Abstract
The “Common European Framework of Reference for Languages” (CEFR 2001) is a tool that is widely used as an instructional basis for all foreign language education in Europe. From 2010 to 2012, the European Commission funded the project “European Benchmarks for the Chinese Language” (EBCL) in order to develop benchmarks and competence descriptors for Chinese as a Foreign Language on the basis of the CEFR. After a short introduction of the project (which is freely accessible on the web) which mainly deals with the introductory levels A1 and A2, the article discusses several aspects and conflicts that occurred during the adaptation of CEFR competence descriptors for EBCL reading competence in Chinese on different levels of competency, and gives some examples. In the last chapter, the newly designed descriptors for “Graphemic and Orthographic Competence” in Chinese are introduced in detail.

1. Introduction: Reading Chinese – the perspective of the learner
Reading competence is the foreign language competence that has for centuries been in the foreground of foreign language teaching, whilst oral competences and writing fell to the wayside. Apart from a lack of linguistic skills on the part of the teachers, this was mainly due to geographic distance, nation-state thinking and a primarily knowledge-oriented education system. It was only in the last fifty years and in the course of globalisation that oral foreign-language competency was deemed as equally important. In addition, more recent developments of foreign-language instruction place greater emphasis on interactive and intercultural competences. Thus, reading competence is no longer at the centre of foreign language teaching, but instead is only one of several target competences of modern language education.

Just like the access to the literature of a certain culture used to be the core of language teaching, Chinese language education also mainly focussed on reading competence even into the 1990s, whereby traditional sinology concentrated on Classical Chinese (文言 wényán). It was only during the
second half of the 20th century that Chinese Studies began to attach more importance to reading competence in the modern Chinese language which had come to represent the standard of written language in China since the 1920s. The intensified contact with contemporary China, especially since the 1980s, led to an enormous increase of Chinese language courses all over Europe in general. At the same time, demands for oral Chinese language competence grew in all areas of education, yet without taking into account the special features of the Chinese language and thus without placing greater emphasis on the teaching of Chinese in comparison to the teaching of Common European languages. All available information leads to the assumption that in Europe most Chinese language teaching that is not part of academic courses in Chinese Studies takes place merely at a very elementary level.

Intermediate and Advanced level Chinese in Europe is thus mainly taught at university institutions offering courses in Chinese Studies, which still often focus on the reception of written language texts. Thus, reading competence has been and in many places still is the core of advanced Chinese language education. As noted in Guder (2005), German M.A. graduates and Ph.D. candidates of Chinese Studies considered their reading competence in Chinese to be significantly higher than their other linguistic competences (i.e. level B2-C1 of Common European Framework of Reference for Languages CEFR; oral interaction B2, writing only B1 in average). This is the case with most institutional language learning settings, but also has to do with the fact that philological university courses generally focus strongly on the understanding of written texts.

At the same time, the reading of Chinese texts involves a great deal of frustration for the learners over a long period of time, since authentic Chinese texts do not only contain unknown words (similar to languages using phonographic writing systems) but also a number of unknown characters, the meaning (and pronunciation) of which is not known to the learner. Accordingly, a study on the reading competence in Chinese courses at schools found that even learners in their second or third year of learning are not yet able to understand the content of even the simplest Chinese short stories (Diao 2012).

Thus, this study does not focus on the ability to read the (Chinese) mother tongue, which assumes oral language competence and is generally acquired at primary school age, when the human brain still has significantly different principles of processing. Regarding Chinese as a Foreign Language at the primary school level, there are no empirical data available as of yet. Since, in Europe, Chinese as a Foreign Language is hardly taught systematically before the age of 12, this project is also based on teaching of learners who have started learning Chinese at the age of 12–14 at the earliest.

The desired “Reading Competence for Chinese as a Foreign Language”, which is intended to be made scalable and describable within the scope of
this sub-project of EBCL, is thus challenged to prevail in the tension field described.

1.1. The project “European Benchmarks for Chinese Language” (EBCL)

The proposal of the project “European Benchmarks for Chinese Language” (EBCL) was made against the background of the promotion of multilingualism in Europe with the recognition that less widely used languages (from a European perspective) needed more attention. As there exists an urgent need to raise the awareness of learners and professionals of the linguistic differences between Chinese and European languages as well as diversity in the functionalities of the Chinese language in a European context, the main aim of the project was to create a benchmark framework for Chinese, based upon the Common European Framework of Reference for Languages (CEFR). These benchmarks were developed for people in Europe who learn Chinese either as part of their lifelong learning pursuit, or for professional purposes or as part of school activities as a common portable framework of standards.

The CEFR splits language competence into six levels (A1, A2, B1, B2, C1 and C2), often referred to as elementary, intermediate and advanced levels. In comparison, most European Universities regard a language level of B2 as minimal requirement for attending a university in the respective language area (in humanities even level C1). C2 level stands for a near-native competence in the respected language and is widely regarded as prerequisite to work as a professional language interpreter.

