



<http://dx.doi.org/10.16926/sn.2022.18.11>

Received: 1.09.2022

Accepted: 30.10.2022

MARIIA IVANYTSKA

<https://orcid.org/0000-0002-4870-431X>

(Taras Shevchenko National University of Kyiv, Ukraine)

e-mail: ivanytska@ukr.net

OKSANA PAVLYCHKO

<https://orcid.org/0000-0002-1991-3809>

(Taras Shevchenko National University of Kyiv, Ukraine)

e-mail: pavlychko@gmail.com

TETIANA MALIK

<https://orcid.org/0000-0001-5275-6406>

(Taras Shevchenko National University of Kyiv, Ukraine)

e-mail: atmalik21@gmail.com

MYKOLA KARPIK

<https://orcid.org/0000-0002-4550-075X>

(Taras Shevchenko National University of Kyiv, Ukraine)

e-mail: karpik.mykola@gmail.com

CORRELATION BETWEEN SEMANTICS AND SYNTAX IN GERMAN: SUBORDINATE CLAUSE ANALYSIS

How to cite [jak cytować]: Ivanytska, Mariia, Pavlychko, Oksana, Malik, Tetiana, Karpik, Mykola. "Correlation between semantics and syntax in German: Subordinate clause analysis". *Studia Neofilologiczne. Rozprawy Językoznawcze*, vol. 18, 2022, pp. 209–221.

Korelacja między semantyką a składnią w języku niemieckim: Analiza zdań podrzędnie złożonych

Abstrakt

Niniejsze opracowanie dotyczy strukturalno-semantycznej relacji zdań złożonych w języku niemieckim i ma na celu ustalenie aktualnej i potencjalnej korelacji między strukturami składniowymi a znaczeniem semantycznym słowa. Materiał językowy zdań złożonych pozwala uzasadnić hipotezę, że istnieje korelacja między semantyką a składnią tekstu. W wyniku ana-

lizy statystycznej udowodniono, że semantyka czasowników jest powiązana nie tylko z aktantami, ale także z pewnymi rodzajami określników. Staje się to widoczne w wysokiej częstotliwości kolokacji czasowników należących do pewnych grup leksykalno-semantycznych z pewnymi typami zdań podrzędnych.

Nasza hipoteza jest testowana statystycznie i dlatego możemy ustalić typowe modele zdań złożonych w języku niemieckim. Modele zdań złożonych są kompilowane na podstawie stałych (częstych) elementów – znaczników zdań podrzędnych, które pomagają ustalić ich typowe wzorce.

Słowa kluczowe: zdanie złożone, orzeczenie, zdanie podrzędne, znaczenie, walencja, struktury składniowe.

Abstract

This study considers the structural-semantic organization of complex sentences in German and is aimed at establishing the actual and potential correlation between syntactic structures and semantic meaning of lexical units. The linguistic material of complex sentences makes it possible to substantiate a hypothesis between the correlation of semantics and syntax of the text. It has been statistically proven that the semantics of the verb is related not only to the actants, but also to certain types of specifiers. This becomes apparent in the high frequency of collocation of the verbs belonging to certain lexical-semantic groups with certain types of subordinate clauses.

Our hypothesis is tested statistically, and therefore enables us to establish typical models of complex sentences in the German language. The complex sentence models are compiled on the basis of constant (frequent) elements – the subordinate clause markers, which help to establish their typical patterns.

Keywords: complex sentence, predicate, subordinate clause, meaning, valency, syntactic structures.

1. Introduction

Among all levels of language structure syntax and semantics are most closely connected with the cognitive process and communication. The reflection of this relationship is vividly expressed in the semantics of the main unit of syntactic level – in a sentence, particularly in its semantic-syntactic structure.

Recent developments in modern science have heightened the need to focus more not on isolated phenomena but on investigating them in their interaction. The present study of syntax and semantics relationships is in congruence with the general tendency to a comprehensive analysis of communication units as a complex, which, in turn, is a part of a more profound problem – the relation of language and thought to objective reality.

In the present study, we will analyse the structural-semantic organization of complex sentences in German. The purpose of this investigation is to determine existing and potential correlation between syntactic structures and semantic meaning of the lexical units that make up these structures. The

importance of the present study is that it contributes to the subject matter of the mechanism of realization of systemic relations of different language levels.

