

Corpus analysis of phraseology in an A1 level textbook of German as a foreign language

Análisis basado en corpus de fraseología en un libro de texto de alemán
como lengua extranjera de nivel A1

María Rosario Bautista Zambrana

Universidad de Málaga. mrbautista@uma.es

Received: 25/05/2017. Accepted: 06/10/2017

Resumen: El objetivo de este artículo es analizar hasta qué punto el libro de texto de alemán como lengua extranjera *DaF kompakt A1* (Sander et al., 2011) cumple con las recomendaciones del *Marco Común Europeo de Referencia para las Lenguas* (Consejo de Europa, 2001) con respecto a la competencia léxica y la competencia sociolingüística en actividades de comprensión y expresión, en concreto en lo concerniente a unidades fraseológicas. En este sentido, nos hemos centrado en las fórmulas fijas y las estructuras fijas presentes en un corpus formado por los materiales del libro de texto, y hemos comprobado si esas expresiones fijas se corresponden con las competencias fraseológicas y sociolingüísticas que se esperan en el Marco para un estudiante de lengua alemana de nivel A1. Con este fin, hemos compilado un corpus con los materiales de comprensión y expresión del libro de texto, formado por tres subcorpus: uno con los textos escritos, otro con los textos orales, y un tercer subcorpus formado por ejercicios. Hemos llevado a cabo un análisis cuantitativo (por medio de *AntConc* 3.4.4 [Anthony, 2016]) y *kJNgram* [Fletcher, 2007]), y uno cualitativo. Nuestros resultados apuntan a que el libro de texto se ajusta a las recomendaciones del Marco.

Palabras clave: corpus; fraseología; alemán como lengua extranjera; Marco Común Europeo de Referencia para las Lenguas; nivel A1.

Abstract: This paper aims to analyse the extent to which the textbook for German as a foreign language *DaF kompakt A1* (Sander et al., 2011) complies with the recommendations of the *Common European Framework of Reference for Languages* (Council of Europe, 2001) (hereafter CEFR) in respect to lexical competence and sociolinguistic competence in receptive and productive activities, specifically with regard to phraseological units. In this respect, we have focused on sentential formulae and fixed frames present in a corpus containing the textbook materials, and we have checked whether

» Bautista Zambrana, María Rosario. 2017. "Corpus analysis of phraseology in an A1 level textbook of German as a foreign language". *Quaderns de Filologia: Estudis Lingüistics* 22: 13-32. doi: 10.7203/qf.22.11298

those fixed expressions correspond to the phraseological and sociolinguistic competences that are expected in the Framework for an A1 level student of German language. To this end, we have compiled a corpus of the textbook receptive and productive materials, made up by three subcorpora: one for the written texts, one for the oral texts, and a third subcorpus containing exercises. We have performed a quantitative analysis (by means of *AntConc* 3.4.4 [Anthony, 2016] and *kJNgram* [Fletcher, 2007]), and a qualitative one. Our results suggest that the textbook complies with the recommendations of the CEFR.

Keywords: corpus; phraseology; German as a foreign language; Common European Framework of Reference for Languages; A1 level.

1. Introduction

This paper is based on the premise that much of the language we use is based on ready-made multi-word combinations, following Sinclair's idiom principle (Sinclair, 1991: 110):

the principle of idiom is that a language user has available to him a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.

A considerable amount of literature has been published following this approach, as well as resources such as the *Academic Phrasebank* (Morley, 2017), which draws on the above-mentioned insight:

It is now accepted that much of the language we use is phraseological in nature; that it is acquired, stored and retrieved as pre-formulated constructions (Bolinger, 1976; Pawley and Syder, 1983). These insights began to be supported empirically as computer technology permitted the identification of recurrent phraseological patterns in very large corpora of spoken and written English using specialised software (e.g. Sinclair, 1991). (Morley, 2017: 5)

This insight has important implications for language teaching and learning. We consider that learning phraseological units is essential for basic level language learners, and that their teaching should start from the very beginning, at the basic levels. As O'Keeffe et al. (2007: 46) state for the case of chunks or clusters¹:

(...) the vocabulary syllabus for the basic level is incomplete without due attention being paid to the most frequent chunks, since many of them are as frequent as or more frequent than single items which everyone would agree must be taught.