The project was realized through the collaboration of colleagues in departments of Chinese studies at four European universities (SOAS - School of Oriental and African Studies London, Freie Universität Berlin, Université de Rennes, Università La Sapienza in Rome) that have experience in and a high profile of the learning and teaching of the Chinese language, together with some related institutions and high schools across Europe. The project lasted from 2010 to 2012 and produced a collection of descriptors and examples that may be used as a first tool to describe Chinese language competence concerning listening, speaking, oral interaction, reading and writing on the levels A1, A1+, A2 and A2+ of the CEFR in different action domains and with a variety of example tasks.¹

2. Research on reading

2.1. General research on reading

Reading is a highly complex, interactive and creative process, whereby meaning is derived from a combination of text information and readers’ expectations. It consists of several, parallel phases:

1. Decoding of words by analysing and identifying words and word meanings,
2. extraction of meaning at sentence level through an analysis of the syntactic and semantic connections between words, and
3. constitution of the meaning of the text by analysing the correlations and creating connections between the individual sentences and paragraphs in context.

In both mother-tongue and foreign language reading processes of languages using phonographic writing systems, phonological recoding (on the basis of already existing oral language skills) plays a central role at the lowest level of decoding; in this process, graphemes and grapheme combinations are assigned certain phonological units, which help decode words and phrases.

At the other end of the reading process, we find the text as a whole, which is the most complex level of all reading processes and highest in hierarchy. Its goal is the understanding of entire texts, not just words and individual sentences, which, in the end, is the central target of foreign-language reading competence. Propositions the reader creates at sentence level result in macro propositions at text level. In this regard, the creation of coherences, text-constituting characteristics (headings, paragraphs, examples, etc.) as well as the previous knowledge and world wisdom of the reader play a decisive role. The reader must be able to establish meaningful correlations and structure the text into semantic units on his or her own.

It has come to be widely accepted that there are two opposing processes that are equally essential for reading. On the one hand, there is decoding and recognition (data-driven, bottom-up process); on the other hand, there is the construction of meaning (expectation-driven, top-down process; Grotjahn 1995: 535; Bimmel 2002: 116). A complex interaction between those processes is needed in order for the reader to actually comprehend what he or she is reading.

Foreign-language reading is an activity that can have different functions: To read in a foreign language can serve the purpose of gathering information, of entertainment, of improving language skills or even of making judgments. Accordingly, there is a variety of different reading competence tasks in teaching material for languages using phonographic writing systems: apart from questions on texts that belong to very different text types (instruction manu-
als, chats, personal invitations, information signs, scientific papers) or discontinuous texts or charts (maps, tickets, comics), we often find exercises asking the student to find suitable headings, identify key information or do matching tasks in order to practice a rather broad spectre of both text types as well as reading strategies (global, selective, detailed):

- **Scanning**: Looking for certain information.
- **Skimming**: Skim through a text to get a general impression and an overview, cursory/global reading: processing the main contents of the text.
- **Detailed reading**: Detailed understanding of the entire text.
- **Argument-based reading**: Consultation of further material for a more in-depth analysis of the contents of the text (“argumentierendes Lesen”, see Lutjeharms 2010: 11).

It is evident that all these reading strategies originate from the context of languages using phonographic writing systems. We might assume that those strategies are also available to an experienced reader of Chinese, as the Chinese writing system refers to phonological units (syllables) and to a considerable extent contains phonologic information as well, but it is obvious that the use of said strategies in a writing system with several thousand different graphemes shows a complexity that goes beyond the occidental understanding of “reading”.

### 2.2. Research on reading Chinese as a Foreign Language

As the grapheme–phoneme relation that constitutes word recognition does not exist in Chinese in the same way (though some authors regard Chinese writing just as a “deep” orthography like English and French orthography, e.g. Hu and Catts 1998), “reading” in the context of Chinese as a Foreign Language is, even at beginners’ level, subject to processes which cannot be compared to those of languages using phonographic writing systems.

As far as research on reading in the context of Chinese as a Foreign Language is concerned, there are few findings that go beyond the perception level of individual characters and words. Studies on reading comprehension in Chinese as a Foreign Language, especially those conducted in China, generally focus on the comprehension of textual details in secondary and upper secondary education and were conducted primarily with native speakers of Japanese and Korean, who have a culturally different access to the Chinese writing system. The few studies conducted on the (Chinese) reading behaviour of “occidental” learners of Chinese do show, however, that 1) for a fluent reading process, the frequency of a character is more decisive than its complexity (Sergent and Everson 1992), and 2) in the course of Chinese language acquisition, the reading process will become more and more similar
to that of a native speaker (Liu, Perfetti and Wang 2006) and that c) a high oral competence in Chinese has a positive influence on the reading ability (Ptaszynski 2009).

