

2. Expression-Syntax: Distribution of word forms

2.1 The contrast to usual Syntax

Anyone hearing the term »syntax« in the frame of our study should completely forget his or her knowledge based on traditional 'syntax of sentences/clauses' and he or she should forget what is understood by »syntax« in the field of »Generative Grammar«. For in both fields observations in the domain of expression variation are always intermingled with questions of content analysis (stating such categories as *singular, plural, present, definite [article], subject, predicate, clause, person* etc.).

My understanding and use of the term (expression-)syntax is totally different. It can be summarized in the following three points:

- »Syn-tax«, a term coming from ancient Greek, literally means nothing other than: »putting together«. Not indicated is *what* should be »put together«. That must be defined separately.
- As mentioned in the previous section the level of expressions, of the word forms, is a completely independent one. That can be shown without difficulty by the theory of signs or by communication theory. One can even use systems theory. The independence of the level of expressions entails that the use of the computer here find a simple and homogeneous level with the consequence that the computer is automatically able to detect much data. The aim of »(Expression-)Syntax« is to analyze the distribution of word forms in the text / in a corpus. Any semantic intrusion will be excluded during data mining. Every text reveals already its individual shape on the level of the »syn-tax/combination of word forms«.
- Such a narrowly defined »Syntax« needs a conceptual frame i. e. a completion by methods that aim at analyzing the meanings of the text: these are »Semantics« for analyzing the literal meaning of single clauses and »Pragmatics« for content analysis on the text level, including indirect, second meanings.⁹

In the following I 'll describe the vocabulary of chapter I without any intrusion of semantic features. Primarily I'll describe the German version.¹⁰ I'm interested in the structure that is perceived primarily by any reader. For before a reader begins to reconstruct the text meaning, he or she has perceived the structured repertory of word forms in that text. And the reader has been impressed and steered by that. Such effects shall now be recognized.

2.2 Distribution of word forms in the German version

2.2.1 Special vocabulary in restricted domains of the text

First we are searching for text domains that contain many word forms not occurring in the rest of

⁹ I comprehensively exemplified – together with others – that concept of analysis and interpretation in describing the Joseph story of the Hebrew Bible: SCHWEIZER (1995).

¹⁰ But I'll add references to the Spanish original.

the text. *Hapax* of course fulfill that condition from the beginning, because they occur only once in the whole text. Additionally, we are interested in *Multi-Types* the repetition of which occurs within a limited distance.¹¹

The results are:¹²

maximum 1: Word forms 250 – 450

maximum 2: Word forms 450 – 740

minimum 1: Word forms 850 – 1375

minimum 2: Word forms 1376 – 1900

maximum 3: Word forms 1900 – 2000

Obviously, shortly after the text beginning the translator offers two areas, which vocabulary that, to a remarkable extent, doesn't occur again in the rest of the chapter.

The statements about the physical appearance Don Quixote's are focused in this way, as well as the question of the origins of the name and the obsession with reading »books of chivalry« (till »verwickelten Redensarten«).

maximum 2 is identical with the domain that describes the difficulties of understanding (»Aristotle himself«) extending to the question which knight should be evaluated as the best one, and finally to the information that Don Quixote is reading day and night.

Following are two areas that contain extremely few word forms exclusively occurring here. It's a long textual passage (twice as long than the two »*maximums*«).

First – *minimum 1* – the topics are activities and reveries of knights, the production of a helmet – out of pasteboard.

That area is followed by a Latin citation (‘*tantum pellis . . .*’) adorned with some proper names – a clear *interim maximum* –: special vocabulary in a very limited area.

minimum 2 entails the search for a name for the horse and for the gentleman himself. That area ends just after the résumé (14.1) and before the topic »knight and Lady«.

The final area again reveals special vocabulary limited to a small area of the text:

The thematically concern is the giant Caraculiambro: he should confirm having been defeated by the permanently underestimated Don Quixote. The final aim of that phantasy: the »sweet lady« is to be impressed.

