

## Variation in Language: How to Characterise Types of Texts and Communication Strategies between Orality and Scripturality. Answers given by Koch/Oesterreicher and by Biber

Wolfgang Raible  
(Freiburg im Breisgau)

### Abstract

Over the past few decades, spoken and written texts, as well as their mutual relationship, has come more and more into the focus of linguistic research. Among the different approaches, the present author would like to compare above all two: a largely theory-driven approach, and another that is more data-driven. The theory-driven approach may be characterised by the names of the late Peter Koch (1951–2014) and Wulf Oesterreicher (1942–2015), the data-based one by Douglas Biber (b. 1952). The research objects of Biber are mostly English texts, whereas Koch and Oesterreicher were anchored in the world of Romance languages. It might be interesting that there were practically no relationships between Koch-Oesterreicher and Biber. Koch took note of Biber only in a late publication, and in a superficial and rather unsatisfactory way<sup>1</sup>. This is all the more regrettable since, as will be shown, both approaches led and lead to comparable results, and that they are both based on a comparable theoretical background.<sup>2</sup>

### 1. Koch/Oesterreicher

The approach of Koch and Oesterreicher (henceforth K&Oe) originated during the conception phase of a collaborative research centre (German: *Sonderforschungsbereich*) whose topic was the area of tension between orality and scripturality (German: *Spannungsfelder zwischen Mündlichkeit und Schriftlichkeit*). The authors presented their approach for the first time during one of the internal preparatory meetings. It was immediately received in a quite positive way by the –largely non-linguistic– audience. Later on, it became one of the most important points of the research program itself.<sup>3</sup>

Oesterreicher, during his studies at Tübingen deeply influenced by Eugenio Coseriu (1921–2002), always endeavoured to avoid the ambiguity of concepts, convinced that, speaking with Francis Bacon, *veritas potius emergit ex errore quam ex confusione* ('Truth can more easily emerge from error than from confusion', *Novum Organum* II,xx). This is why, above all, one distinction is fundamental for the approach of K&Oe. Referring to a book by the late Ludwig Söll (1931–1974), *Gesprochenes und geschriebenes Französisch* (1974), they make a strict distinction between 'medium' and 'conception'.

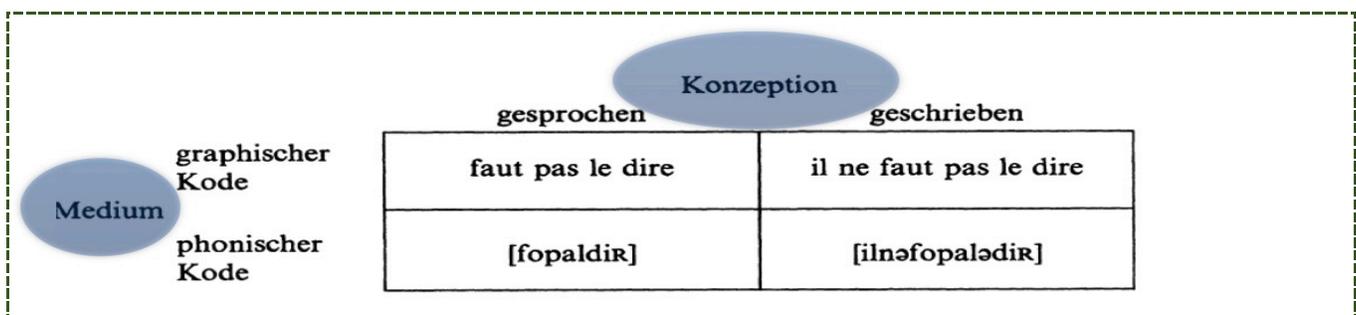


Figure 1: The distinction between 'medium' and 'conception' according to Söll (1974, 18).

A medium may be either 'phonic code' or 'graphic code'. A conception may be either 'spoken' or 'written'. This leads to four possible cross-classifications as seen in the scheme above. For instance, a university lecture is in 'phonic' code, but it is conceptually 'written'.

<sup>1</sup> Oesterreicher & Koch 2016, 82-84.

<sup>2</sup> The part devoted to Biber will be somewhat longer, since most readers in non-anglophone European countries will not be overly familiar with the methods he applies. Hence a (critical) review will perhaps be welcome.

<sup>3</sup> Point two among a total of the seventeen points of the program reads "Skalarer Charakter konzeptioneller Schriftlichkeit" (scalar nature of conceptual scripturality).

The transcript of a live discussion is in ‘graphic code’, but seen from its conception it is rather ‘spoken’.<sup>4</sup> The use of *rather* shows at the same time that ‘conceptually written’ and ‘conceptually spoken’ are necessarily scalar concepts. In fact, Söll enumerated a wealth of attempts to characterise what *écrit* (‘written’) could mean in opposition to *parlé* (‘spoken’): *cultivé, langue littéraire, soigné, prepared, objective* (‘cultivated, literary style, well-kept, prepared, objective’), and so on. This is the point where right from the beginning K&Oe clearly go beyond Söll:

“A closer look onto this twofold distinction shows that the relation between phonic and graphic code should be understood as a strict dichotomy, whereas the polarity between ‘spoken’ and ‘written’ stands for a continuum of conceptual possibilities with innumerable shades and gradations.” (K&Oe 1985, 17)<sup>5</sup>

### 1.1 A threefold lucky find catapults K&Oe into a scholarly orbit

K&Oe see this polarity as a continuum that they illustrate by various text genres, not style labels like Söll, from private conversations to administrative or legal texts. What comes into play here is Koch's background: the influence that Coserius' thinking exerts on Oesterreicher corresponds to the influence of the late Brigitte Schlieben-Lange (1943-2000) on Koch, resulting in the importance the latter attributed to pragmatics. One of Schlieben-Lange's most influential writings at the time was *Traditionen des Sprechens. Elemente einer pragmatischen Sprachgeschichtsschreibung* (‘Traditions of Speaking. Elements of a pragmatic historiography of language’, 1983). Being major elements of the communicative economy of a society, textual genres have their pragmatic settings: They mostly come in series, with one text being the model for another, not without typical (and inescapable) changes during diachrony.

Witness the genre ‘scientific article’ that started in its current form in the 17th century in the *Philosophical Transactions of the Royal Society*. Not only do such texts presuppose a specific readership, a specific kind of wording with a specific vocabulary, but also a specific sequence of textual parts: They start with (1) a passage describing the state of the art, (2) the author's discovering a topic up to now not at all or at least not sufficiently treated by research, (3) proposing new ideas that could fill the gap. After having described the experiments performed according to these ideas, there comes (4) a discussion of the results. The final part tends to be an (5) outlook on what still has to be done in that field of research.<sup>6</sup>

This example clearly shows the conceptual framework such a text is based on and makes evident that its conception is totally ‘written’. Interpreting the opposition between ‘conceptually written’ vs ‘conceptually spoken’ as a *scale of text genres* (not as a series of style labels) is the first and major advance of K&Oe compared to Söll's basic insight. It is curious that this tends to go unnoticed.

Now to the first of two other features that made the model of K&Oe quite famous among German speaking scholars (and in some Romance speaking countries). This feature consisted in giving the two ends of the conceptual scale catchy and memorable names: *Sprache der Nähe* and *Sprache der Distanz* i.e. ‘language of proximity’ and ‘language of distance’ (in French, K&Oe chose the terms *immédiat communicatif* and *distance communicative*). These quite suggestive terms proved to be immediately understandable to a large, but not necessarily linguistic public. It should be clear, though, that these metaphors only sum up a whole series of possible scales extending behind them: active, ongoing processing vs objectification, private vs public communication, spontaneous vs prepared communication, low vs high information density, and so on. – “Grâce à leur caractère métaphorique, ces deux termes englobent la totalité des paramètres conceptionnels” (‘Thanks to their metaphorical character these two terms encompass all conceptual parameters’; K&Oe 2001, 586).<sup>7</sup>

The second of the two other features that made the model of K&Oe famous is its visualisation. There is an axis extending between *Nähe* and *Distanz*. One might want to place this axis in a rectangle horizontally dividing it (cf. Figure 2). Above this axis, one could imagine the graphic realisation, and below, the phonic one, that is, the ‘medium’.

<sup>4</sup> Later on, the authors realised that this distinction between medium and conception was already made by Charles Bally (1865–1947) as well as by Aurélien Sauvageot (1897–1988), a French Finno-Ugrist of high renown and of the highest originality. One might be tempted to think of an apparent paradox formulated by Jorge Luis Borges: A great author creates his precursors (“El hecho es que cada escritor crea a sus precusores”. Borges, *Inquisiciones*, “Kafka y susprecursores”).

<sup>5</sup> “Beigenerer Betrachtung dieser doppelten Unterscheidung stellt sich heraus, daß das Verhältnis von phonischem und graphischem Kode im Sinne einer strikten Dichotomie zu verstehen ist, während die Polarität von *gesprochen* und *geschrieben* für ein Kontinuum von Konzeptionsmöglichkeiten mit zahlreichen Abstufungen steht.”

<sup>6</sup> Biber himself (Biber/Conrad 2009, 157–166) observes changes in scientific articles. The ARCHER-Corpus (1600–1999), created by Biber and Finegan, contains some specimens.