¹ As O'Keeffe et al. (2007: 63) explain, there are many terms to describe the phenomena of multi-word vocabulary or chunks: some of these terms are *lexical phrases* (Nattinger and DeCarrico, 1992), *routine formulae* (Coulmas, 1979), *formulaic sequences* (Wray, 2000, 2002), *chunks* (De Cock, 2000), as well as *restricted collocations*, *fixed expressions*, or *multi-word units/expressions*. Throughout this paper we will use the generic terms *phraseological units* and *fixed expressions*, and when referring to our specific object of study, *sentential formulae* or *fixed frames*.

Bearing this in mind, this paper aims to analyse the extent to which the textbook for German as a foreign language *DaF kompakt A1* (Sander et al., 2011) complies with the recommendations of the *Common European Framework of Reference for Languages* (Council of Europe, 2001) (hereafter CEFR) in respect to lexical competence and sociolinguistic competence in receptive and productive activities, specifically with regard to phraseological units. This textbook was selected because we have been using it in several courses at our University since the academic year 2013/2014, with good results and wide acceptance among lecturers and students.

The CEFR describes *lexical competence* as the knowledge of, and ability to use, the vocabulary of a language, and it consists of lexical elements and grammatical elements. The lexical elements comprise, according to the CEFR, single word forms and *fixed expressions*: the latter consist of several words and are used and learnt as wholes (CEFR, 2001: 111). They include sentential formulae, phrasal idioms, fixed frames, phrasal verbs, compound prepositions and fixed collocations. We will focus in this paper on sentential formulae and fixed frames.

Sentential formulae are not defined explicitly in the CEFR, but are described as including three kinds of expressions: direct exponents of language functions such as greetings (e.g. Eng. *How do you do?*, *Good morning!* and deut. *Guten Morgen!*, *Nett*, *Sie kennenzulernen*), proverbs and relict archaisms. We have focused on the first type, direct exponents of language functions, and have looked for minimal communicative units, that can function as autonomous sequences². As for the language functions involved, they are presented in the CEFR (2001: 126) as part of the functional competence³:

1.1 imparting and seeking factual information:

- identifying
- reporting
- correcting

² In this sense, we consider that sentential formulae are *phraseological statements* ('enunciados fraseológicos'), as defined by Corpas Pastor (1996): they are autonomous speech sequences, minimal communicative units, stated with a distinct intonation.

³ These language functions are called specifically *microfunctions* and are defined as "categories for the functional use of single (usually short) utterances, usually as turns in an interaction." (CEFR, 2001: 125)

- asking
 - answering
- 1.2 expressing and finding out attitudes:
- factual (agreement/disagreement)
 - knowledge (knowledge/ignorance, remembering, forgetting, probability, certainty)
 - modality (obligations, necessity, ability, permission)
 - volition (wants, desires, intentions, preference)
 - emotions (pleasure/displeasure, likes/dislikes, satisfaction, interest, surprise, hope, disappointment, fear, worry, gratitude)
 - moral (apologies, approval, regret, sympathy)
- 1.3 suasion:
- suggestions, requests, warnings, advice, encouragement, asking help, invitations, offers
- 1.4 socialising:
- attracting attention, addressing, greetings, introductions, toasting, leave-taking
- 1.5 structuring discourse⁴:
- (28 microfunctions, opening, turntaking, closing, etc.)
- 1.6 communication repair
- (16 microfunctions)

Fixed frames, on the other hand, are described as expressions “learnt and used as unanalysed wholes, into which words or phrases are inserted to form meaningful sentences” (CEFR, 2001: 111), e.g. Eng. *Please may I have ...* or Deut. *Könnte ich bitte ... haben?* Fixed frame is another name for *phrase frame*, which Römer (2009: 150) defines as “sets of n-grams which are identical except for one word, e.g. *at the end of*, *at the beginning of*, and *at the turn of* would all be part of the p[hrase]-frame *at the * of*.”

Lexical competence is associated in the CEFR with the scale of “Vocabulary range”; its descriptor for the A1 level points also to phraseological competence: “Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.”