The interpretation of this data¹³ may point to the insight that the German translator not only is using a vocabulary with a high percentage of variations (see above: word statistics), but he even structures the distribution of word forms cleanly. After an inconspicuous beginning two areas soon follow with special vocabulary. They attract attention and stay in mind. The translator then grants the readers some relaxation (by well-known vocabulary). But that phase of relaxation is interrupted like a flash of lightning in the area *tantum pellis . . .* In this way an impending habituation in the process of text reception is avoided. At the end the number of special word forms increases again. That requires the increased attention of the readers in comparison with the immediately preceding parts.¹⁴ Thus a possibly waning interest is again stimulated by means of the variation of word forms.

11 In our computer program I used rigid parameters: I took the twentieth part of the text (105 words); as factor had been chosen: 21, i. e. in a domain of 105×21 words a certain word form is not allowed to occur another time. In other words: by using such a factor the end of the text will be jumped over. The following results will be confirmed even if one varies the parameters.

12 See Appendix II.

13 Looking at the Spanish original – see Appendix II – shows that the structure is simpler there: *maximum 2* in the German version does have a correspondence in the Spanish original (whereas the first maximum is marked only slightly). The same with the two minimums: Only *minimum 2* has a correspondence in the original. The clear final-*maximum* in the German version is not as distinct in the original or it has a correspondence somewhat earlier.

14 So it would be misleading to pay attention only to contents and the imaginative construction. One even has to observe the structure of the medium (*expressions*).

2.2.2 Increase in vocabulary

The information about the *token / type*-Relation can be used to point to areas of the text that introduce relatively many new *types* that didn't appear in the text before. These are equally sections of the text that specifically capture the attention of the readers.

At the beginning of any text – of course – every word form is new. But it is unclear when an author will interrupt that phase of the concentrated introduction of new word forms, when he or she – so to speak – will modify the ascending flight to a normal horizontal one. In the German version that crucial point is around word number 136.

Words like *Tartsche, Lanzengestell, Windhund, Schüssel, Suppe, Kuh, Hammelfleisch, Fleischkuchen, Überbleibseln, Knochenreste, Linsen, Täubchen . . .* etc. do not sound like words used very frequently in everyday life. That section of the text ends with *Bauerntuch . . . vom feinsten*.

The author is plunging the readers in the details of rural life. According to GÜNTER GRASS the first page of a novel is decisive. The question is whether the reader's attention can be captured? Cervantes seemingly succeeds in doing that with all the culinary and practical details.

The next position where especially many new word forms are introduced is found in the area word number 206 – 245.

It describes the stuttering related to the last name: *Quijada – Quesada – Quijano* with additional new word forms.

We might expect that after a certain introductory area a text doesn't offer further findings, because the main vocabulary has been introduced and further new word forms will only appear dispersedly, no longer in high density. But that assumption is not 'natural law'. Even in an advanced stage our text reveals several areas of highly increase in vocabulary:

Position 850 – 860. The area is not only interesting, even dramatic, because of unexpected *contents*. At the same time – and much easier to grasp – there is an accumulation of new words:

I'm going to underline the new words – from the beginning up to this point: *er könne nicht aufkommen gegen den Ritter vom flammenden Schwert, der mit einem einzigen Hieb zwei grimmige ungeheure Riesen mitten auseinandergewehau.*

In the region of word number 1166 – disregarding function words – the reader could almost guess the story on the basis of the new words:

vornahm – Reinigung – Rüstungsstücken – Urgroßeltern – gehört – Rost angegriffen – Schimmel überzogen, seit langen Zeiten – Winkel hingeworfen – vergessen – reinigte.

At Position 1377 – 1390 a quasi-veterinary description of the horse is striking (*Hufen – Steingallen – Groschen Pfennige – Gebresten*).