<sup>7</sup> The catchiness of the terms is largely confirmed by the reception of the model: It is commonly dealt with under exactly this name, *Nähe und Distanz* (cf. Feilke/Hennig 2016). At the same time, the scheme of is criticised time and again precisely due to its use of these metaphors (e.g. “What is oral in conceptual orality?” Cf. Zeman (2016)).

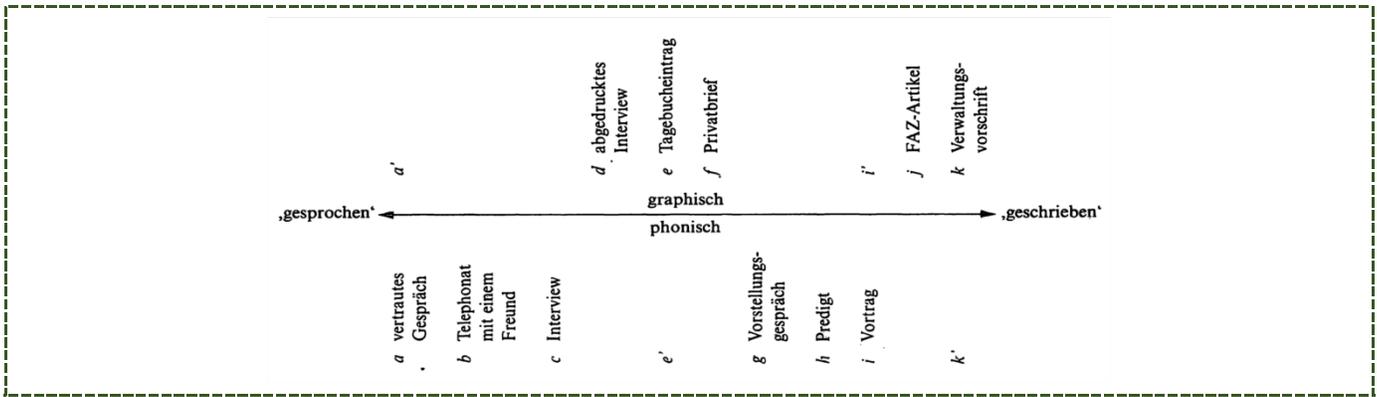


Figure 2: The scale between *Nähe* and *Distanz* seen in a horizontal view: Above are the graphic, below the phonic realisations.(K&Oe 1985, 18).

Instead of this solution, the authors drew the axis for the extension of the conceptual space from the upper left to the lower right edge of the rectangle, thus transforming it into a parallelogram, after having brought the axis once more into a horizontal position (cf. Figure 3). Now the space of the parallelogram on top of the axis would symbolise the increasing probability of a text genre’s being situated on the axis to be implemented in graphic code, whereas the space below would show the decreasing probability of its being orally coded. A conceptually written legal text on the right end will be realised in graphic mode; a spontaneous dialogue situated at the left end will, most probably, be orally conceived and orally realised.

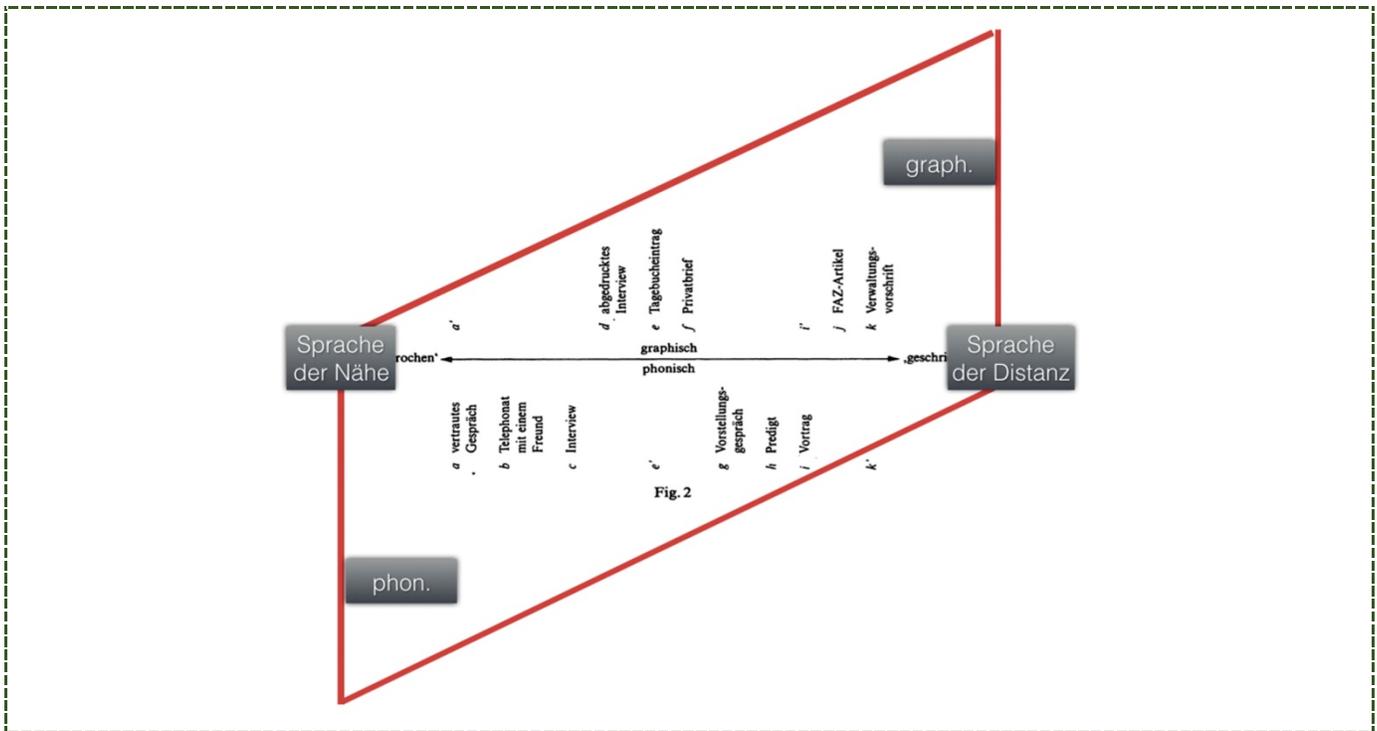


Figure 3: The resulting parallelogram.

### 1.2 K&Oe in Orbit

In its best-known and most developed form, the model took the shape shown below (cf. Figure 4): The two lists of features between brackets on top show the ‘conditions of communication’, the *Kommunikationsbedingungen*.

The corresponding lists on the bottom of the scheme are called *Versprachlichungsstrategien*, ‘strategies of verbalisation’. The small letters (a)–(k) stand exemplarily for genres illustrating the scheme.<sup>8</sup> It should be expected that

<sup>8</sup> (a) Spontaneous conversation between friends; (b) phone call to a friend; (c) interview; (d) printed interview; (e) diary entry; (f) private letter; (g) personal interview; (h) sermon; (i) keynote; (j) newspaper article; (k) administrative regulation.

practically all these strategies of verbalisation and conditions of communication cannot but find their counterparts in the empirically based analyses of Biber.

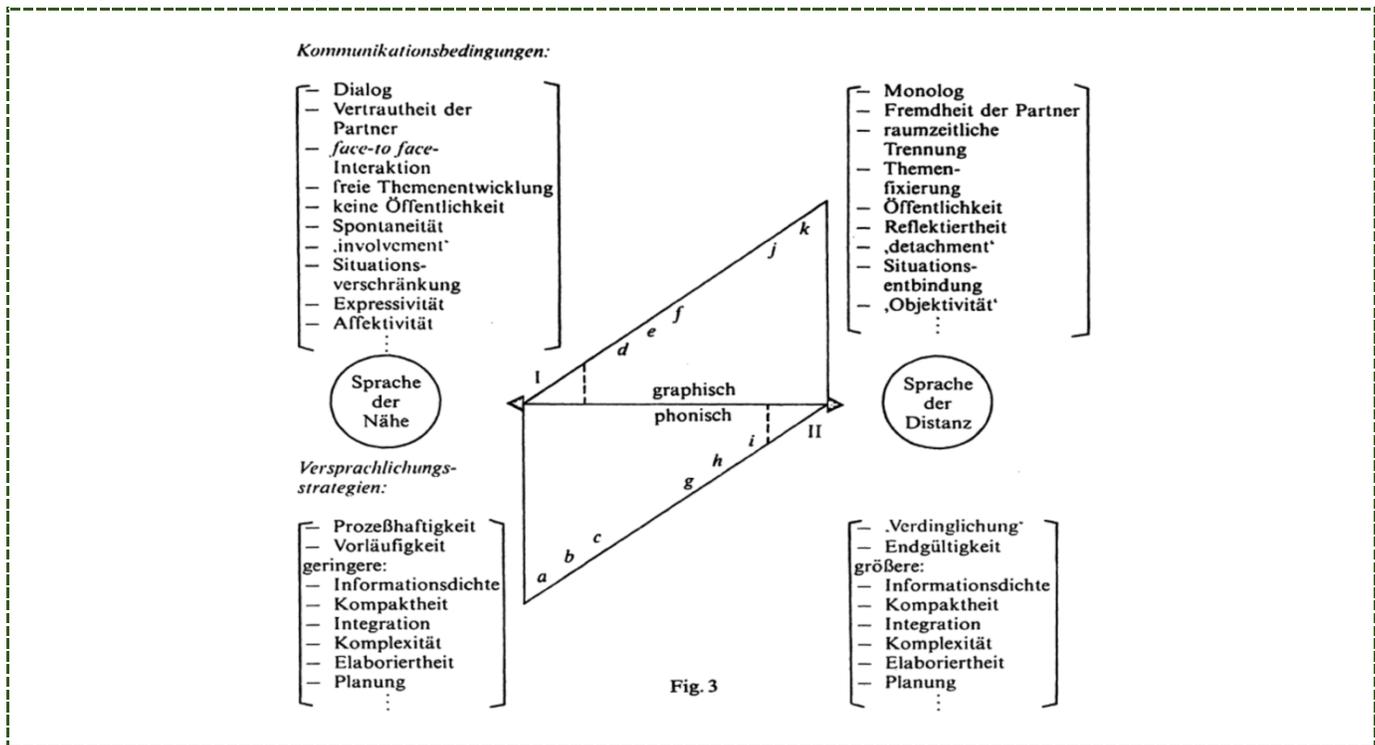


Figure 4: The model of Koch & Oesterreicher in its best-known form<sup>9</sup>.

