortarten im Text. Hřebíček, Luděk (ed.): *Glottometrika 11*. Trier: wvt. 142–156.



HONDA, Akiko, 1996: Daigaku Koogi ni okeru kotoba no danjosa [= Gender differences in language use in university lectures]. *Kotoba* 16, 15.

HŘEBÍČEK, Luděk, 1996: Word Associations and Text. Schmidt, Peter (ed.): *Glottometrika 15. Issues in General Linguistic Theory and the Theory of Word Length*. Trier: wvt. 96–101.

JAKOP, Nataša, 2005: Opis pomena pragmatičnih frazemov s pragmatičnimi kategorijami. Kržišnik, Erika/Eismann, Wolfgang (eds.): *Frazeologija v jezikoslovju in drugih vedah / Phraseology in linguistics and other branches of science*. Ljubljana: Filozofska fakulteta. 167–182.

JUCKER, Andreas H./SMITH, Sara W./LÜDGE, Tanja, 2003: Interactive aspects of vagueness in conversation. *Journal of Pragmatics* 35, 1737–1769.

OVERSTREET, Maryann, 1999: *Whales, Candlelight and Stuff Like That. General Extenders in English Discourse*. Oxford: Oxford University Press.

OVERSTREET, Maryann, 2005: And stuff *und so*: Investigating pragmatic expressions in English and German. *Journal of Pragmatics* 37/11, 1845–1864.

PALACIOS MARTINEZ, Ignacio M., 2011: “*I might, I might go I mean it depends on money things and stuff*”. A preliminary analysis of general extenders in British teenagers’ discourse. *Journal of Pragmatics* 43/9, 2452–2470.

PARVARESH, Vahid/TAVANGAR, Manoochehr/RASEKH, Abbas Eslami, 2010: General extenders in Persian discourse: Frequency and grammatical distribution. *Cross-cultural Communication* 6/3, 18–35.

POUTSMA, Hendrik, 1916: *A Grammar of Late Modern English*. Groningen, NL: P. Noordhoff.

PRODROMOU, Luke, 2008: *English as a Lingua Franca: A Corpus-based Analysis*. New York/London: Continuum.

RUZAITĚ, Jūratė, 2010: Translation equivalents of vague language items: A study of general extenders in parallel corpus. *Kalby studijos / Studies about Languages* 16, 33–38.

SCHIFFRIN, Deborah, 1987: *Discourse Markers*. Cambridge: Cambridge University Press.

SCHOURUP, Lawrence, 1999: Discourse markers. *Lingua* 107, 227–265.

TAGLIAMONTE, Sali A./DENIS, Derek, 2010: The stuff of change: General extenders in Toronto, Canada. *Journal of English Linguistics* 38/4, 335–368.

TERRASCHKE, Agnes, 2010: *Or so, oder so, and stuff like that* – general extenders in New Zealand English, German and in learner language. *Intercultural Pragmatics* 7/3, 449–469.

VERDONIK, Darinka/KAČIČ, Zdravko, 2012: Označevalci odprte propozicije v govornem diskurzu. *Slavistična revija* 52, 267–281.

VERDONIK, Darinka/ZWITTER VITEZ, Ana, 2011: *Slovenski govorni korpus GOS*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

WIERZBICKA, Anna, 1991: *Cross-cultural Pragmatics: The Semantics of Human Interaction*. Berlin: de Gruyter.

WINTER, Joanne/NORRBY, Catrin, 2000: *Set marking tags 'and stuff'*. *Proceedings of the 1999 Conference of the Australian Linguistic Society*. < <http://www.als.asn.au/proceedings/als1999/winter&norrby.pdf>>.

ZIPF, George K., 1935: *The psycho-biology of language. An introduction to dynamic philology*. Boston: Mifflin.

ZIPF, George K., 1949: *Human behavior and the principle of least effort. An introduction to human ecology*. Cambridge, Mass.: Addison-Wesley.

---

**ZORA**  
**97**

**Phraseologie  
im Wörterbuch  
und Korpus**

---

**Phraseology  
in Dictionaries  
and Corpora**

Herausgeber / Editors  
**Vida Jesenšek**  
**Peter Grzybek**

Maribor, Bielsko-Biala, Budapest, Kansas, Praha  
2014