Sociolinguistic competence, on the other hand, is concerned with the knowledge and skills required to deal with the social dimension of language use, as the CEFR (2001: 118) explains. There are two areas

⁴ We can find the complete lists of microfunctions for structuring discourse and for communication repair in *Threshold Level 1990* (van Ek and Trim, 1991).

here closely related to phraseology: linguistic markers of social relations and politeness conventions. The former comprises the following types of expressions, many of which are fixed (CEFR, 2001: 118):

- use and choice of greetings:
 - on arrival, e.g. *Hello! Good morning!*
 - introductions, e.g. *How do you do?*
 - leave-taking, e.g. *Good-bye . . . See you later*
- use and choice of address forms:
 - frozen, e.g. *My Lord, Your Grace*
 - formal, e.g. *Sir, Madam, Miss, Dr, Professor* (+ surname)
 - informal, e.g. first name only, such as *John! Susan!*
 - informal, e.g. no address form
 - familiar, e.g. *dear, darling*; (popular) *mate, love*
 - peremptory, e.g. surname only, such as *Smith! You (there)!*
 - ritual insult, e.g. *you stupid idiot!* (often affectionate)
- conventions for turntaking
- use and choice of expletives (e.g. *Dear, dear!, My God!, Bloody Hell!, etc.*)

Politeness conventions, for their part, include the following types of expressions (many are as well phraseological in nature) (CEFR, 2001: 119):

1. 'positive' politeness, e.g.:
 - showing interest in a person's well being;
 - sharing experiences and concerns, 'troubles talk';
 - expressing admiration, affection, gratitude;
 - offering gifts, promising future favours, hospitality;
2. 'negative' politeness, e.g.:
 - avoiding face-threatening behaviour (dogmatism, direct orders, etc.);
 - expressing regret, apologising for face-threatening behaviour (correction, contradiction, prohibitions, etc.);
 - using hedges, etc. (e.g. 'I think', tag questions, etc.);
3. appropriate use of 'please', 'thank you', etc.;
4. impoliteness (deliberate flouting of politeness conventions), e.g.:
 - bluntness, frankness;
 - expressing contempt, dislike;
 - strong complaint and reprimand;

- venting anger, impatience;
- asserting superiority.

There is a scale related to sociolinguistic competence, “Sociolinguistic appropriateness”, and it includes a descriptor for the A1 level which mentions phraseological aspects: “Can establish basic social contact by using the simplest everyday polite forms of: greetings and farewells; introductions; saying please, thank you, sorry, etc.”

This study is specifically centered on receptive activities (reception) and productive activities (production). The former include reading and listening activities (CEFR, 2001: 65-71). For the A1 level there are not any descriptors for listening activities that include any reference to fixed expressions, but we do find some descriptors about reading that mention phraseology: in “Overall reading comprehension” it is recommended for the A1 level that the learner can “understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases and rereading as required”. In the section “Reading for orientation” we find that the learner “Can recognise familiar names, words and very basic phrases on simple notices in the most common everyday situations.”

Production, on the other hand, includes speaking and writing activities. With respect to oral production, there is one descriptor for the A1 level that mentions phraseology: in “Overall oral production” it is proposed that the learner “can produce simple mainly isolated phrases about people and places.” As for writing activities, the descriptor “Overall written production” includes the recommendation that the A1 level learner “can write simple isolated phrases and sentences”, while the descriptor “Creative writing” mentions that the learner “can write simple phrases and sentences about themselves and imaginary people, where they live and what they do.”

The specific objective of this paper has been to study the sentential formulae and fixed frames present in a corpus containing the receptive and productive materials of the textbook *DaF kompakt A1*, and to check whether those fixed expressions correspond to the phraseological and sociolinguistic competences that are expected in the Framework for an A1 level student of German language. The remaining part of the paper proceeds as follows: in Section 2 we present the methodology that we have followed to carry out this study, while in Section 3 the results of

the quantitative and qualitative corpus analysis are laid out. Finally, Section 4 offers the discussion of the results, and Section 5 some concluding remarks.