## 2. Douglas Biber

### 2.1 Going the empirical way, or, how to accomplish the same result with a different set of instruments

The approach of Douglas Biber is fundamentally different from that of K&Oe. I shall illustrate this claim by means of Biber (1986), a seminal paper published in *Language*, and by Biber’s highly influential monograph *Variation Across Speech and Writing* (1988). Additionally, I will refer to a book published in 2009 (Biber & Conrad 2009).

Biber’s approach is statistical, the statistics being based on empirical data coming from individual texts he takes from well-known corpora of English. In Biber (1986), they encompass 16 different text genres – here again, text genres play a basic role – that “represent a broad range of spoken/written situational possibilities” (1986, p. 390). Altogether, Biber takes into account 41 linguistic features in his paper (1986). Two years later he increases the number of features to 67 and the number of genres to 23.

The corpora used in Biber (1986) and (1988) are largely identical, as is shown in the following table:

Written: The LOB Corpus (+ professional letters) (Biber 1986, 390)			Written: Genres 1–15 from the LOB corpus (Biber 1988, 67)		
number of texts			number of texts		
1.	Press reports	44	1.	Press reportage	44
2.	Editorial letters	27	2.	Editorials	27

<sup>9</sup>‘Conditions of communication’ (*Kommunikationsbedingungen*): dialogue vs monologue; partners know one another vs partners without mutual knowledge; face-to-face situation vs separation in space and time; free development of theme vs fixed theme; private vs public; spontaneous vs non spontaneous; involvement vs detachment; relation to situation vs intersubjective fixation of situation; expressive, affective vs objective. – ‘Strategies of verbalisation’ (*Versprachlichungsstrategien*): ongoing process vs reification; provisional character vs irreversibility; lower vs higher informational density; lower vs higher compactness; lower vs higher integration; lower vs higher complexity; less vs more elaborate; lower vs higher degree of planning.

			3.	Press reviews	17
			4.	Religion	17
3.	Skills and hobbies	38	5.	Skills and hobbies	14
4.	Popular lore		6.	Popular lore	14
			7.	Biographies	14
5.	Official documents	30	8.	Official documents	14
6.	Academic prose	80	9.	Academic prose	80
7.	Belles lettres	77			
8.	General fiction	29	10.	General fiction	29
			11.	Mystery fiction	13
			12.	Science fiction	6
			13.	Adventure fiction	13
9.	Romantic fiction	29	14.	Romantic fiction	13
			15.	Humor	9
			16.	Personal letters	6
10.	Professional letters	10	17.	Professional letters	10
Spoken: The LL corpus			Spoken: from London-Lund corpus		
11.	Face-to-face conversation	57	18.	Face-to-face conversation	44
12.	Telephone conversation	20	19.	Telephone conversation	27
13.	Interviews	23	20.	Public conversations, debates, and interviews	22
14.	Broadcasts	19	21.	Broadcast	18
15.	Spontaneous speeches	9	22.	Spontaneous speeches	16
16.	Planned speeches	9	23.	Planned speeches	14
>1 million words		545	Approx. 960,000 words		482
Biber 1986, Table 1. Distribution of text samples			Biber 1988, Table 4.2. Distribution of texts across 23 genres		

**Table 1: The text genres taken into account in Biber (1986) and (1988)**

The biggest difference lies in Belles lettres, a genre only taken into consideration by Biber (1986).<sup>10</sup> In his analyses, Biber relies on the instrument of factor analysis, which comprises two fundamental steps:

- Clustering of the linguistic features tagged in the texts into groups that frequently co-occur.
- Search for underlying factors and their interpretation as textual dimensions, through assessment of the communicative function most widely shared by the features grouped on each factor. These factors are supposed to be the ‘hidden’ factors that ‘organise’ the data observed.

Biber uses principal factor analysis (PFA), slightly different from factor analysis proper. Yet, he is never overly explicit on the subject – among all of his numerous writings, the most detailed information about the methods used to compute the correlations and to extract the factors still remains the already mentioned early book (Biber 1988).

Let us now consider the method in the narrower sense. The first step of each factor analysis must be the normalisation of the frequencies to a text length of, e.g., 1000 words. This leads to the frequencies of linguistic features: Maximum value, minimum value, range, and standard deviation – nouns, verbs, adjectives, prepositions, conjunctions, down-toners, hedges, emphatics, and so on. We need this information for the corpus as a whole and for individual genres distinguished as such in the corpus.

<sup>10</sup>Note that Biber does not explain what he understands by ‘Belles lettres’. Perhaps they correspond in part to Mystery, Science, and Adventure fiction in 1988.

An example: For the present tense, we find in Biber's 1988 corpus an average of 77.7, a maximum of 182.0, a minimum of 12.0 (i.e. a range of  $182-12 = 170$ ) and a correspondingly high standard deviation of 34.3 (in relation to 1000 words of text). The data points for this feature are thus distributed over a wide range of values.

Central is then the matrix with the correlation coefficients calculated for the whole series of linguistic features. The correlation coefficient expresses the covariance of two variables divided by the product of their standard deviations. Such a Pearsonian correlation coefficient takes the form of a number between -1 (totally negative linear correlation) and +1 (totally positive correlation), with zero as no linear correlation at all.

In the case of Biber (1988), the correlation coefficients result in a 67x67 matrix extending over nearly 10 pages (Biber 1988, 270–279). The matrix is even more impressive as it shows the same parts over and under the diagonal separating its two halves: Every combination of features necessarily appears twice.

With regard to the contents of Biber's (1988) matrix, already at first sight, we recognise some interesting correlations: For instance, the existence of first person pronouns strongly correlates with the occurrence of second person pronouns (this should not come as a surprise, though), but also with demonstrative pronouns, the use of *do* as a pro-verb, causal subordinators, particles, contractions, deletion of *that*, private verbs (*to think*, *to feel*, etc.). On the other hand, first person pronouns show a negative correlation with prepositions, attributive adjectives, word length, nouns, agentless passives, etc.

## 2.2 Being at the mercy of statistics applications

The large number of linguistic features taken into account by Biber makes computing, i.e., using one of the available statistics packages, unavoidable. In the case of Biber (1988), for instance, nearly 4,500 correlation coefficients have to be calculated.<sup>11</sup>

There is no doubt that from now on, at least, we will be entirely at the mercy of one of the available statistics packages: Such an application will extract the maximum amount of shared variance among the variables for each of the factors it has discovered. (We can even tell the application, for instance SAS, SPSS, nowadays R or FACTOR, the number of factors it should extract for us). The aim of this procedure being, as we remember, to reduce the vast number of variables to a very limited number of *underlying factors* or, as Biber mostly prefers to say, *dimensions*.

“In a factor analysis, a large number of original variables (in this case, the linguistic features) are reduced to a small set of derived variables (the factors). Each factor represents some amount of variation in the original data that can be quantitatively summarised or generalised – a grouping of variables that cooccur with a high frequency in the data. However, only the first few factors are likely to account for non-trivial amounts of the shared variance, and thus be worth further consideration. In the present case, it was determined that five factors account for non-trivial amounts of variance; these were hence retained for further analysis.” (Biber 1986, 392)

Let us have a look at Dimension (Factor) 1, visualised in Table 2. The degree of variance covered by a linguistic feature is seen as the squared correlation coefficient of the respective feature, the so-called ‘R square’ ( $R^2$ ). Thus .79 in Table 2 tells us that 79% of the *yes/no* questions in the corpus load on this factor.<sup>12</sup> Negative numbers tell us that, for instance, the role of word length is markedly low in this context: Hence, there is a negative or inverse relationship between these two variables. This characterises the textual features we see at the opposite end of the scale of Dimension (Factor) 1. A load of .35% or less is considered negligible in the present context.