3. **Statistic data on character and word frequencies**

Contrary to the expectations of reading instruction in the 1970s and 1980s, empirical research has shown that the reader has to be familiar with at least 95% of the words to understand a text (Laufer 1997; Hu and Nation 2000; Qian 2002). In order to deduce the meaning of words from the context and to learn unknown vocabulary implicitly – important goals of foreign language reading instruction – it is assumed that 97% of the words must be completely understood (Swanborn and de Glopper 1999, see Tschirner 2005). According to Laufer (1997), for the English language, a vocabulary comprising the 5,000 most frequent lexemes is necessary to understand 95% of the words of an average newspaper text or specialist essay and to be able to read it with comprehension and without losing too much time – a vocabulary scope which is regularly mentioned in connection with level B2 of the CEFR (Tschirner 2010).

A recent Taiwanese calculation of spoken words produced the following similar results: 95% of any text is made up of 4,653 words, for 97%, 7,100 words are needed (Academia Sinica 2012). Also with regard only to spoken language, according to the *Xiandai Hanyu Pinyin Cidian* (Beijing Yuyan Xueyuan Yuyan Jiaoxue Yanjusuo 1986), which still is the most precisely documented survey on word frequency in Chinese texts, only 3,067 words are needed to reach this 95%. Yet, in its survey of written text bodies, the *Xiandai Hanyu Pinyin Cidian* identifies 7,838 lexemes needed to reach 95%.

If we try to apply the findings from Western foreign language research, namely the idea that a reader needs to know 95% to 97% of a text vocabulary to a graphemic level (apparently not done before by current research), we need data to determine the number of characters needed to cover those 95% or 97% of lexical units. Both comprehensive studies on character frequency reveal a number of 1,500 to 2,000 characters (Da 2005: 1,566 / 1,966 characters; *Xiandai Hanyu Pinyin Cidian*: 1,358 / 1,706 characters).2

---

2 The new framework plan for Chinese as a Foreign Language of the PR China claims that a reader could understand 98.21% of any Chinese text knowing only 1,800 characters (or 92.1% with only 900 characters; Zhongguo Guojia Duiwai Hanyu Jiaoxue Lingdao Xiaozu Bangongshi and Jiaoyubu Shekesi 2010). The source of this data remains obscure, though.
Table 1: Available data on word and character amount of 95% resp. 97% of a Chinese text

<table>
<thead>
<tr>
<th>Source</th>
<th>95% of a text characters / words</th>
<th>97% of a text characters / words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laufer (not related to Chinese)</td>
<td>? / 5,000</td>
<td></td>
</tr>
<tr>
<td><em>Xiandai Hanyu Pinlü Cidian</em> (1986; whole corpus)</td>
<td>1,358 / 7,838</td>
<td>1,706 / ?</td>
</tr>
<tr>
<td>Da 2005</td>
<td>1,566 / ?</td>
<td>1,966 / ?</td>
</tr>
<tr>
<td><em>Xiandai Hanyu Pinlü Cidian</em> 1986 (spoken language)</td>
<td>? / 3,067</td>
<td>? / (List ends with 96.65% = 4,000)</td>
</tr>
<tr>
<td>Academia Sinica 2012 (spoken language)</td>
<td>? / 4,653</td>
<td>? / 7,100</td>
</tr>
</tbody>
</table>

Source: Author’s own compilation and calculation.

All this suggests that in order to understand authentic written texts, the knowledge of at least 1,500 characters is needed in addition to a corresponding vocabulary of about 5,000 words (see also Guder 2007).

Independent of an exact definition of the number of characters required, there is an additional problem in that many Chinese characters convey multiple meanings, as can be seen from the example below:

花 huā flower; blossom / pattern; design /
fireworks / essence; cream / wound /
courtesan; prostitute / (Surname) / spend;
expend / flowery; florid / profligate /
blurred / false / randy; lecherous

生 shēng to give birth to; to bear / to
grow / to exist; to live / to be born / liv-
ing; alive / unripe; green / raw; uncooked /
unprocessed; crude / unfamiliar; strange /
stiff; mechanical / existence; life / pupil;
student; scholar

Even if an experienced reader would know all the semantic concepts that have, over centuries, developed from one and the same morpheme or character, he or she would still not be able to grasp the meaning of the combination of the two morphemes in 花生 huāshēng, without having learned that it denotes “peanut”.

Thus, as can be seen again and again in practical reading instruction, the “knowledge” of certain characters does not necessarily lead to an understanding of the corresponding texts, for many characters have different functions and meanings depending on their lexical and situational context (as described in Moser 2010; examples in German in Wang and Guder 2011: footnote 6).