2. Methodology

We have followed a quantitative and a qualitative methodology. In order to perform the linguistic analysis that we have set out to do, we have compiled a corpus of the *DaF kompakt A1* textbook materials, made up by three subcorpora: one for the written texts (letters, e-mails, advertisements, text messages, biographies, news...), one for the oral texts (transcriptions of conversations and monologues, mostly voice messages), and one for the exercises; all of these texts were taken both from the *Kursbuch* ('coursebook') and the *Übungsbuch* ('workbook'). In the case of the spoken and the written subcorpora, we decided to include only complete texts, while for the exercise subcorpus, we selected those activities that contained sentences or at least some type of fixed expressions; in this way, exercises focusing exclusively on single word forms or morphology were left out. The formulation and instructions of the exercises, as well as the grammar reference sections and vocabulary lists, were left out too.

The textbook is a compact method, containing relatively few written texts, a moderate amount of oral texts, and a substantial number of exercises. Thus, the written subcorpus includes 26 texts, containing 2620 tokens and 929 types (type-token ratio 35,46%); the oral component comprises 81 texts, containing 7936 tokens and 1449 types (type-token ratio 18,26%); and the exercise subcorpus is made of 215 texts (each one representing a different task), containing 10250 tokens and 1620 types (type-token ratio 15,8%). As we can see, there is greater lexical variety in the written subcorpus, whereas the exercise subcorpus has the lowest ratio, which means that many of its words occur repeatedly.

We have performed the quantitative analysis by means of *AntConc* 3.4.4 (Anthony, 2016) and *kJNgram* (Fletcher, 2007). We have used the Cluster/N-Gram function of *AntConc* to extract all 2-, 3-, 4- and 5-word n-grams from each corpus. We established a normalised threshold of 250 occurrences per million words for each corpus, which resulted in a minimum threshold of two for the spoken corpus, and of only one occurrence for the written corpus. Even though it might seem a very

low absolute threshold, it is actually a high normalised threshold, which can be justified by the fact that we are dealing with very frequent word combinations, relevant for basic level language learners. The exercise subcorpus, on the other hand, was used for comparison purposes, so all the n-grams extracted in the previous steps were searched for later in this subcorpus.

Afterwards, we employed *kfNgram* to extract all 2- to 6-word phrase frames, i.e. n-grams which are identical except for a single word, from each corpus. We expanded the number of words (*n*) to 6, as we noticed that in that way some more relevant frames could be extracted. As for the options specified, it is worth noting that in order to generate lists of phrase-frames, the programme relies on previously-produced lists of wordgrams (n-grams) with values of *n* of 2 or greater; that is why we generated in the first place as many n-grams as possible, by setting the minimum frequency of occurrence to 1.

As for the qualitative methodology, we examined all n-grams and phrase frames extracted from the oral and the written subcorpora to see which ones complied with the definition of *sentential formulae* and *fixed frames* as proposed by the CEFR, and then compared the results with the n-grams and phrase frames extracted from the exercise subcorpus, so as to check whether the phraseological units laid out in the receptive materials were later practised in the productive sections. In this sense, we could define our work as *corpus-based*, as Storjohann (2005: 8-9) describes:

From this repository, appropriate material is extracted to support intuitive knowledge, to verify expectations, to allow linguistic phenomena to be quantified, and to find proof for existing theories or to retrieve illustrative samples. It is a method where the corpus is interrogated and data is used to confirm linguistic pre-set explanations and assumptions. It acts, therefore, as additional supporting material.

Thus, we have used the corpus to find pre-defined linguistic structures: *sentential formulae* and *fixed frames*. As we mentioned above, both are types of *fixed expressions*, which consist of several words and are used and learnt as wholes (CEFR, 2001: 111). In this way, we have selected those n-grams which fulfilled the conditions to be a sentential formula and complied with any of the language functions listed above.

As for the fixed frames, we followed the same approach: to focus on those that corresponded to minimal communicative units, and that complied with any of the language functions cited above.

The study on sociolinguistic competence, on the other hand, was carried out by reviewing all the sentential formulae that we had previously extracted from the spoken and the written subcorpora, and by determining which ones could meet the criteria to constitute a linguistic marker of social relations, or an expression of politeness. The results were then compared with the expressions found in the exercise subcorpus.