The linguistic material of complex sentences makes it possible to examine the intersection of all language levels, functions and aspects and to test a hypothesis of the correlation of semantics and syntax of the text.

2. Hypothesis and course of research

To investigate semantic-syntactic relations in the German language, we have written out 5766 complex sentences from the German fiction books by multiple authors. The material for research were 30 texts of German-language writers of the post-war period – parts of novels and short stories, from which all complex sentences were selected by the continuous sampling method.

The authors were chosen according to the following principle: representatives of different generations, men and women, authors from different regions of Germany, Austria and Switzerland, as well as the former GDR, so that the sample was representative. Among the authors: H. Böll, G. Grass, F. Dürrenmatt, A. Seghers, B. Apitz, E. Strittmater, C. Wolf, H.v. Doderer, E. Aichinger u.a.

The analysis being used is based on statistical methods (Baayen; Field et al.; Gries; Albert & Marx).

The importance of quantitative description of a language is found in the works of Baudouin de Courtenay and Ferdinand de Saussure; however mathematical methods have been actually applied only since about 1950–1960th.

Linguistic analysis, performed at present, tends to be quantitative for the most part. Modern linguistics effectively employs statistical methods and techniques in language research. This is predetermined by a number of reasons: the development of linguistic disciplines that need statistical methods of research (sociolinguistics, psycholinguistics and others); current need for text processing; application of mathematical methods provides deeper insights into relationships in language. Statistical laws, should be noted, are characterised by a rather broad sphere of application: they regulate all complicated systems and language in particular.

The advantages of statistical methods are as follows: they enable to make a valid analysis of linguistic data and verify linguistic hypotheses; allow to quantify a huge range of phenomena and to generalise the data obtained with reliable measures; determine objective and significant relationships of the events under study. A statistical approach provides adequate methods for modelling the events investigated.

In our research, we use statistical methods that test linguistic hypotheses, which may also be quantitative.

This study is designed to assess the hypothesis that the semantics of the main clause verbs (predicates) determines not only the actants that are used in the main sentence, but also the subordinate clause type. Consequently, one can use probabilistic methods and construct semantic-syntactic models of complex sentences that can be useful both for combinatory dictionary developers and as patterns for machine translation.

3. Frequency distribution of different types of subordinate clauses

To analyze the collected data, all the complex sentences have been classified according to the functional-semantic types of clauses, namely, they are classified into 18 types of subordinate clauses: (1) subject, (2) predicative, (3) object, (4) attributive, (5) purpose, (6) concessive, (7) time 1 (simultaneous), (8) time 2 (nonsimultaneous), (9) comparative 1 (unreal), (10) comparative 2 (real), (11) reason, (12) restrictive, (13) relative, (14) clauses of place, (15) result, (16) manner, (17) condition 1 (real), (18) condition 2 (unreal).

Lexical units of each part of the complex sentences (main clause and subordinate clause) are analysed, and first of all, the semantics of predicates and subjects, as well as objects and adverbial modifiers in certain types of sentences. The main clause verbs (predicates) are considered in particular detail, because they form the core of the semantic structure of a whole complex sentence.

All the verbs are classified according to the semantic principle, that is, verb groups with common or similar lexical meaning and similar combinability are singled out, the so-called lexico-semantic groups. The basis for the inclusion of a number of verbs into a separate lexico-semantic group is a seme – the common element of meaning of these verbs. The presence of a common semantic component in the meaning of the verbs of lexico-semantic paradigm is determined on the basis of a componential analysis together with the context.

Thus, each lexico-semantic group consists of verbs that have a common integral semantic component and typical qualifying (differential) components in their meanings, and are also characterized by similarity of combinability and wide development of functional equivalence.

We suggest the following semantic groups of predicates:

- (1) Existence, being (*sein, leben, fehlen, also es gibt*)
- (2) Implicated dynamic state (*tun, machen, misshandeln, erfüllen, wirken, geschehen, vollbringen*)
- (3) Possible and necessary action (*können, dürfen, sollen, müssen, also haben + zu + Infinitiv*)