Let us now have a look at Biber's two versions of Dimension (Factor) 1. 1986 Factor 1 is called “Interactive [top of the table] vs. Edited text”, 1988 “Involved [top] vs. Informational Production”. The numbers behind the linguistic features are the ‘weight’ or ‘factor loading’ of that feature for the factor. The numbers in the 1986 list are already ‘R squared’, the numbers in the right list (1988) are Pearsonian coefficients followed by ‘R squared’ numbers between brackets. By definition, the first factor should explain the highest degree of the variance of the texts.

<sup>11</sup>A correlation coefficient can be interpreted in many ways. When we think of the unit circle we learned in mathematics class, the co-sine varies between -1 and +1, just like the correlation coefficients:  $\cos(0^\circ)$  is 1,  $\cos(90^\circ)$  is 0,  $\cos(180^\circ)$  is -1. The correlation coefficient could then be considered as the angle between two vectors. But there are even more possibilities (e.g. points on the surface of a unit sphere above the unit circle); in any case, the mental idea of a multidimensional and multivariate space overtaxes the perceptive faculty and imagination of an ordinary mortal.

<sup>12</sup>To ‘load on a factor’ is the usual jargon of factor analysis.

Biber 1986, 393: Factor 1 Numbers = R <sup>2</sup>		Biber 1988, 102: Factor 1 Numbers in brackets = R <sup>2</sup>	
yes/no questions	.79	private verbs	.96 [.92]
that clauses	.76	that deletion	.91 [.83]
final prepositions	.68	contractions	.90 [.81]
pro-verb <i>do</i>	.67	present tense verbs	.86 [.74]
contractions	.67	2nd person pronouns	.86 [.74]
<i>I/you</i>	.62	<i>do</i> as pro-verb	.82 [.67]
general hedges	.61	analytic negation	.78 [.61]
<i>if</i> clauses	.56	demonstrative pronouns	.76 [.58]
WH-questions	.52	general emphatics	.74 [.55]
pronoun <i>it</i>	.49	1st person pronouns	.74 [.55]
other adv. subordinators	.48	pronoun <i>it</i>	.71 [.50]
specific emphatics	.46	<i>be</i> as main verb	.71 [.50]
demonstrative BE/WH	.42	causative subordination	.66 [.44]
WH-clauses	.41	discourse particles	.66 [.44]
general emphatics	.41	indefinite pronouns	.62 [.38]
(present tense	.42)	general hedges	.58 [.34]
(infinitives	.35)	amplifiers	.56 [.31]
word length	-.71	sentence relatives	.55 [.30]
type/token ratio	-.65	WH questions	.52 [.27]
		possibility modals	.50 [.25]
		WH clauses	.47 [.22]
		final prepositions	.43 [.18]
		(adverbs	.42 [.18])
		(conditional subordination	.32 [.10])
		nouns	-.80 [-.64]
		word length	-.58 [-.34]
		prepositions	-.54 [-.29]
		type/token ratio	-.54 [-.29]
		attributive adjs	-.47 [-.22]
		(agentless passives	-.39 [-.15])
		(past participial WHIZ	-.38 [-.14])
		deletions	
		(present participial	-.32 [-.10])
		WHIZdeletions)	
“Table 2: Summary of the factorial structure of 41 linguistic features” [here for Dimension (Factor) 1 (1986)]. <sup>13</sup>		“Summary of the factorial structure” of 67 linguistic features, [here: for Dimension (Factor) 1 (1988)]	

**Table 2: Summary of the factorial structure for Factor 1 (Biber1986 and 1988)**

### 2.3 Problems with naming the factors

If we conduct a factor analysis based on text corpora, the application will extract factors, but it is up to us to give them a name corresponding to the ordering function we see behind the variables loading on a factor. In this case, the author *interprets*, (i.e. in a hermeneutic operation) after an exemplary discussion (Biber 1986, 394sq.), Dimension (Factor) 1 as ‘Interactive [listed above at the top]’ vs ‘Edited text [bottom]’. In the 1988 book, after a still more profound discussion (Biber 1988, 104–108), the name chosen for Dimension (Factor) 1 is ‘Involved’ vs ‘Informational Production’, the interpreted phenomena remaining more or less identical.

That is to say, the most important factor in the 1986 and 1988 corpora corresponds more or less to the scale of K&Oe between *Sprache der Nähe* and *Sprache der Distanz*.

It may be interesting to visualise Dimension 1 in relation to the text genres it covers. In the case of the 1986 corpus, the procedure leading to a scale of text genres does not appear obvious. In 1988, the procedure is quite understandable. In Biber (1988), there are tables showing the mean frequencies of linguistic features found in the genres taken into account.<sup>14</sup> This allows Biber to create, among other things, a table entitled “Descriptive statistics for specialised sub-genres” (Biber 1988, 181–184) showing, for instance, that Dimension 1 in Telephone

<sup>13</sup> Features in parentheses are repeated loadings, and are not used in the computation of factor scores.

<sup>14</sup> “Mean frequency counts of all features in each genre”; Appendix III in Biber (1988, 246–269).

Conversations with personal friends has a mean of 40.8 (a minimum value of 25.7, a maximum of 52.9, hence a range of 27.2, and a standard deviation of 8.6). Thus we might order these genres according to their mean values for Dimension 1 as the following shows, combining pp.128 and 185 of Biber (1988). We understand Biber's comment, based on his interpretation of the resulting scale, that "this is rather an extremely powerful factor representing a very basic dimension of variation among spoken and written texts in English" (Biber 1988, 104).

40	Telephone conversations
	Business telephone conversations
35	Face to face conversations
...	
30	Disparate telephone conversations
...	
20	Personal letters
...	
...	
...	
5	Romantic fiction
	prepared speeches (high percentage of 1st person pronouns)
...	
0	
	sports broadcasts
	mathematics academic prose
	general fiction
	broadcasts
...	
-5	
	non-sport broadcasts
	science fiction
	religion
-10	popular lore; editorials; hobbies
	cultural press reportage
	biographies
	press reviews
	academic prose; press reportage
-15	spot news reportage
	humanities academic prose
	financial press reportage
	natural sciences academic prose
-20	official documents

**Table 3: Mean scores of Dimension 1 for each of the genres 'Involved' vs 'Informational Production' (Biber 1988, 128, 185)**

The genres enumerated in Table 3 illustrate the large presence of newspaper and broadcast texts in Biber's corpus. At the same time, this scale shows that a distinction according to medial orality or scripturality would be irrelevant and unrealistic, since orally realised types of texts are scattered over the entire length of the table. What is crucial is the conceptual aspect: "There do, however, seem to exist some differences in the potential form of speech and writing, due to the different cognitive constraints on speakers and writers" (Biber1988, p. 160; 107 sq.).

#### 2.4 A variety of downstream factors

The result of a Factor analysis is not just one factor, the one onto which the highest number of variables load. As an instrument of 'multidimensional analysis', this method leads to a larger number of dimensions (factors). The following table is an overview over the results yielded by three of Biber's analyses.

The first three Dimensions (Factors) resulting from Biber's multidimensional factor analyses of three text corpora <sup>15</sup>		
Biber 1986:392 sqq. >1 million words	Biber 1988:122 sqq. ≈ 960.000 words	Biber 2009:228 2.7 million words
The composition of the two text corpora is similar		Texts belong to university registers
1: Interactive vs. Edited text	1: Involved vs. Informational Production	1: Oral vs. Literate Discourse
2: Abstract vs. Situated Content	2: Narrative vs. Non-Narrative Concerns	2: Procedural vs. Content-Focused Discourse
3: Reported vs. Individual Style	3: Explicit vs. Situation-Dependent References	3: Reconstructed Account of Events

In all three factor analyses the first dimension is identical, while its denomination, being subject to Biber's interpretation, slightly varies. The downstream dimensions merit some comment, though, because they are not totally identical (although the author underlines the contrary for the 1986 and the 1988 results).

Let us, e.g., have a look onto such genres in the 1988 analysis where Dimension 2 ("Narrative vs. Non-Narrative Concerns") is strongly represented (in parentheses are the maximum values): General Fiction (15.6), Romantic Fiction (11.7), Adventure Fiction (10.5), Mystery Fiction (10.3), Popular Lore (9.2), Biographies (8.0). Least frequent occurrences are in Press Editorials (1.8), Official Documents (-1.5), Personal Letters (1.7), Broadcasts (-0.6). According to the texts illustrating this dimension, we have above all past tenses at one end of the dimension, whereas texts on the other end could be called descriptive. This recalls something like Harald Weinrich's *'Erzählte vs Besprochene Welt'* (in French, *'récit'* and *'commentaire'* – Weinrich<sup>6</sup>2001). For Dimension 3 ("Explicit vs. Situation-Dependent References") the strongest expression is found in Academic Prose (18.6), Official Documents (13.4), Professional Letters (12.4), Press Reviews (10.3), Spontaneous Speeches (9.7) – but not in Fiction genres (-1.2, -1.3, 1.0) or Face-to-Face conversations (1.6).

This means indeed that downstream dimensions reflect the specific content of the text genres they belong to, here narrative concerns and the intention to express absolute –not relative– references in factual texts. All this shows once more the interpretation process behind the denomination of dimensions.