3. Results

We extracted n-grams and phrase frames following the criteria mentioned above, and classified the results in two groups: those related to lexical competence, and those related to sociolinguistic competence.

3.1. *Lexical competence*

We explored the spoken and the written subcorpora separately, in order to detect differences in spoken and written discourse, so we will offer differentiated results.

3.1.1. Spoken subcorpus

From the spoken subcorpus of *DaF kompakt A1* we extracted 60 sentential formulae and 23 fixed frames. We classified the sentential formulae according to the number of words in the n-grams, and noted down which language function (LF) was being fulfilled. Here are some examples of 2-, 3- and 4-grams⁵:

⁵ We did not find any relevant 5-grams.

Rank	Freq.	N-gram	LF
5	20	guten Tag ⁶ ('good morning/afternoon')	1.4
6	18	vielen Dank ('thank you very much')	1.2
25	10	auf Wiedersehen ('goodbye')	1.4
31	9	auf Wiederhören ('goodbye'[telephone])	1.4
47	7	das geht ('it is possible')	1.1/1.2
1	4	wie geht's? ('how are things?')	1.4
23	3	das ist alles ('that's everything')	1.1
57	3	weißt du was? ('you know what?')	1.2
105	2	das klingt gut ('that sounds good')	1.2
134	2	es geht so ('so-so')	1.2
1	4	wie geht's dir? ('how are you?')	1.2
4	3	kann ich Ihnen helfen? ('can I help you?')	1.3
21	2	das geht leider nicht ('unfortunately that is not possible')	1.1/1.2
69	2	können Sie mir helfen? ('can you help me?')	1.3
108	2	wie geht es dir? ('how are you?')	1.2

The 60 sentential formulae that we have found in the oral corpus fulfil the following language functions, as described by the CEFR (2001):

1.1 imparting and seeking factual information	11
1.2 expressing and finding out attitudes	36
1.3 suasion	5
1.4 socialising	10
1.5 structuring discourse	2
1.6 communication repair	1

Out of the 60 sentential formulae extracted from the oral subcorpus, 45 of them are found in the exercise subcorpus, occurring at least once. 22 of them occur three or more times.

In respect to the fixed frames, we classified them according to the number of words and we noted down their language function. Below are some fixed frames of 2-, 3-, 4- and 5-grams.

⁶ Our search was not case-sensitive, but we have capitalized the nouns in these tables of results.

Fixed frame	Total freq.	Nr. of varieties	LF
bis * ⁷ ('see you *')	13	6	1.4
soll ich * ⁸ ('shall I *')	4	4	1.3
das macht * ('that's [price]')	3	3	1.1
wie geht's *? ⁹ ('how are *?')	7	4	1.2
was ist mit *? ⁹ ('what about *?')	3	3	1.1
ich hätte gern * ('I'd like *')	3	2	1.2
mir geht es * ('I am *')	2	2	1.2
wie komme ich zum * ⁹	3	3	1.1

The 23 fixed frames that we have found in the spoken corpus comply with the following language functions:

1.1 imparting and seeking factual information	9
1.2 expressing and finding out attitudes	9
1.3 suasion	2
1.4 socialising	3
1.5 structuring discourse	0
1.6 communication repair	0

As we observe, most of the fixed frames are used to impart and seek factual information, or are related to expressing and finding out attitudes. As in the case of the sentential formulae, we have been barely able to find expressions for structuring discourse or repairing communication.

Out of the 23 fixed frames found in the oral subcorpus, 13 of them appear in the exercises.

3.1.2. Written subcorpus

From the written subcorpus of *DaF kompakt A1* we extracted 25 sentential formulae and four fixed frames. We classified the sentential formulae according to the number of words in the n-grams, and noted down

⁷ Only with nouns or adverbs expressing a point of time in the future, for instance: *bis Montag* ('see you on Monday'), *bis später* ('see you later').

⁸ This phrase frame is actually not only completed by adding one word, but more, but we decided to include it given its function: to propose something.