- (4) Deliberate and intended action (*versuchen, wollen, streben, bemühen, wünschen, suchen*)
- (5) Possession, belonging (*haben, halten, besitzen, gebrauchen, benutzen*)
- (6) Physical effect on the object (*berühren, schieben, streicheln, reiben, schlagen, brechen, prügeln, morden, putzen*)
- (7) Movement, moving (*kommen, sinken, setzen, kriechen, fliehen, übersiedeln, springen, fahren, geraten, fallen, sich neigen, sich wenden*)
- (8) Sensory perception (*sehen, hören, merken, fühlen, beobachten, empfinden, lauschen, spähen*)
- (9) Mental activity (*meinen, denken, entschließen, überlegen, erfahren, schätzen, halten für, erkennen, wagen, irren, vergessen, kennen*)
- (10) Speech, information, communication (*sagen, mitteilen, besprechen, tönen, fragen, nennen, schwätzen, prophezeihen*)
- (11) Emotional state, feelings, emotional perception (*lieben, hassen, gefallen, hängen an, sich freuen, gern haben, erleben, sich vergnügen, fürchten*)
- (12) Phasing (*beginnen, wiederholen*), i.e. the verbs of the beginning, completion continuation and regularity of the action
- (13) Static state (*sich befinden, stehen, übernachten, warten, sitzen, hocken, bleiben, ragen, stecken*)
- (14) Result of action, goal achievement (*gelingen, finden, stoßen auf, erreichen, begegnen, gewinnen, siegen, verwirklichen, erfinden*)
- (15) Delivery, receiving, transfer (*bringen, bekommen, holen, senden, erhalten, führen, kriegen, tragen, senden*)
- (16) Moral, intellectual influence on the object (*überwinden, bezwingen, nachgeben, stürmen, widerstehen, begünstigen, hindern, fesseln, unterdrücken, befehlen, bitten, lassen, verzeihen*)
- (17) Subject state change (*werden, sterben, sich vergrößern, sich teilen, versteinen, verstümmeln, verschwinden, entstehen, erlöschen*)
- (18) State (also verbs denoting state of nature, physiological processes, generative verbs) (*zittern, atmen, bluten, leiden, dursten, hungern*)
- (19) Relative (*gleichen, entsprechen, ähneln*)
- (20) Qualifying: compound nominal predicates (*sein + Adjektiv*)
- (21) Verbs with complex semantics that denote abstract actions in which two or three semes can be dominant (*bezeichnen, zeigen, lesen, zeichnen, schreiben, sparen, winken, lernen, prüfen, dokumentieren*)
- (22) *Sein + N* that conveys the identification of the fact of existence

The predicates of each group are characterized according to their semantic features and valency abilities.

Our data processing involves the following stages: 1) to examine the frequency use of lexico-semantic groups of predicates in different types of complex sentences (using a chi square χ^2 test); 2) to establish a relationship between frequency use of lexico-semantic groups of predicates and definite

lexico-semantic groups of subjects and adverbs (employing quantitative criteria); 3) to determine characteristic semes of the lexical units that make up the sentences in order to form typical models of complex sentences.

The statistical data obtained on the frequency use of every lexico-semantic group of predicates in each sentence type are presented in Table 1.

The analysis of the data presented allows to find out whether there are characteristic relationships between types of complex sentences and lexico-semantic groups of predicates in them.

The use of a chi square (x^2) test evaluates if semantic and syntactical characteristics in a sentence are correlated in any way.

First of all we want to find out whether the difference between the frequencies in Table 1 is significant. This can be done with the help of the x^2 -test.

To do this it is necessary to make some more alternative tables of frequency distribution. Let us take, for example, complex sentences with time clauses (Temporal 1). It is assumed that time clauses are related to verbs of movement. For this purpose Table 2 is compiled.

To calculate the x^2 criterion the following formula is used

$$x^2 = \sum \frac{(a \times d - b \times c)N}{(a + c)(b + d)(a + b)(c + d)}$$

The value obtained is $x^2 = 35,31$. The significant correlation between the verbs of movement and time clauses (Time 1) exists when $x^2 > 3,84$ (Levickij 1989: 7).

The results of the analysis of each lexico-semantic group in every type of subordinate clause show that the difference between the frequencies in Table 1 is significant (Table 3).

If such relationship exists, then the degree of semantic compatibility of the verb with syntactic structure is calculated with the help of the coefficient K (critical value of K : $0 < K < 1$).