A more developed and detailed description of Biberian Factor analysis would show that the downstream dimensions can be further manipulated or adjusted by a specific statistical operation called 'factor rotation'. It makes the distance and the separation between dimensions become clearer. Instead of explaining the extremely theoretical procedure of rotation, I would like to draw the attention to another phenomenon. K&Oe regard their basic scale as a metaphorical container for a whole series of underlying oppositions: "thanks to their metaphorical character these two terms [the ends of the K&Oe scale] encompass all conceptual parameters" (Koch & Oesterreicher 2001, p. 586.) A sample from the grid of parameters characterising a situation of text production is quite informative in this respect:

Some of K&Oe's Parameters characterising the communicative behavior of partners with respect to the situational context (see above, note 9)	
(4) anchored in action and situation	not anchored in action and situation (4)
(5) depending on situation	independent of the actual situation (5)
(6) presence in space and time	separation in space and time (6)

We can see that Biber's 1988 Dimensions 2 ("Narrative vs. Non-Narrative Concerns") and 3 ("Explicit vs. Situation-Dependent References"), as well as Dimension (2) of 1986 ("Abstract vs. Situated Content") correspond to the parameters (4) to (6) in the list of K&Oe. That is they fit into the framework of K&Oe. This is a property that would also apply to other downstream dimensions of Biber.<sup>16</sup> Their factor loading and thus their significance are usually rather small.

<sup>15</sup> Additional Factors: Biber 1986, 390: Factor 4& 5: "The interpretations of the dimensions underlying these factors are open to refinement, and require further validation". Biber 1988,4: Overt Expression of Persuasion; 5: Abstract vs. Non-Abstract Information; 6. On-Line vs. Informational Elaboration. Biber/Conrad 2009: 4. Teacher-Centered Stance.

<sup>16</sup> Dimension 4 in Biber 1988, "Overt Expression of Persuasion", with its relatively modest loadings (extending here between +3 and -4) resembles something like a condensed extract (under the heading of 'persuasion') parallel to Dimension 1.

Nevertheless, we are advised not to overlook the fact genres attributed to a certain dimension stand in relation to other genres, since they share properties of more than one single dimension: “The relations among any two genres [...] will be a relatively complex comparison of the genres with respect to all dimensions” (Biber 1988, p.168). In other words, neither linguistic features nor their correlations create a mark of distinction, a “unique selling point” for a genre. In my view, the explanation is simple: All texts have to make use of basic grammatical features.

## 2.5 The central role of text genres for both K&Oe and Biber

Both K&Oe and Biber start from the assumption that textual genres (or discourse traditions, as Koch would have preferred) are of fundamental importance. Let me quote in this context an insight from Mikhail Bakhtin [1953] (1986):

“We know our native language –its lexical composition and grammatical structure– not from dictionaries and grammars but from concrete utterances that we hear and that we ourselves reproduce in live speech communication with people around us. We assimilate forms of language only in forms of utterances and in conjunction with these forms. The forms of language and the typical forms of utterances, that is, speech genres, enter our experience and our consciousness together, and in close connection with one another. To learn to speak means to learn to construct utterances (because we speak in utterances and not in individual sentences, and, of course, not in individual words). Speech genres organise our speech in almost the same way as grammatical (syntactical) forms do. We learn to cast our speech in generic forms and, when hearing others’ speech, we guess its genre from the very first words; we predict a certain length (that is, the approximate length of the speech whole) and a certain compositional structure; we foresee the end; that is, from the very beginning we have a sense of the speech whole, which is only later differentiated during the speech process. If speech genres did not exist and we had not mastered them, if we had to originate them during the speech process and construct each utterance at will for the first time, speech communication would be almost impossible.” (Bakhtin 1986,78f.)

While the genres of Biber are real genres, forming a part of his corpora of real English texts, the genres taken into account by K&Oe are essentially theoretical. This is why these authors must necessarily refrain from discussing concrete linguistic features of texts, while Biber registers all these features in detail, his analyses being built exactly on this wealth of data. Hence the theoretical results of Biber’s approach are drawn from concrete English texts, for instance, the insight that behind a dimension like Dimension 1 we have to assume a conceptual attitude, whereas the medial realisation of a genre is of minor importance (Biber 1988, p. 107 sq., 160). In contradistinction to this, K&Oe foster a purely theory-driven model thought to hold for texts of more than one single language.

In spite of this fundamental difference, the scale K&Oe establish between *Sprache der Nähe* und *Sprache der Distanz*, admittedly from the beginning a conceptual, cognitively based scale, seems to be entirely comparable and even equivalent to both Biber’s Dimension 1 in 1986 and 1988, and Dimension 1 in Biber/Conrad (2009).<sup>17</sup>

K&Oe 1985	<i>Sprache der Nähe</i>	<i>Sprache der Distanz</i> <sup>18</sup>
Biber 1986	Interactive Texts	Edited Texts
Biber 1988	Involved Production	Informational Production
Biber& Conrad 2009	Oral Discourse	Literate Discourse

Hence we have to suppose in all these cases an identical underlying ‘cognitive constraint’ at work. This holds all the more as Biber, working with other co-authors on different foreign language corpora: Spanish, Korean, and Somali (Kim/Biber 1994; Biber et al. 2006) always discovers the same dimension 1.

All in all, there is a striking resemblance between both K&Oe and Biber as to their results, although their starting points are diametrically opposed: theory-driven vs data-driven. Even the critiques addressed to K&Oe mostly hold for Biber, too: Multimediality, nowadays a central issue, does not play any role whatsoever. A theory of the media is entirely lacking.<sup>19</sup> Sociolinguistic aspects are not at issue.<sup>20</sup> What is central for both K&Oe and Biber is register variation, i.e.

<sup>17</sup>This already results from the text genres from which the properties of the dimensions (despite their slightly different names) are derived. – It should be added that the corpus analysed in the 2009 book of Biber & Conrad is totally different: It is nearly three times bigger (2.7 million words), even consisting of different text registers (texts used in university teaching and student administration: the TOEFL 2000 Spoken and Written Academic Language Corpus; T2K-SWALI Corpus).

<sup>18</sup>In French: *Immédiat communicatif* vs. *Distance communicative*; In Spanish: *Inmediatez comunicativa* vs. *Distancia comunicativa*.

<sup>19</sup> Cf., e.g., Raible (2010; 2014), who casts an unusual light on the media discussion. Dürscheid (2016) is perfectly right in recommending abstention from discussing this topic in relation to K&Oe.

textual genres in general, not dialect variation. And as has been said, the goal of Biber/Conrad (2009) is to improve academic writing in general, based on the knowledge of its genres and on guidelines you can find while analysing a large text corpus stemming from this domain.

### 3. The missing link between Koch/Oesterreicher and Biber

It is about time to interpret these results against an (up to now invisible) common background. It goes without saying that text genres play a central role for both K&Oe and Biber. K&Oe illustrate their axis between *Sprache der Nähe* and *Sprache der Distanz* with a series of genre names supposed to be located at certain positions on this scale. Biber illustrates his dimensions with analogous scales showing the genres in positions that correspond to the mean frequency counts of linguistic features in each genre.<sup>21</sup> The difference is, we repeat, that Biber's genres come from corpora and exist there as real texts. Now the question is, what could be the *theoretical background* for a relationship between an ordered series of both real and hypothetical text genres?

The answer comes in a detour leading to an author both K&Oe and Biber do not consult in this context: Karl Bühler with his *Sprachtheorie*, published in 1934. Given the history of the (non-)reception of Bühler, Biber could not know him in his early writings – the first English translation is from 1990. K&Oe know Bühler, citing him in another context, although not referring to his *Vierfelderschema* ('four-celled pattern'), the third of the four axioms he outlined.<sup>22</sup>

This third axiom of Bühler is precisely the background of interest in the present context. Bühler starts by evoking how linguistics has to do with four aspects of language, two introduced by Wilhelm von Humboldt, and two by Ferdinand de Saussure.

“Wilhelm von Humboldt spoke of *energeia* and *ergon*, the linguist Saussure picked up the opposition used in French between *la parole* and *la langue* (in English *speech* and *language*) in order to make it the topic of a *linguistique de la parole* running parallel to the traditional *linguistique de la langue*. Since Humboldt there has been practically no expert of any stature who did not sense that something quite remarkable had been touched with the terms *energeia* and *ergon*, and there has been none since Saussure who has not reflected on *la parole* and *la langue*. But neither the old nor the new pair has become properly productive among the fundamental linguistic concepts. Even today occasional attempts are made, sometimes in psychological terms, sometimes in epistemological terms, to claim priority for one of the two elements of the pair *energeia* and *ergon* (Bühler [1934] 1990,57).”<sup>23</sup>

Both pairs of terms being equally important, Bühler creates his famous four-celled pattern (Table 6) in order to unite both approaches in a productive way. Especially the higher level of formalisation has been subject to different interpretations.

Later on in his text, Bühler replaces the Humboldtian, non-Aristotelian opposition between *énérgéia* and *érgon* by the more adequate Aristotelian terms *práxis* and *poíesis*<sup>24</sup>. -- In our context, the most important fact is that, contrary to one's initial impression, the terms on the lower and the higher levels of formalisation should be seen as positions on a scale, not as an exclusive opposition:

<sup>20</sup>Cf. Biber (2009, 264): “Quantitative sociolinguists (especially in the variationist tradition developed by Labov, Trudgill, and others) generally disregard register variation, instead focusing on dialect variation.” – The book edited by Biber/Finegan (1994) is no counterexample.