⁹ In English: 'how do I get to *?'.

which language function (LF) was being fulfilled. Here are some examples of 2-, 3- and 4-grams:

Rank	Freq.	N-gram	LF
6	8	liebe Grüße ('kind regards')	1.5
38	3	du weißt ('you know')	1.2
180	2	sehr gern ('I'd love to')	1.3
3	3	hast du Lust? ('do you feel like <i>it/doing it</i> ?')	1.3
873	1	Gott sei Dank ('thank God')	1.2
1589	1	mit freundlichen Grüßen ('yours sincerely')	1.5
2390	1	wie geht es dir? ('how are you?')	1.2/1.4
2423	1	wir grüßen euch herzlich ('we send our best wishes')	1.5
950	1	hast du Zeit und Lust? ¹⁰	1.3
2011	1	so geht es nicht weiter ('it cannot go on like this')	1.2

The 25 sentential formulae that we have found in the written subcorpus fulfil the following language functions, as described by the CEFR (2001):

1.1 imparting and seeking factual information	2
1.2 expressing and finding out attitudes	6
1.3 suasion	8
1.4 socialising	4
1.5 structuring discourse	7
1.6 communication repair	0

As we can see, the sentential formulae fulfil varied functions, being suasion and structuring discourse the most common.

Out of the 25 sentential formulae detected, 12 are found also in the oral subcorpus, and 15 in the exercise subcorpus (and nine of them occur three or more times).

With regard to the fixed frames, we classified them according to the number of words and we noted down their language function. Below are the fixed frames that we were able to extract (2-, 3-, and 5-grams):

¹⁰ In English: 'do you have time and feel like it?'

Fixed frame	Total freq.	Nr. of varieties	LF
liebe * ('dear *')	9	8	1.5
lieber * ('dear')	5	5	1.5
danke für * ('thanks for *')	2	2	1.2
* gefällt mir sehr gut ¹¹	2	2	1.2

As we can see, two of the fixed frames fulfil the function of structuring discourse, and the other two are used to express attitudes.

Two of the fixed frames found in this subcorpus are present in the exercises: *liebe ** and *lieber **.

3.2. Sociolinguistic competence

We explored the spoken and the written subcorpora separately, so we will offer differentiated results.

3.2.1. Spoken subcorpus

We analysed the sentential formulae that we extracted from the corpus in order to establish which ones could meet the criteria to act as linguistic markers of social relations or as politeness conventions. We found that 10 expressions can be considered linguistic markers of social relations, and all of them are 2-grams. Below are some examples:

Rank	Frequency	Expression
128	4	bis später ('see you later')
245	3	grüß Gott ('hello')
248	3	guten Morgen ('good morning')
252	3	herzlich willkommen ('welcome')
721	2	oh je ('oh dear')

Most of these expressions are related to the use and choice of greetings (on arrival and leave-taking). We also find one expletive (*oh je*). 8 of these expressions are present also in the exercise subcorpus.

As for the politeness conventions, we detected 34 expressions among the sentential formulae that we had previously extracted. Below are some examples (2-, 3- and 4-grams):

¹¹ In English: 'I like * very much'.

Rank	Frequency	Expression
65	6	hier bitte ('here you are')
67	6	ja, gern ('with pleasure')
70	6	kein Problem ('no problem')
92	5	gern geschehen ('my pleasure')
143	4	freut mich ('pleased to meet you')
282	2	tut mir leid ('sorry')
324	2	wie Sie wollen ('as you like')
1	4	wie geht's dir? ('how are you?')
4	3	kann ich Ihnen helfen? ('can I help you?')

Most of these expressions are related to positive politeness (*wie geht's dir*, *freut mich*), while a few correspond to negative politeness (*tut mir leid*). We also find clusters for expressing 'please' or 'thank you' (*vielen Dank*; *nein, danke*). 26 of these expressions are found also in the exercise subcorpus.

3.2.2. Written subcorpus

We determined that nine sentential formulae from the written subcorpus can be considered linguistic markers of social relations. Below are some examples:

Rank	Frequency	Expression
899	1	grüß dich ('hello')
46	2	viele liebe grüße ('lots of love')
1935	1	seid herzlich begrüßt ('best wishes')

Three of these linguistic markers are found in the oral subcorpus, while six are also present in the exercise subcorpus.