As it is well known, the significant value of x^2 only shows that the divergence between the frequencies analysed is significant. But the value of x^2 can measure neither the connection nor the differences between the investigated features. This can be measured with the help of the Chuprov's contingency coefficient K (Urbach 359, 360) as its formula contains the value of x^2 . The following formula is used for our multicell tables:

$$K = \frac{\sqrt{x^2}}{\sqrt{N\sqrt{(r-1)(c-1)}}},$$

where N is the number of observations (the number of sentences studied in Table 1 – $N = 5766$), r is the number of lines in the table, and c is the number of columns.

With the help of this formula, it is possible to find out in which table the contingency of features is higher and in which of them the divergence between the empirical frequencies is greater.

The data obtained are presented in Table 4.

On the basis of such analysis it is possible to use a modelling technique that is realized by the record of frequency use of certain lexical meanings that make up syntactic structures.

Having carried out such calculation for each lexico-semantic group in every type of subordinate clause, we have established significant correlation between semantic and syntactic properties of subordinate clauses.

4. Results of the research

Our results demonstrate that certain semantic classes of lexical units correspond to certain syntactic units equal to the structural parts of complex sentences. The relationship between the verbs of motion in the main clause and subordinate clauses of place and time has been established, as well as between the verbs of mental activity as predicates of the main sentence and object subordinate clauses. Verbs of motion require adverbial subordinate clauses to denote directions; static verbs require clauses of place, verbs of state and behaviour – modal actants. The valency of static verbs cannot include a adverbial modifier of direction, and nominal predicates cannot be combined with clauses of place. It is important to note that the statistical method enables to determine implicit relationships that should be interpreted.

The present investigation has tested our hypothesis statistically, and therefore we can establish typical models of complex sentences in the German language. For example, in the main clauses of time 1 (simultaneous) complex sentences, the predicates expressed by the verbs of movement ($K=0,08$), change of state ($K=0,06$), static state ($K=0,04$) and the physical state of the subject ($K=0,05$) are most frequent. We carried out similar calculations for subordinate clauses and determined their most frequent patterns:

S(N+antr) P(V+ movement U(PB+ direction), U(NS+ time) + change of state)

S(N+abstr) P(V+ state U (PB+ place), U (NS+ time) +phase)

S(N+antr) P(V+ phase) O(N+abstr) U(NS+ time)

where S – Subjekt; N – Nomen; P – Prädikat; U – Umstandsbestimmung; PB – Präpositionalbestimmung; NS – Nebensatz; +antr – anthroponym-HEME; +abstr – abstract-HEME.

We have found that semes +movement, +moving, +change, +state and +phasing are combined first of all with the syntactic structures of temporal categories and denoting place.

Table 1. Frequency distribution of main clause predicates

Lexico-semantic groups of predicates/ Complex sentence type	Existence	Dynamic state	Possible action	Intended action	Possession	Physical effect	Movement	Perception	Mental activity	Community	Emotional state	Subject state	Phrasing	Static state	Result of action	Delivery	Sein + Adjektiv	Moral influence	Abstract action	State	Relative	Sein + N	TOTAL
Subjekt	16	19	19	1	0	3	25	5	7	6	5	23	2	7	8	0	105	31	5	11	9	41	348
Objekt	3	14	107	35	4	5	3	135	336	152	19	8	3	4	12	14	22	40	19	11	2	0	948
Prädikativ	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	35
Attribut	43	22	31	15	31	33	66	42	49	52	20	18	4	37	23	40	27	33	24	15	7	31	663
Vergleich 1	54	23	4	3	2	17	27	34	6	51	7	15	3	14	3	15	22	11	9	46	0	0	366
Vergleich 2	2	6	11	4	3	7	20	8	2	14	9	3	3	7	5	9	30	10	9	9	1	1	173
Kausal	6	24	40	12	14	15	54	17	31	55	59	19	7	15	10	25	75	51	23	22	5	3	582
Temporal 1	28	20	27	20	17	44	123	61	44	56	25	53	15	43	16	31	74	58	25	54	1	13	848
Temporal 2	2	13	15	11	4	16	38	40	8	15	3	15	18	19	9	12	8	27	11	22	0	5	281
Final	2	15	12	6	5	15	7	0	8	14	0	1	1	2	1	21	8	21	9	1	0	2	151
Konzessiv	3	8	15	4	4	8	22	25	19	20	8	4	3	4	4	4	28	11	6	5	1	1	207
Lokal	11	4	9	3	1	3	31	6	1	5	1	2	1	17	3	6	5	3	4	5	1	0	122
Folge	3	1	3	1	4	26	25	8	7	17	8	11	2	13	5	4	72	12	5	15	2	4	248
Bedingung 1	12	14	83	19	14	17	37	11	22	25	13	14	2	13	12	27	28	38	29	10	1	7	448
Bedingung 2	5	8	15	1	4	4	15	4	5	3	2	9	0	3	5	7	15	19	8	1	0	2	135
Einschränkung	1	2	1	4	2	1	4	0	3	4	0	3	0	1	0	0	6	5	1	0	0	0	38
Modal	0	8	12	3	2	11	24	7	3	15	0	1	6	1	6	8	3	17	12	8	0	0	147
Je desto	0	0	1	0	1	0	3	1	3	0	0	10	1	1	0	0	2	1	0	2	0	0	26
TOTAL	225	201	405	142	112	225	524	374	554	504	179	209	71	201	122	223	530	388	199	238	30	110	5766