The question as to whether the above mentioned data allows for the conclusion that the knowledge of about 2,000 characters equals “reading competence” in Chinese has not yet been answered. Yet we can assume that the knowledge of 1,500–2,000 characters and 5,000 lexemes consisting mainly of
those characters would correspond to level B2 in terms of reading competence in Chinese as well.3

4. Special features concerning reading processes of Chinese as a Foreign Language

4.1. Implications of the writing system in general

As described, the Chinese writing system, in comparison to almost all other writing systems in the world, is not a primarily phonographic script consisting of a grapheme inventory of less than 100 elements, but is a morphematic-syllabic script. This means that, unlike phonographically transcribed languages that tend to assign a grapheme to each phoneme, the Chinese script assigns a grapheme to each morpheme. Accordingly, the number of graphemes in Chinese amounts to several thousands, and, due to the lack of a clear grapheme-phoneme relation, a phonetic recoding of unknown words is only possible when the learner has reached a relatively high level of literacy and a high (near native) language competence.

An additional problem in this field is that of proper names: whilst in all languages sharing an alphabet writing system, not only so-called “international vocabulary” are widely understood, but even unknown names of persons and places can be phonetically decoded. The recognition or identification of a proper name in a Chinese text depends on the familiarity of the reader with the respective characters and their pronunciation. Furthermore, non-Chinese proper nouns and names are often subject to major phonological changes in order to adapt them to Chinese phonology (e.g. 悉尼 Xīní Sydney, 雅典 Yǎdiǎn Athens, 黑格尔 Hēigé’ěr Hegel, 伏尔泰 Fú’ěrtài Voltaire; see 5.2. point 2).

This phenomenon substantially impedes reading comprehension, especially when identifying names of places and persons. As a result, a variety of exercises that are rather popular and widely used for reading practice of phonographically transcribed languages at the elementary level (identification of persons and places on tickets, information brochures, travel documents, etc.) cannot be used for the instruction of Chinese reading at the same level.

Beyond this rather influential general difference, however, the complexity of reading Chinese texts is further complicated by other features that are often hidden behind the obviously different writing system as will be discussed in the next section.

---

3 This means also a strong contradiction to the CEFR levels that are widely proposed for the Hanyu Shuiping Kaoshi (HSK) / Chinese Proficiency Test which claim HSK level 4 with a vocabulary of only 1,200 lexical items as equivalent to CEFR B2 level. (cf. FaCh 2010).
4.2.  Spoken vs. written language

All language cultures have different registers for spoken and written language, which manifest themselves on various linguistic levels such as choice of words, syntax and idiomaticity. Several studies (Norman 1988; Rosner 1992) prove that this diglossia is a salient feature of Chinese, which can for one thing be explained by the size of the realm of Chinese culture and also by the strong conservatism of the then-used written language (文言 wényán), the implications of which can be seen until this day. In Chinese, the vocabulary used even for the simplest information signs, advertising brochures or letters is often different from that used in colloquial speech, which, in turn, also implies the use of different (and often less frequent) characters. Some frequent examples:4

<table>
<thead>
<tr>
<th>English</th>
<th>Colloquial Chinese</th>
<th>Written Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>where?</td>
<td>在哪里 zài nǎlì</td>
<td>何处 héchù</td>
</tr>
<tr>
<td>from here</td>
<td>从这里 cóng zhèlǐ</td>
<td>由此 yǒucǐ</td>
</tr>
<tr>
<td>if</td>
<td>如果 rúguo</td>
<td>若 ruò</td>
</tr>
<tr>
<td>for (expressing finality)</td>
<td>为了 wèile</td>
<td>以便 yǐbiàn</td>
</tr>
<tr>
<td>seems to</td>
<td>好像 hǎoxiàng</td>
<td>显得 xiānde</td>
</tr>
</tbody>
</table>

Source: Author’s own compilation.

This means that even for a reading competence of the simplest authentic written texts in Chinese, the learners often have to acquire a separate set of characters, a fact that is often used as a further argument for the focus on oral language in beginners’ courses. Conversely, this further complicates the process of acquiring reading competence even with regard to the simplest authentic texts.5

4.3.  Text typology

“Texts” can be both oral as well as written texts. Especially at the beginner’s level and in all foreign languages, oral texts (especially dialogues) are often presented in their written form to illustrate certain linguistic phenomena or to test linguistic knowledge.

Yet, as far as “reading competence” in the strict sense of the word is concerned, such oral texts do not seem relevant, for in real situations, they are

4 A very concise monography on this phenomenon in Chinese is Sun 2012.
5 Thus, the Taiwanese “Test of Chinese as a Foreign Language” TOCFL (2012) only works with authentic texts starting at level 3 (corresponding to B1).
hardly ever put into writing (an exception would be meeting minutes, screen-
plays and theatrical plays). The CEFR itself limits reading comprehension at
levels A1 and A2 to “simple texts” such as postcards, letters, brochures, signs,
newspaper articles describing events and only expands to more complex text
types relating to everyday life for levels B1/B2 (instructions, argumentative
texts etc.).

Thus, the number of text types that learners at levels A1 and A2 mainly
should be able to deal with remains manageable. Nevertheless, we have to
take into account the central problem as mentioned in 4.2 above, namely the
Chinese written language (书面语 shùmiànyǔ). And this, in the end, is ano-
ther one of the reasons why even today Chinese language instruction takes
place almost without authentic written text types for a long period of time.