Starting with position 1825 there are new: *Baum – Blätter – Frucht – Seele – argen Sünden willen – gutes Glück*

The giant Caraculiambro's admission of having been defeated (position 1926) till Don Quichote's joy over his speech and the discovery of his Lady again increase the vocabulary. The increased informational value on the level of expressions matches the fictional content excellently: different, very surprising insights are reported.

2.2.3 Stagnation of the vocabulary

On the level of word distribution already a text would be boring that shows steady and uniform values under the accents that are relevant in expression syntax. This has not been the case up to now during our analysis of the Cervantes text. This impression continues. The aspect *stagnation of the vocabulary* simply is a counter term to the previous aspect and does not include a stylistic evaluation. *Stagnation* could be understood as pejorative. But that's not my position. For, if an author structures his or her vocabulary in such a way, that distinct areas of a dense introduction of new words can be detected, then he or she needs complementary areas, where extremely few new words are introduced. That is not only a necessity on behalf of statistics. It is even a necessity for the readers: they need phases of rest where mainly known words are repeated and so are intensified. An entire text consisting of a permanent and high increase in vocabulary would affect a permanent overburden of the readers, causing to become disinterested and stop reading.

After the highly informative introductory section a first phase of rest begins at position 300:

Surely at first some nouns are new (*Übung, Verwaltung, Vermögens, Wißbegierde, töricht Leidenschaft, Morgen, Ackerfeld*). But the accompanying function words are more numerous. The information – formulated indeed inconspicuously – follows that the gentleman had brought many books about chivalry to his home and they delighted him. The number and type of words form a phase of rest on the level of word choice.

When Cervantes alludes to *Aristotle* and says that even that philosopher would not have been able to detect anything new in the cited texts, his own language is strikingly noninformative (beginning with position 471 »studierte sich ab, um sie zu begreifen . . .«). A very nice convergence of expression level and that of content.

It can be expected that the phases of stagnation within the text will increase: Beginning with position 658 50 words are relevant. The barber's opinion about the »Knight of Phoebus« is reported there. The stagnation of vocabulary creates the impression: the barber's opinion was not exactly based on real information.

Beginning with position 897 (the speech about the giant Morgante) till end of the text a very long passage is dominated by word stagnation. As shown above that section will be interrupted by short and noticeable word increases again and again. But from here on Cervantes has basically introduced the main vocabulary needed for the remainder of the chapter. Nearly the first half of the chapter passes before that dramatic change takes places, that occurs in any text.¹⁵

2.3 Synopsis of the Results

Summarizing the different types of results we have a well formed profile of chapter I:

Phase 1 (word number 1 – 850): The author starts his text with a high increase in vocabulary. Without delay he captures the full attention of the readers. Beginning with word number 250 a locally restricted vocabulary is dominating. But reading should not become too strenuous, so three brief areas of word stagnation are integrated, so to speak as areas of rest.

¹⁵ We should keep in mind what had been described under 1.3: Nearly half of the entire vocabulary (2112) is *different* words (919): 43,51 % – a very high value.

Phase 2 (word number 850 – 1900): The inclination for innovations at the beginning is gradually being replaced by greater calmness. The increase in vocabulary is diminishing regularly. Therefore even a locally limited special vocabulary cannot be detected – with a single brief exception. The excitement of the beginning has changed into a quieter narration. Only briefly – in order to avoid habituation – small regions with high word increases are encountered there. More striking is the clear rupture at word number 1370, where word increase and locally restricted vocabulary converge.

Phase 3 (beginning with word number 1900) again has its own profile, but that doesn't return totally to the structure of the beginning (the stagnation continues to be valid in principle): By a short and striking locally limited vocabulary (combined with the increase in vocabulary) the information value on the level of word forms is significantly higher. In this way the author avoids an impending habituation to the vocabulary introduced earlier: the final part of chapter I will be perceived attentively. That's due to the structure of the vocabulary and not only due to Dulcinea . . .

6. APPENDIX II: (Expression-) Syntax

German version:

Binnenwortschatz



Spanish original:

Binnenwortschatz