<sup>21</sup> Since texts belonging to a genre are not clones of one another, one cannot give them a fixed position on a scale, neither along Dimension 1 nor along the K&Oe scale: There will necessarily be a certain bandwidth. This is illustrated in Biber's (1988, 172 sqq.) Figure 8.1, “Spread of scores along Dimension 1”.

<sup>22</sup>Bühler, then a Vienna based internationally renowned psychologist, with links to the Prague Linguistic Circle, published the book in question in 1934. After his emigration to the USA, the book was not available in Germany or Austria. Until 1950 it existed only in a Spanish translation (by Julián Marías Aguilera) – which made Spanish speaking scholars, e.g. Eugenio Coseriu and Klaus Heger, privileged persons knowing this seemingly ‘exotic’, but in fact extraordinary author. Only by 1965 did there appear a reprint of the 1934 edition in Germany, initiating a kind of a pragmatic turn in German linguistics. In 1983 there followed an Italian translation. Only by 1990, 56 years after its first publication, was Bühler eventually available in English (transl. Donald Fraser Goodwin, published by Benjamins).

<sup>23</sup> Some are inclined to think that the notions are Aristotelian. In Aristotle, *énérgéia* is opposed to *dynamis*, not to *érgon*. *Enérgéia* means ‘real’ in opposition to ‘virtual, possible’ (*dynamis*).

<sup>24</sup> The relevant passage in Aristotle's *Nicomachean Ethics* (1040a) makes a distinction between simply ‘doing’ and intentionally ‘making’ something.

	Subject-related phenomena (Humboldt: <i>enérgeia</i> )	Phenomena that have an intersubjective fixation (Humboldt: <i>érgon</i> )
Lower level of formalisation ( <i>Parole</i> )	<b>Speech Action</b> ( <i>Sprechhandlung</i> )	<b>Language Works</b> ( <i>Sprachwerk</i> )
Higher level of formalisation ( <i>Langue</i> )	<b>Speech Acts</b> ( <i>Sprechakt</i> )	<b>Language Structures</b> ( <i>Sprachgebilde</i> )

**Table 6: Karl Bühler's [1934] (1990) four-celled pattern.**

“Things must first be divided in this manner according to the two highest ordering aspects, praxis and poesis, for it is only after making the division that it will become properly possible to see the factual intertwining of these guiding threads in practised, cultivated speaking [...]” (Bühler 1990, p. 63).

The Bühlerian text further shows us that this scale is to be seen as *conceptual*. This is illustrated by the change a playing child undergoes during ontogenesis.

“There is a considerable, palpable difference between games of activity [*Handlungsspiele*] and games of production [*Werkspiele*]; for in the former what is actually supposed to happen with and to the material is only fleetingly and symbolically implied. But then the child makes headway and learns to regard the product of his activity as a work (and this is by no means a matter of course). The first hint that this will happen is when it stops after the fact to look at and admire and to get others to admire what has come to be from its manipulations; [...] Looking back at what is finished or at what has been finished by chance is a stimulus for the child at play, and the decisive phase follows, the phase *in which the result of the activity is anticipated in a conception and thus begins to regulate the operation on the material prospectively and in which finally the activity does not come to rest before the work is completed* (my italics).

Likewise, in principle, one who is engaged in producing a language work does not speak as one engaged in practical action does; for all of us there are situations in which the problem of the moment, the task at hand is solved by speaking directly from within the life situation: speech actions. And there are other situations in which we work productively on the adequate formulation of a given stuff, and produce a language work. Thus, the feature that must be highlighted in the concept “speech action”, the feature without which it is inconceivable, is that speaking is completed (or fulfilled) to the extent that it performs the task of solving the practical problem in the situation.” (Bühler [1934] 1990, p. 62)

The scalarity is even encouraged in the vertical dimension of the four-celled pattern itself. Between a lower and a higher level of formalisation (or abstraction), an intermediate level can be placed: If the lower level is for tokens and the higher one for systemic units, then the middle level can be used for the types representing the tokens. And if the units to be represented on the lower level are real texts (belonging to a genre), the intermediate level is for these genres as text types.

Here is the enlarged four-celled pattern with the added intermediate level, this level being illustrated by a scale from K&Oe themselves (K&Oe 1985, 18):

	Subject-related phenomena (von Humboldt: <i>enérgeia</i> )	Phenomena independent of a subject, having an intersubjective fixation (von Humboldt: <i>érgon</i> )
Lower level of formalization Texts as tokens Saussure: <i>parole</i>	<b>Speech Action</b> ( <i>Sprechhandlung</i> )	<b>Language Works</b> ( <i>Sprachwerk</i> )
Intermediate level of formalisation Texts as types		
Higher level of formalisation	<b>Speech Acts</b> ( <i>Sprechakt</i> )	<b>Language Structures</b> ( <i>Sprachgebilde</i> )

**Table7: Four-celled pattern with an intermediate layer for text genres as types.**

The analyses of Biber, based on the linguistic features of his corpus consisting of text genres, show with a hardly to be surpassed clarity that individual genres quite selectively use the possibilities offered by the system units on the higher level. On the right hand side of the above scale, the genres of Biber do not make use of hedges, discourse particles, general emphatics; instead the type/token ratio and word length are high, attributive adjectives are frequent. As is shown by the quote in the following note, Wulf Oesterreicher himself loved this kind of style in his writings.<sup>25</sup>

On the content level, note 25 with the text cited from Oesterreicher adds complex syntax to the linguistic features of interest we should take into account (see above, Section 2.2).<sup>26</sup>

Learning how to write is an extremely arduous task. The genera on the right side of the above scale must be acquired with difficulty in a long learning process. No ordinary mortal will be able to write a French judgment without further ado. It takes French jurists about two years until they master the task of writing a coherent text of three or four pages with up to 10 levels of subordination – in one single sentence (Krefeld 1985).

**3.1 In diachrony, literacy tends to shift the right end of the scales**

In 2009, Biber and Conrad dedicate a chapter to the changes in text genres over time (2009:143-176). From this point of view, it makes sense to deepen the topic somewhat against the common background of both conceptions. The conceptually demanding genres at the right end of Dimension 1 (Biber) or the *Nähe/Distanz* scale of K&Oe are not made overnight. Instead they are the result of a long, continual process. Scholars of Romance philology all know the Oaths of Strasbourg, considered to be the first text in Old French. They are the exact copy of a Latin pattern (we have analogous oaths in Latin), the wording being mostly rather Old French, but the overall syntax and its articulations being borrowed from Latin. Such a text was not conceivable in Old French at that time (Raible 1994).

Another example are scientific texts. For centuries, Latin remained the domain of such productions. Nicole Oresme (b. before 1330, d. 1382), one of the most original thinkers of 14th-century Europe, for instance, still wrote his treatises in Latin. At the request of Charles V, he translated some of his texts into Middle French, among them *De origine, natura, jure et mutationibus monetarum*, i.e. *Treatise on the origin, nature, law, and alterations of money*.<sup>27</sup>

<sup>25</sup> “Vielmehr ist die anthropologische und sprachtheoretisch-kommunikationstheoretische, also universalistische Fundierung des sprachlichen Gesamtgeschehens in der Sprechfähigkeit ebenso impliziert wie auch die Bedeutung der aktuellen, individuellen Diskurs- und Textexemplare mit ihren jeweiligen Kontexten, die letztlich als empirische Basis aller linguistischen Aussagen zu Sprachlichem gelten müssen.” Oesterreicher, quoted from Feilke/Hennig (2016, 33).

<sup>26</sup> In the 2009 book, there is a kind of to-do list, where complex syntax could figure: “Linguistic features that might be investigated in a register analysis” (Biber/Conrad 2009, 78–82).

<sup>27</sup> With this text, Oresme became the founder of Political Economy, a discipline that even owes its name to him. L. Wolowski, *Traictie de la première invention des monnoies* de Nicole Oresme. Paris 1864, Reprint from 1976.

Since the respective text exists in both Latin and Middle French, it is worthwhile looking at the differences between the two versions. Among other things, one observes a systematic doubling of expressions. To an existing expression in Middle French, Oresme adds another one derived from Latin. He does this 47 times for attributive adjectives, 73 times for nouns, 5 times for prepositions, and 82 times for verbs. In Modern French, both terms have survived, one of them being, in general, more technical. Some examples include *prendre gain et émolument, laquelle controverse et débat, l'origine et commencement, exciter et esveiller, tollerance et souffrance, marchandises et denrées, aparcevable et sensible, appetible et convoitable*.<sup>28</sup> The examples make us understand the role that a high type/token ratio, the relative number of nouns and verbs, not to mention complex syntax, play in highly literate genera. Last but not least, we see once again the importance of attributive adjectives.

It goes without saying that the long, continual process (in literate societies) that results in the development of text genres near the pole of communicative distance, is linked at the same time to equally developed corresponding instruments on the higher level of formalisation, the units of *la langue*. More often than not, the relevant texts are official, legislative, or administrative acts, belonging to juridical genera, such as patent specifications, testaments, judgements and the like – all under the heading of “official documents” in Biber’s approach. Those familiar with the development of Creole languages know that the instruments corresponding to such genres on the level of *la langue* disappear as soon as literacy disappears. For examples, cf. Raible (1994), Michaelis (1994), or Kriegel (1996). For examples from Louisianan or Acadian French, cf. Stäbler (1995) and Wiesmath (2006).