For the EBCL project, it firstly needed to be ascertained that the text
types mentioned in the CEFR also and to the same degree exist in the Chi-
inese language. This was affirmed, but the question whether those texts pre-
sented a level of difficulty similar to that determined by the CEFR, was di-
sputed. It was shown that there were three further limiting factors (apart from
the always existing problem of fragmentary knowledge of characters) regard-
ing the intelligibility of various “simple” text types:

a) the difficulty with proper names and nouns as mentioned further
above (4.1.),
b) the use of written Chinese/literary language (4.2.), and
c) a seemingly culture-specific, rather limited use of pictorial material,
charts and other visual support mechanisms in certain text types.

4.4. Competing writing system standards for Chinese

Ever since the simplification of the Chinese writing system in the PR China
in the 1960s, there have been two writing systems for Chinese: Taiwan, Hong
Kong and many Chinese communities all over the world use the so-called
“traditional characters” (which are also required when dealing with any Chi-
inese text written before 1964), whilst the PRC back then defined the re-
formed “abbreviated characters” as standard. The changes made in the script
reform concern about 40% of all characters. About half of those 40% consist
of 14 very regular simplifications, which can be taught in both directions in
no time. The remaining 20% have to be learned by heart individually but are,
in most cases, graphically based on their original characters.

Though in recent times both systems overlap concerning their use (the
Taiwanese admit that they use some abbreviated characters when handwrit-
ing, and the use of traditional characters tend to convey quality and academia
on the mainland), most institutions teaching Chinese as a Foreign Language
(not those teaching Chinese as a heritage language!) have adopted the abbreviated characters as their standard writing system.

The popular debate as to which of the two systems is more suitable for Chinese in foreign language education is primarily shaped by political viewpoints. No reliable empirical data is available. Traditional characters tend more often to convey etymographic information whereas simplified characters might be easier to recognize visually and undoubtedly are easier and quicker to write by hand as well.

In general it seems useful to focus on one of the two systems when writing whilst acquiring passive reading competence in the other (which also enables digital writing in both systems using Pinyin). With regard to the EBCL project, it was agreed upon that the learner should be aware of both systems. However, even though a familiarity with both systems is desirable, it is regarded as sufficient to have reading and writing competence in only one of the two systems.

4.5. Consequences for teaching reading of Chinese in contrast to European languages

As criticised by Lutjeharms in 2004, until now, reading intention and motivation have been hardly considered in cognitive psychology research on reading, although the reading intention has a strong and decisive influence on the reading process. This is especially true for Chinese language education, where common reading tasks consist almost exclusively of fictitious dialogues (which, as mentioned under 4.3., are not written texts per se) or very simplified stories and mainly informative non-fictional texts. Those “texts” by no means represent the range of written material which learners in an everyday Chinese-speaking environment find themselves confronted with. Compared to similar tasks in foreign languages using phonographic writing systems, reading tasks in traditional Chinese textbooks at the elementary level show a significantly smaller vocabulary. This is due to 1) the characteristics of the Chinese script, 2) the above mentioned difficulties with proper names, and 3) the fact that the methodical approach of confronting the learner with unknown words and characters (especially in the context of the learning target of global reading comprehension) is, to this day, still unknown in language teaching in China: unknown vocabulary is always presented in dry lists in order of their appearance in the respective text, without reference to their frequency, their usage in colloquial vs. written Chinese, or to their morphemes, synonyms or antonyms.

This combination of the special characteristics of the writing system in combination with a traditional, strongly receptive teaching approach leads to an oversimplification of reading comprehension in Chinese language instruction, which does not correspond to the intellectual capacities of the learners,
be it secondary school pupils or adults. So-called “reading tasks” at elementary level are focussing neither on the addressee nor the text, but merely on a certain progression.⁶

For developing benchmarks for Reading Chinese, the functional approach of the EBCL project needed to take all crucial problems with reading competence mentioned above into account. This implied that any reading task had to be assigned only with authentic written text material.

5. Conceiving descriptors for reading and orthographic competence in the EBCL project

5.1. Developing the EBCL Descriptors for “reception written” (A1-A2)

In its main part of work, the EBCL project has developed or adapted more than 120 can-do-descriptors with examples on listening comprehension, speaking and spoken interaction as well as more than 60 descriptors on reading and writing competences, all with various examples in Chinese language. In accordance with the focus of this paper, the following part of this essay deals with the development of the descriptors for “reception written” (i.e. reading) as an example for the general development of the whole EBCL project.

When developing the EBCL can-do descriptors for “reception written” the following resources were referred to:

- Common European Framework of Reference for Languages (Council of Europe 2001),
- Bank of descriptors for self-assessment in European Language Portfolios “ELP” (Günter Schneider and Peter Lenz, University of Fribourg/Switzerland, 2004),
- European Association for Quality Language Services bank of descriptors (EAQUALS) which provides descriptors for “plus levels” as well as strategies,
- Japan Foundation ‘Can do’ statements: Japanese Standard for Japanese Language education based on the CEFR.