On the other hand, we found eight expressions that qualify as politeness conventions, such as the following:

Rank	Frequency	Expression
915	1	guten Appetit ('enjoy your meal')
1862	1	stimmt's? ('right?')
2390	1	wie geht es dir? ('how are you?')

Five of these politeness conventions have been detected also in the oral subcorpus, whereas six are present in the exercise subcorpus.

4. Discussion

In respect to lexical competence, we have divided our results into two groups, oral texts and written texts, and have compared them with those of the exercise subcorpus.

The spoken subcorpus contains 60 sentential formulae and 23 fixed frames, whereas 45 of the formulae (75%) and 13 of the frames (56,5%) are practised in the exercises.

As for the written subcorpus, we have found quite a few sentential formulae (25; 15 of which in exercise subcorpus, 60%). Only four fixed frames have been extracted, whereas two of them (50%) are practised in the exercise section. Some of the sentential formulae are present as well in the oral subcorpus: 12 (48%).

If we consider the results of the oral and the written subcorpora as a whole, we obtain a total number of 73 sentential formulae and of 27 fixed frames. 51 of these sentential formulae are found in the exercise subcorpus (69,86%), while we detect 15 fixed frames (55,55%).

Given these results, and taking into account the amount and variety of sentential formulae and fixed frames that we have encountered, we can state that both the oral and the written subcorpora comply sufficiently with the recommendations of the CEFR in respect to lexical competence: “Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.” Even though the written component is quite small and therefore probably not representative enough of the German language at a basic level, this shortcoming is offset by the fact that a part of its fixed expressions are present also in the oral and exercise subcorpora. Considering this fact, we may state that this subcorpus contributes to the compliance with the reading descriptors in the CEFR: the learner “can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases and rereading as required” and the learner “can recognise familiar names, words and very basic phrases on simple notices in the most common everyday situations.”

Regarding productive activities, although we have used the exercise subcorpus mainly for comparison purposes, we can draw some interesting conclusions: a majority of fixed expressions from the receptive materials are present in this subcorpus, as we stated above, and their frequency is relatively high; while the total number of sentential for-

mulae (tokens) for the oral and written subcorpora is 261 and 43, respectively, we can find 198 tokens in the exercise section. On the other hand, there are 161 fixed frames (tokens) in the spoken component, 18 in the written one and 125 in the exercise subcorpus. This fact allows us to state that the productive component complies sufficiently with the descriptors “Overall written production” (the A1 level learner “can write simple isolated phrases and sentences”) and “Creative writing” (the learner “can write simple phrases and sentences about themselves and imaginary people, where they live and what they do”).

Another significant finding is that we have detected a noticeable difference between the written and the oral subcorpora, in spite of the coincidences that we have mentioned above. In the written subcorpus there are more phraseological units for structuring discourse (*mit freundlichen Grüßen*) and for expressing suasion (*hast du Zeit und Lust?*), while in the oral subcorpus there are more fixed expressions for expressing and finding out attitudes (*das kling gut*), for socialising (*wie geht's?*) and for imparting and seeking factual information (*das geht*).

Regarding sociolinguistic competence, we have also distinguished two groups, oral texts and written texts, and have compared the results with those of the exercise subcorpus.

In the oral subcorpus there are 10 linguistic markers of social relations and 34 politeness conventions (respectively, 8 [80%] and 26 [76,47%] in the exercises). On the other hand, in the written subcorpus there are nine linguistic markers of social relations and eight politeness conventions (respectively, six [75%] and six [75%] in the exercises).

Given the number and variety of expressions extracted from the oral subcorpus (and to a lesser extent, from the written subcorpus), we can state that *DaF kompakt A1* complies with the descriptor of sociolinguistic appropriateness for the A1 level, as formulated in the CEFR (2001: 122): “Can establish basic social contact by using the simplest everyday polite forms of: greetings and farewells; introductions; saying please, thank you, sorry, etc.” The written subcorpus has yielded very limited results, but this can be partially compensated by the fact that the phraseological units are also present in the oral subcorpus (33% of linguistic markers and 62,5% of politeness conventions) and in the exercises of the book, so we can consider that they are sufficiently repeated, in such a way that students are exposed to them.