Another example of the enrichment of *la langue* by demands are literary genres in their full development.<sup>29</sup> Novelists of the 17th to 18th centuries were not overly interested in describing situations of communication. A simple *said (s)he* accompanied an utterance and could rarely be seen as corresponding to highly variable situations. All that changes with Romanticism. Look at the following passage from one of the novels of a later, well known author:

— “Oh !” *répondit froidement le colonel en relevant la tête par un mouvement de fierté, “si je succombe, je saurai mourir, mais en compagnie.”*

Là, le vieillard avait disparu. Les yeux de l’homme énergique brillaient rallumés aux feux du désir et de la vengeance. (Balzac, *Le Colonel Chabert*)

Here, two points are remarkable: Authors of this generation want to describe the tone, the speech melody, the gestures, the facial expressions, the body language of their heroes: New description techniques surface, and literary authors become experts in describing multimodal dialogue situations. Second, the author normally refrains from using no more than the traditional speech act verbs (in our quote is still one of them, *répondre*). In Romantic novels one author, Victor Hugo, may use up to 80 different verbs (such as *écrire, s’écrier, murmurer, bégayer, dénigrer*) in such contexts. This is why, in Modern French, we actually have between 400 and 500 speech act verbs (cf. Mocken 2014). In English we observe the same phenomenon: A considerable number of English speech act verbs appear for the first time in Early New English: *to acknowledge, to advocate, to assert, to concede, to remind, to apologise, to question, to request*; or even only in New English: *to remark, to retort, to state, to accept, to guarantee, to volunteer* (cf. Traugott 1987).

Literacy may become a motor of language change, creating features such as new techniques of junction (cf. Raible 1992; 2001), new possibilities of forming heads of object clauses, of compounding adjectives to create complex noun groups, complex syntax (Karlsson 2010), and the like.<sup>30</sup>

### 3.2 An additional, universalistic level

It is time to complete and round off the overall picture. When Biber/Conrad (2009) report on studies done for other languages, we read:

“In many respects, there are similar register patterns across languages. For example, multidimensional studies of register variation in Spanish, Korean, Somali, and English have all identified a first dimension with similar linguistic features and similar differences among registers. In all four languages, this dimension identifies a fundamental opposition between ‘oral’ registers and ‘literate’ registers. [...]”

<sup>28</sup> Cf. also Stempel (1987).

<sup>29</sup> Based on the ARCHER-Corpus, Biber/Conrad (2009) examine genre changes in, for instance, the domain of the English novel.

<sup>30</sup> It should not go unmentioned that the majority of the examples mentioned in section 3.1 are due to working with the conceptual framework of K&Oe.

“Given that these languages are widely divergent in their linguistic/typological characteristics and in their sociocultural contexts, there is no methodological bias in these analyses that would have resulted in a first dimension with these characteristics. Rather, it seems likely that this represents a universal pattern of register variation, at least for languages that have an established literacy tradition.

A second candidate for a universal register pattern is the distinction between narrative and non-narrative registers. All four of these languages have a dimension that distinguishes between written narrative registers (e.g., fiction, folk stories) and all other registers. And all four languages use similar linguistic features to define this dimension, including past tense, communication verbs, third person pronouns, and time adverbials.” (Biber/Conrad 2009, 256, 259)<sup>31</sup>

As has been said, Oesterreicher was above all a student of Eugenio Coseriu’s. As such, he was always, beyond the empirical facts of historical languages, interested in language universals.<sup>32</sup> This is why *theory-driven approach* in the case of K&Oe means that it is based on the universals of speech activity. K&Oe cite these universals right from the start of their 1985 paper.<sup>33</sup> Maria Selig (2017), in her contribution to the *Nähe-Distanz-Modell*, refers time and again to the ‘anthropological (if not necessarily media related) parameters regulating the creative, reflexive and social activity of defining the communicative situation’.

Thanks to parallel studies made on other languages, we now see that Douglas Biber (and Susan Conrad), having started from an empirical base, arrive at comparable results, inferring, based on induction, the existence of universal tendencies.

We now would like to get an overall picture of the theoretical background on which we can map both Biber (and Conrad) and K&Oe. To this end, the model we started from when Karl Bühler was introduced, must be supplemented by a further, universal level (cf. Table 8).

	Subject related phenomena (von Humboldt: <i>énérgiea</i> )	Phenomena independent of a subject, having an intersubjective fixation (von Humboldt: <i>érgon</i> )
Lower level of formalisation: Texts as tokens (Saussure: <i>parole</i> )	<b>Speech Action</b> ( <i>Sprechhandlung</i> )	<b>Language Works</b> ( <i>Sprachwerk</i> )
Intermediate level of formalisation: Texts as types		
Higher level of formalisation: (Saussure: <i>langue</i> )	<b>Speech Acts</b> ( <i>Sprechakt</i> )	<b>Language Structures</b> ( <i>Sprachgebilde</i> )
Highest (universal) level of formalisation (abstraction)	e.g. <i>Sprachhandlungsprogramme</i> (patterns of speech activity) of Hansjakob Seiler, Coseriu’s Universals of Speech Activity	

**Table 8: The enlarged four-celled pattern of Karl Bühler, now with an intermediate level of texts as types, and an additional universalistic level.**

<sup>31</sup> Once more we may refer to Weinrich (<sup>6</sup>2001).

<sup>32</sup> He edited, for instance, together with the present author, one-half, the Universals part, of the two volumes of *Language Typology and Language Universals* (Haspelmath et al. 2001.)

<sup>33</sup> Coseriu’s univesals: Semanticity, otherness [language is directed to others], creativity, exteriority, discoursiveness.

On the highest level, we have the universal principles of Coseriu, i.e. the anthropological basis, or the *Sprachhandlungsprogramme* spelled out by Hansjakob Seiler as a series of so-called dimensions, abstract principles realised in different forms in historical languages. This means that on the level of a higher formalisation (the level of *la langue*), we may have different solutions or different ‘techniques’ (the wording is Seiler’s) of one and the same pattern of speech activity. Some examples: Adjectives are far from being universal. The category of aspect is expressed on the *langue* level by the form of the direct object in Finnish, in other languages tense/aspect is marked on pronouns (Nordlinger & Sadler 2004), and the like.<sup>34</sup>

Biber, too, implicitly notes the existence of specific techniques on the *langue* level. Consider the following quote from Biber/Conrad (2009, p. 259):

“Somali exploits two grammatical devices not found in English: optative mood, which is marked on verb phrases to express polite directives, and directional pre-verbal particles, which indicate whether the action of the verb is occurring either towards or away from the speaker; functionally, these features fit the social requirements of letters in Somali culture, resulting in this distinctive register pattern.”

In summary, both authors state:

“A synthesis of previous research on spoken and written registers shows three general distributional patterns: (1) linguistic features that are common in informational writing tend to be rare in the spoken registers, and vice versa; (2) spoken registers are surprisingly similar to one another in their typical linguistic characteristics, regardless of differences in communicative purpose, interactivity, and pre-planning; but in contrast (3) written registers have a wide range of linguistic diversity.”

The last point means that when it comes, on the right end of Biber’s Dimension 1 (or ‘Communicative Distance’ in the sense of K&Oe), to complex syntax or high degrees of repeated embedding of clauses into clauses, the same principle on the universal level will lead to different techniques or realisations in different languages. The respective genres all have a similar position on Dimension 1, though, while differing in their linguistic materialisation.

#### 4. Some concluding remarks

1. Totally different approaches have led to comparable ordered scales of text genres (the embodiments of linguistic variation). This result of the comparison between K&Oe and Biber may come as a surprise to some scholars.
2. Before Biber and K&Oe, the question of orality vs scriptuality was a field of confusion in the sense of Francis Bacon (cited above, Section 1), mostly due to the non-distinction between ‘medium’ and ‘conception’. Biber (1986, 385 sq.) aptly enumerates some misunderstandings. Then, the centre of interest shifted from ‘medium’ to ‘conception’. (At a higher level, interest is now increasingly focused on the medium again.)
3. The comparison between Biber and K&Oe has indirectly shown that Anglo-centred research on the subject has a certain advantage over the corresponding research done in non-anglophone European countries: The use of electronic devices for text-based research is promising. In the meantime, there are large databases for European languages other than English, too, even tagged ones.<sup>35</sup>
4. The scales of both K&Oe and Biber are a good framework for the observation of language change. New techniques preferably emerge at the literate end of the scales, in part also in ‘Involved production’ at the other end.<sup>36</sup>
5. Beyond doubt, K&Oe’s *Nähe und Distanz* and Biber’s Dimension 1 are crucial for the classification of genres between orality and scriptuality. Downstream dimensions (Biber) reflect more and more the content of individual analysed genera or groups of genera, hence their overall validity is questionable.<sup>37</sup>
6. While studies like those of Biber show the performance capacity we need for applicable results –Biber/Conrad wish to improve academic writing–, K&Oe have clearly shown the effectiveness of a theoretical approach. Starting from reasonable assumptions and a good theoretical background, they have developed a model that has provided guideline for the research of an extremely successful collaborative research centre (*Sonderforschungsbereich*). Even thirty years later, their model, known under the heading of *Nähe und Distanz*, is stimulating discussions (cf., e.g., Feilke/Hennig (2016)).