The CEFR scales and bank of descriptors by Schneider and Lenz served as the main resource because they had already been validated.

⁶ There are a few rather interesting exceptions (examples to be found in the New Practical Chinese Reader (ed. by Liu Xun 2001ff., Workbook, at the end of each lesson) or in the Chuji Hanyu yuedu kecheng textbook by Zhang Shitao), where learners are confronted with original text material such as public notices, train tickets, road maps, classified ads or websites and can test their (unavoidably) selective reading-understanding ability with authentic material.
In a first step, the A1-A2 can-descriptors of the illustrative scales for reception written as proposed in the CEFR were all adopted and it was then further examined whether

1. prelevels and sublevels resp. intermediate levels had to be introduced,
2. existing CEFR descriptors had to be modified and/or specified,
3. ELP descriptors from the bank of descriptors had to be added, and
4. completely new descriptors had to be created.

5.1.1. Prelevels and sublevels

It became evident soon that, in contrast to the levels of the CEFR, it was necessary to introduce sublevels (A1+, A2+) in order to support learner’s motivation. For reception and production written language, due to the characteristics of the Chinese writing system even a pre-level below criterion level A1 was necessary: The knowledge and competence of dealing with Chinese characters that is needed at the beginning stage required an elementary level in a separate section called “graphemic competence” (comparable with the marginal “orthographic control” section of CEFR). This level, which was labelled A1.1, had to deal with the pre-literal stage of acquiring knowledge about character composition, stroke order and radicals as a prerequisite for any reading and writing skill (see 5.3.).

5.1.2. Modifying existing CEFR descriptors

Since the CEFR has mainly been developed for European languages, some descriptors refer to features that are shared among Indo-European languages and are advantageous for the European learner, e.g.

Can understand short, simple texts containing the highest frequency vocabulary, including a proportion of shared international vocabulary items (RW1: Overall Reading Comprehension, A2.1).

However, as already mentioned above (4.1.), Chinese learners can hardly rely on any cognates and/or “international vocabulary items”. Therefore, some descriptors like these have been modified, here by deleting “including a proportion of shared international vocabulary items”.

Proper names and numbers are usually the most salient words for beginning readers of phonetic writing systems. In Chinese, on the other hand, proper names are often written with less frequently used characters and are therefore acquired at a later stage. Hence, the number of “familiar names” mentioned in CEFR becomes very limited concerning Chinese, as each name first needs to become “familiar” in its sinographemic form:
Can understand very short, simple texts a single phrase at a time, picking up familiar names, words and basic phrases and rereading as required (RW1: Overall Reading Comprehension, A1).

5.1.3. Adding European Language Portfolio descriptors

Some of the existing CEFR descriptors were considered not to be sufficiently specific and precise; therefore European Language Portfolio (ELP) descriptors were drawn from the bank of descriptors (Schneider and Lenz 2004).

Example:

I can skim small advertisements in newspapers, locate the heading or column I want and identify the most important pieces of information (price and size of apartments, cars, and computers) (RW3: Reading For Orientation, A2.2).

In some instances, the ELP descriptors were even further modified to be more specific:

I can follow clear, step-by-step instructions on equipment encountered in everyday life such as simple food preparation instructions (e.g. instant noodles), public telephone, taking out cash or buying a drink from a machine mainly relying on visual support (RW5: Reading Instructions, A2).

5.1.4. Creation of completely new descriptors

When existing CEFR descriptors as well as ELP descriptors were considered insufficient, sometimes new descriptors were created, e.g. “I can understand simple written messages concerning appointments (time and date), e.g. in sms phone messages, emails etc.” (RW2: Reading Correspondence, A1) and “I can find time and price related information encountered in public places (price tags, opening hours)” (RW3: Reading for Orientation, A1). These newly created descriptors have not been validated yet and were therefore proposed merely on a tentative base.

Finally, textual samples which were mainly taken from real-life situations were added to every descriptor in order to illustrate the reading task. To do this, there was made a compilation of photos and written documents from a wide range of text types and topics. These were then examined as to their suitability for a certain language level and assigned intuitively to an existing CEFR descriptor.

After the texts had been assigned, reading tasks were formulated for various example texts. Those tasks made it very clear that the competence to be described does not mean understanding every detail, but very often encompassed either global understanding or selective reading.
5.2. General conclusions concerning reading competence

It was agreed upon that the CEFR descriptors for “reading competence” can, in principle, be applied to the Chinese language. Yet, three crucial limitations must be mentioned:

1. In addition to the necessary vocabulary, the learner also has to be familiar with a corresponding amount of characters. Whereas all European languages have close and mostly regular sound-grapheme connections, the connection between spoken language and written graphemes is comparably weak in Chinese, so that in order to acquire reading competence in Chinese, a phonetic system as well as a largely independent graphemic system has to be mastered at least passively. This leads to a significantly longer learning time in comparison to languages using an alphabetic writing system.