Table 2. Frequency distribution of lexico-semantic groups of movement and other lexico-semantic groups in the main clauses

Type of clause	Group of verbs	Movement	Other verbs	Total
Time 1		123 <i>a</i>	725 <i>B</i>	848
Other sentences		401 <i>c</i>	4517 <i>D</i>	4918
Total		524	5242	5766 <i>N</i>

5. Conclusion

The analysis of one of the main units of communication – a complex sentence, shows the relevance of the study of the relationship between semantic and structural aspects of the sentence. It helps to reveal the laws of semantic-syntactic meaning formation, the structure and functions of hypotaxis.

The main clause predicate determines the pattern of the complex sentence, but this is not limited to valency relations. Our statistical study confirms that the semantics of the verb is connected not only with the actants but also with certain types of specifiers manifested in the high-frequency joint use of verbs of certain lexical-semantic groups with a kind of subordinate clauses.

The complex sentence models are compiled on the basis of constant (frequent) elements – the subordinate clause markers, which help to establish their typical patterns. Such models are beneficial when learning a foreign language and can also be used for machine translation. The existence of a wide range of transition between language models, types of sentences, signifies the openness and dynamism of the language system.

References

- Albert, Ruth, and Nicole Marx. *Empirisches Arbeiten in Linguistik und Sprachlehrforschung. Anleitung zu quantitativen Studien von der Planungsphase bis zum Forschungsbericht*. 3., überarb. und aktual. Aufl. Narr, 2016.
- Altmann, Gabriel, and M.H. Schwibbe. *Das Menzerathsche Gesetz in informationsverarbeitenden Systemen*. Olms, 1989.
- Baayen, R. Harald. *Analyzing Linguistic Data. A Practical Introduction to Statistics using*. Cambridge University Press, 2008.
- Field, Andy, and Jeremy Miles and Zoë Field. *Discovering Statistics Using*. Sage Publications, 2012.
- Gries, Stephan Th. *Statistics for Linguistics with R. A Practical Introduction*. 2., überarb. und erw. Aufl. De Gruyter Mouton, 2013.
- Field, Andy. *An Adventure in Statistics. The Reality Enigma*. Sage Publications, 2016.
- Hřebíček, Luděk. *Text in Communication: Supra-Sentence Structures*. Brockmeyer, 1992.
- Hřebíček, Luděk. *Text Levels. Language Constructs, Constituents and the Menzerath-Altmann Law*. WVT, 1995.
- Hřebíček, Luděk. „Persistence and Other Aspects of Sentence-Length Series”. *Journal of Quantitative Linguistics*, 4, 1997, 103–109, Routledge is a member of Taylor & Francis Group.

-
- Ivanitska, Maria. L. *Do pitannya pro spivvldnoshennya semantiki I sintaksisu* (na materlall skladnopldryadnih rechen nImetskoYi movi. Chernivtsi.) (Doctoral dissertation.) 1994.
- Levickij V.V. *Statisticheskoe izuchenie leksicheskoy semantiki*. 1989.
- Perebijnis, Valentina I. (Ed.). *Statystycni parametry styliv*. Naukova Dumka, 1967.
- Urbach, V. Ju. *Biometriceskie metody*. Nauka, 1964.
- Wickham, Hadley and Garrett Grolemond. *R for Data Science*, O'Reilly, 2017.
- Yule, G.U. "On a Sentence Length as a Statistical Characteristics of Style in Prose". *Biometrika*, 30, 1939, pp. 363–390.