<sup>34</sup> With this extension on four levels we are not far away from Oesterreicher (2016, 58). But being a student and follower of Coseriu, he favors a three-level representation, omitting the crucial level between the universal and the level of genres as types (his historical level), placing discourse traditions and historical language parallel to each other.

<sup>35</sup> For some time, France had a certain advance in this domain, thanks not least to Robert Martin (b. 1936) and his activities at the Nancy centre, today called ATILF.

<sup>36</sup> This is an important aspect of the famous *Sprachausbau* according to Heinz Kloss (1978).

<sup>37</sup> This is why approaches like those of Landert/Jucker, making parts of the basic dimension of K&Oe three dimensional (Landert/Jucker 2011, 1432), have only a limited value.

7. Although Karl Bühler published his *Sprachtheorie* back in 1934, he still deserves our articular attention. As a psychologist and a specialist both in child psychology and the psychology of thinking, he developed, among other things, new and unaccustomed perspectives on language and communication, for instance the importance of cognitive processes for the production of texts.
8. Even as children we understood more than we could express ourselves. This also applies to adults. Most of us cannot write a legal text, but we think at least that we can understand it. An interesting question now would be how much of the information-rich text genres at the end of Dimension 1 (the pole of “Kommunikative Distanz”) can be understood by people with less language training? This could have some didactic consequences: Writing to be understood – one of the goals Douglas Biber has always pursued.

### Cited Authors

- Bakhtin, Mikhail M. (1986). The Problem of Speech Genres (originally published in 1953). In: Caryl Emerson & Michael Holquist (eds), *Speech Genres and Other Late Essays*. Austin: University of Texas Press.
- Biber, Douglas (1986). Spoken and Written Textual Dimensions in English: Resolving the Contradictory Findings, *Language* 62 (2), 384--414.
- Biber, Douglas (1988). *Variation Across Speech and Writing*, Cambridge Univ. Press.
- Biber, Douglas & Edward Finegan (eds) (1994). *Sociolinguistic Perspectives on Register*. Oxford Univ. Press.
- Biber, Douglas (1995). *Dimensions of Register Variation. A Cross-Linguistic Comparison*. Cambridge Univ. Press.
- Biber, Douglas, Mark Davies, James K. Jones, and Nicole Tracy-Ventura (2006). Spoken and written register variation in Spanish. A multi-dimensional analysis. *Corpora* 1, 7--38.
- Biber, Douglas & Susan Conrad (2009). *Register, Genre, and Style*. Cambridge Univ. Press.
- Bühler, Karl (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fischer.
- Bühler, Karl (1990). *Theory of Language. The representational function of language*. Transl. by Donald Fraser Goodwin in collaboration with Achim Eschbach. Amsterdam & Philadelphia: Benjamins.
- Dürscheid, Christa (2016). Nähe, Distanz und neue Medien, in: Feilke & Hennig (2016), 581--624.
- Feilke, Helmut & Mathilde Hennig (eds) (2016). *Zur Karriere von 'Nähe und Distanz'*. Berlin: de Gruyter.
- Haspelmath, Martin, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible (eds) (2001). *Language Typology and Language Universals -- Sprachtypologie und sprachliche Universalien -- La Typologie des langues et les universaux linguistiques*. An International Handbook – Ein internationales Handbuch -- Manuel international. Handbücher zur Sprach- und Kommunikationswissenschaft, vols 20.1 and 20.2. Berlin: de Gruyter.
- Karlsson, Fred (2010). Syntactic recursion and iteration. In: Harry van der Hulst (ed.), *Recursion and Human Language*. Berlin/New York: Mouton de Gruyter, 43-67.)
- Kim, Young-Jin & Douglas Biber. (1994). A corpus-based analysis of register variation in Korean. In: Biber & Finegan (1994), 157--181.
- Kloss, Heinz. (1978). *Die Entwicklung neuer germanischer Kultursprachen seit 1800*. Düsseldorf: Schwann.
- Koch, Peter & Wulf Oesterreicher (1985). Sprache der Nähe -- Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36, 15--43.
- Koch, Peter & Wulf Oesterreicher (2001) Gesprochene und geschriebene Sprache. Langage parlé et langage écrit, in: *Lexikon der Romanistischen Linguistik*, Tübingen: Niemeyer, vol. I, 2 *Methodologie*, Article 62, 584--627.
- Krefeld, Thomas. (1985). *Das französische Gerichtsurteil in linguistischer Sicht. Zwischen Fach- und Standardsprache*. Frankfurt, Bern [etc.]: Peter Lang.
- Kriegel, Sibylle (1996). *Diathesen im Mauritius- und Seychellen-Kreol*. Tübingen: Narr.
- Landert, Daniela & Jucker, Andreas H. (2011). Private and public in mass media communication: From letters to the editor to online commentaries. In: *Journal of Pragmatics* 43, 1422--1434.
- Michaelis, Susanne (1994). *Komplexe Syntax im Seychellen-Kreol. Verknüpfung von Sachverhaltsdarstellungen zwischen Mündlichkeit und Schriftlichkeit*. Tübingen: Narr.
- Mocken, Susanne (2014). *Informatik erschließt Literatur: Der 'discours citant' in der französischen Romanliteratur zwischen 1750 und 1920*. Freiburg: Universität, Online-Resource. NIHIN: New Ideas in Human Interaction. Studies. <https://www.freidok.uni-freiburg.de/data/9349>
- Nordlinger, Rachel & Louisa Sadler (2004). Nominal tense in crosslinguistic perspective. *Language* 80 (4), 776--806.
- Oesterreicher, Wulf & Peter Koch (2016 [posthumous]). 30 Jahre 'Sprache der Nähe – Sprache der Distanz'. Zu Anfängen und Entwicklung von Konzepten im Feld von Mündlichkeit und Schriftlichkeit. In Feilke & Hennig (2016), 30-125.
- Raible, Wolfgang (1992). *Junktion: eine Dimension der Sprache und ihre Realisierungsformen zwischen Aggregation und Integration*. Heidelberg: Winter. [A short English version is Raible (2001)]

- Raible, Wolfgang (1994). Literacy and Language Change. In: Svetla Čmejrková, František Daneš, and Eva Havlová (eds), *Writing vs Speaking. Language, Text, Discourse, Communication*. Proceedings of the Conference Held at the Czech Language Institute of the Academy of Sciences of the Czech Republic, Prague, October 14--16, 1992. Tübingen, 111--125.
- Raible, Wolfgang (1996). Relatinisierungstendenzen. In: Günter Holtus, Michael Metzeltin, and Christian Schmitt (eds), *Lexikon der Romanistischen Linguistik* Vol. II, 1. Latein und Romanisch. Historisch-vergleichende Grammatik der romanischen Sprachen. Tübingen: Niemeyer.
- Raible, Wolfgang (2001). Linking clauses. In: Haspelmath & al. (2001), 590--617.
- Raible, Wolfgang (2010). SEMIOSE: Die Leistung von Medien und ihren Gattungen. *Poetica* 42, 1--28.
- Raible, Wolfgang (2014). Medium and genre. In: Monika Fludernik and Daniel Jacob (eds), *Linguistics and Literary Studies – Linguistik und Literaturwissenschaft*. Berlin: De Gruyter, 359--383.
- Selig, Maria (2017). Plädoyer für einen einheitlichen, aber nicht einförmigen Sprachbegriff: Zur aktuellen Rezeption des Nähe-Distanz-Modells. *Romanistisches Jahrbuch* 68, 114--145.
- Söll, Ludwig (1974). *Gesprochenes und geschriebenes Französisch*, Berlin: Schmidt.
- Schlieben-Lange, Brigitte (1983). *Traditionen des Sprechens. Elemente einer pragmatischen Sprachgeschichtsschreibung*. Stuttgart: Kohlhammer.
- Stäbler, Cynthia K. (1995). *Entwicklung mündlicher romanischer Syntax. Das 'français cadien' in Louisiana*. Tübingen: Narr.
- Stempel, Wolf-Dieter (1987). Notizen zu Nicole Oresmes Spracharbeit, in: Wolf-Dieter Stempel & Karlheinz Stierle (eds). *Die Pluralität der Welten: Aspekte der Renaissance in der Romania*. Munich: Fink, 11--37.
- Traugott, Elizabeth C. (1987). Literacy and language change. The special case of speech act verbs. *Interchange* 18: 32--47.
- Weinrich, Harald (<sup>6</sup>2001, <sup>1</sup>1964). *Tempus: Besprochene und Erzählte Welt*. München: Beck.
- Wiesmath, Raphaële (2006). *Le français acadien. Analyse syntaxique d'un corpus oral recueilli au Nouveau-Brunswick, Canada*. Paris: L'Harmattan.
- Zeman, Sonja (2016). Nähe, Distanz und (historische) Pragmatik. Oder: Wie 'nah' ist 'Nähesprache'? In: Feilke/Henning (2016), 423-478.