2. The category “familiar names” which can be found several times within the CEFR must be relativized to the extent that even the names of persons and places which are written in the same way in all languages using an alphabetic writing system cannot automatically be considered “familiar names” in Chinese, which means that reading tasks containing such proper or place names must be rated much more difficult than is the case in the CEFR.

3. Reading tasks at this level usually comprise the targeted search for information (selective reading) or the comprehension of a certain text type (global reading comprehension). Other than that, the assumption was made that the acquisition of Chinese reading ability at this level of competence mainly focusses on the spoken language (口头语 kǒutóuyǔ) and the characters required for that. Learners at beginners’ level (A1/A2) are hardly able to understand authentic written documents due to the amount of characters primarily used in written language, a vocabulary which the learners have not yet acquired at this level (compare the Taiwanese CFL exam TOCFL 2012). In view of the use of Chinese written language in authentic written texts (as mentioned in 4.2.), it was agreed for the EBCL project that lexemes of the Chinese written language in the sense of shūmiànyǔ (书面语) were only to be considered relevant for level B1. As far as reading competence at level A2 is concerned, only a few lexemes that are somewhat frequent especially in the context of information signs should be assigned to this level (此 cǐ here, this /勿 wù do not).

As a consequence of the fact that several written text types were considered to be more difficult than they are in languages included in the CEFR, and because of the limited knowledge of characters at A1-B1 level, it was agreed that the use of dictionaries or electronic tools are needed to solve problems with reading tasks from level A1+ on (see 5.3.1.).
5.3. Orthographic/Graphemic control as prerequisite for reading and writing competence

As previously mentioned, there was a consensus that the characteristics of the Chinese writing system as described above led to certain limitations for the descriptors of the reference framework regarding reading competence and that the desired competence levels could only be reached much later than with European languages. Furthermore, it soon became obvious that certain basic requirements had to be defined for Chinese, which seemed so self-evident for the European languages that the reference framework, focussing exclusively on languages using an alphabetical writing system, did not pay any attention to them. This fact led to the implementation of the aspect of “orthographic control” as a basic precondition for dealing with the Chinese written language.

In the scope of “linguistic competences”, the CEFR also describes “orthographic control”. Since the Chinese orthography consists of a set of rules of disproportionately higher complexity than the orthography of languages using an alphabetic writing system, this part had to be significantly expanded and modified. On the one hand, there was the central question of the role of the alphabetic transcription system Hanyu Pinyin in describing orthographic competences. On the other hand, the graphemic system of the Chinese script itself contains a lot of “orthographic” features (strokes, stroke order, character structure, sub-graphemes, radicals, punctuation marks and the existence of two concurring writing systems within the Chinese language environment), the relevance of which all had to be taken into account individually. So due to the many special characteristics of the Chinese language, the CEFR descriptors on “orthographic control” had to be significantly modified and expanded.

As said, the competences needed for a basic reading and writing knowledge of Chinese called for an A1.1 level even below level A1 for reading and writing, which not only described competences, but also named strategies that could help learners to familiarize effectively with the Chinese writing system. Based on the experiences of various teachers from all institutional areas, the EBCL project decided on expanding “orthographic control” to “orthographic/graphemic control” and dividing this scale into the two separate competence areas of “sinigraphemic competence” (SGC) and “Hanyu Pinyin reading and writing competence” (CPY).

The elements added in the section “suggested knowledge and strategies” should thus be understood as recommendations. On the sinigraphemic side, they also include basic knowledge about stroke order, signific and phonetic character components and, at level A2, the awareness of four different construction principles in which iconicity, semanticity and phoneticity all play a role to varying extents.
5.3.1. EBCL Can-do Statements (CDS) for Graphemic/Orthographic Control (GC) concerning Chinese characters