5. Conclusions

This paper has offered an overview of two competences of the CEFR that are directly related to phraseology: lexical and sociolinguistic competences. We set out to analyse a corpus of the receptive and productive materials of the textbook *DaF kompakt A1*, in order to check whether it contained fixed expressions that corresponded to the recommendations of the CEFR. After analysing the corpus in search for sentential formulae and fixed frames, we can conclude that both the spoken and the written subcorpora (as well as the exercise component) comply with the descriptors laid out in the CEFR. The written subcorpus was very limited in size and therefore did not yield many results in form of fixed expressions, but still they can be considered sufficient, taking into account that a fair amount of the fixed expressions listed occur also in the exercises and in the oral subcorpus.

Apart from the obvious difference in the number of detected fixed expressions, we also determined that the spoken subcorpus contains more expressions related to imparting and seeking factual information, to expressing and finding out attitudes and to socialising, while the phraseological units found in the written subcorpus deal more with suasion and structuring discourse. This is in line with the well-known differences between oral and written discourse.

With respect to sociolinguistic competence, we found numerous expressions that comply with the minimum recommendations for the A1 level. With regard to linguistic markers of social relations there are not significant differences between the spoken and the written subcorpora, but we have observed that politeness conventions are much more prevalent in oral discourse.

Further investigations can be carried out by comparing the results to a general reference corpus of the German language, to determine if these fixed expressions are indeed the most frequent and widely used by German speakers. However, we might encounter the limitation that, to the best of our knowledge, there is not any general reference corpus of the spoken German language that allows the user to create lists of n-grams or clusters.

6. Bibliography

- Anthony, Laurence. 2016. *AntConc* (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. <http://www.laurenceanthony.net/>
- Bolinger, Dwight. 1976. Meaning and memory. *Forum Linguisticum* 1: 1-14.
- Corpas Pastor, Gloria. 1996. *Manual de fraseología española*. Madrid: Gredos.
- Coulmas, Florian. 1979. On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics* 3: 239-266.
- Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- De Cock, Sylvie. 2000. Repetitive phrasal chunkiness and advanced EFL speech and writing. In Mair, C. & Hundt, M. (eds.) *Corpus Linguistics and Linguistic Theory. Papers from ICAME 20*. Amsterdam: Rodopi, 51-68.
- Fletcher, William H. 2007. *KfNgram* [Computer Software]. Annapolis MD: USNA. <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html>
- Morley, John. 2017. *Academic Phrasebank*. Manchester: The University of Manchester. <http://www.phrasebank.manchester.ac.uk/> [Accessed 05/03/2017].
- Nattinger, James & DeCarrico, Jeanette. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- O’Keeffe, Anne; McCarthy, Michael & Carter, Ronald. 2007. *From Corpus to Classroom: language use and language teaching*. Cambridge: Cambridge University Press.
- Pawley, Andrew & Syder, Frances Hodgetts. 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In Richards, J. C. & Schmidt, R. W. (eds.) *Language and Communication*. Longman: New York, 191-226.
- Römer, Ute. 2009. The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics* 7: 141-163.
- Sander, Ilse; Braun, Birgit; Doubek, Margit; Frater-Vogel, Andrea; Trebesius-Bensch, Ulrike; Vitale, Rossana; Behnes, Sibylle; Kotas, Ondrej & Marquardt-Langermann, Martina. 2011. *DaF kompakt A1. Deutsch als Fremdsprache für Erwachsene*. Stuttgart: Klett.
- Sinclair, John. 1991. *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.
- Storjohann, Petra. 2005. Corpus-driven vs. corpus-based approach to the study of relational patterns. In *Conference e-journal, Corpus Linguistics 2005 conference*. Birmingham: University of Birmingham, 1-20.

- <http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx> [Accessed 5/03/2017].
- van Ek, Jan Ate & Trim, John Leslie Melville. 1991. *Threshold Level 1990*. Cambridge: Cambridge University Press.
- Wray, Alison. 2000. Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics* 21(4): 463-489.
- Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.