<table>
<thead>
<tr>
<th>EBCL / CEFR Level</th>
<th>EBCL descriptors</th>
<th>Suggested knowledge and strategies for graphemic/orthographic control</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1.1</td>
<td>E-SGC-A1.1-1</td>
<td>- knows the basic principles of stroke order and stroke direction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- is aware that many characters can be further divided into smaller components</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- knows the main rules of composition of complex characters</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- can distinguish simple (独体字) from combined (合体字) characters</td>
</tr>
<tr>
<td></td>
<td>E-SGC-A1.1-2</td>
<td>- knows the difference between words, characters and components and does not confuse these three categories</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- knows that in di- or polysyllabic words nearly every character has a morphemic function</td>
</tr>
<tr>
<td></td>
<td>E-SGC(R)-A1.1-1</td>
<td>Can tell whether a given text is written in Modern Chinese or Modern Japanese</td>
</tr>
<tr>
<td></td>
<td>E-SGC(R)-A.1.1-2</td>
<td>Understands the function of common Chinese punctuation marks that have the same function as in European languages (.,:!?“”)</td>
</tr>
<tr>
<td>A1</td>
<td>E-SGC-A1-1</td>
<td>Can copy familiar words and characters as well as unfamiliar characters in simple signs or names</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- knows how to use Pinyin input method (refer to Hanyu Pinyin Reading and Writing competence)</td>
</tr>
<tr>
<td>A1</td>
<td>E-SGC-A1-3</td>
<td>Can write down his/her gender, nationality and any date (like one’s birthday) by hand</td>
</tr>
<tr>
<td>A1+</td>
<td>E-SGC-A1+-1</td>
<td>- can tell the difference between semantic components and phonetic components of Chinese characters in general</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Knows how to use dictionaries or electronic devices to look up lexical items of unknown characters</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>EBCL / CEFR Level</th>
<th>EBCL descriptors</th>
<th>Suggested knowledge and strategies for graphemic/orthographic control</th>
</tr>
</thead>
</table>
| A1+               | E-SGC(R)-A1+-1   | Can tell the meaning of about 15 different semantic components even if they are parts of unknown characters. Examples:  
|                   |                  | じ/人 person  
|                   |                  | け/水 water, liquid  
|                   |                  | え 土 earth, locality  
|                   |                  | わ/女 woman / female  
|                   |                  | し/火 fire / (verbs of cooking)  
|                   |                  | か/木 wood  
|                   |                  | て/手 hand / (verbs of doing sth with hand(s))  
|                   |                  | ま/走 (verbs of fast movement)  
|                   |                  | く/口 mouth / (particle of orality)  
|                   |                  | い/言 words / (verbs of speech)  
|                   |                  | て/食 food  
|                   |                  | た/心 heart (words of emotionality)  
|                   |                  | げ/月 moon / flesh (parts of the human body)  
|                   |                  | く/貝 shell / (words of money and trade)  
|                   |                  | は/糸 silk 日 sun / time  
|                   |                  | た/金 metal  
|                   |                  | ま/石 stone &&/艹 grass, plant  
|                   |                  | く/宀 roof  
|                   |                  | - is familiar with certain graphic variants of important semantic components like  
|                   |                  | じ/人  
|                   |                  | て/手  
|                   |                  | て/食  
|                   |                  | た/心  
|                   |                  | い/言  
|                   |                  | け/水 |
| A1+               | E-SGC(R)-A1+-2   | Can tell whether a given text is written in Chinese traditional characters or mainland characters.  
|                   |                  |  |
| A2                | E-SGC-A2-1       | Can copy short phrases on everyday subjects by hand without hesitation - e.g. names of institutions.  
|                   |                  |  |
| A2                | E-SGC-A2-2       | Can write by hand with reasonable graphemic accuracy so that the written characters are understood by other readers or by an OCR device  
|                   |                  | - knows examples of characters that share the same phonetic components  
|                   |                  | - knows the four main categories of characters (in his/her mothertongue): pictographic / indicating (象形字 and 指示字), semantic compound (会意字), phonetic loan (假借字) and semantic-phonetic (形声字)  
|                   |                  |  |
From level B1 on, orthographic competence descriptors are believed to be quite similar to the ones postulated in the CEFR.

### 5.3.2. EBCL Can-do Statements (CDS) for Graphemic/Orthographic Control (GC) concerning Hanyu Pinyin

Apart from the competences related to the Chinese writing system, the transcription system Hanyu Pinyin plays an important role for learning correct Chinese pronunciation as well as for writing Chinese texts on a computer. Hanyu Pinyin, based on the Latin alphabet, has become the leading transcription system for the Chinese language. According to findings of the EBCL project team, there are no more courses or textbooks for CFL courses in Europe that do not use Hanyu Pinyin. Yet, as the written Chinese language is still using characters, standard keyboards also allow for other input methods, and because a purely imitative learning of the Chinese pronunciation by means of another transcription system is conceivable in individual cases, Hanyu Pinyin competence was not defined as a stringent necessity but rather as an explicit recommendation for European learners.

<table>
<thead>
<tr>
<th>EBCL / CEFR Level</th>
<th>EBCL descriptors</th>
<th>Suggested knowledge and strategies for graphemic/orthographic control</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2</td>
<td>E-SGC(R)-A2-1</td>
<td>Understands the function of Chinese punctuation marks <em>dunhao</em> (、), <em>shenglüehao</em> (......), <em>shuminghao</em> (《》) and <em>zhuanminghao</em> (﹒﹒﹒)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EBCL / CEFR Level</th>
<th>EBCL descriptors</th>
<th>Suggested knowledge and strategies for graphemic/orthographic control</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1.1</td>
<td>E-GOCPY-A1.1-1</td>
<td>Can read and understand familiar words and sentences when written in Pinyin with tone marks</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- knows the function of letters and tone marks of the Pinyin alphabet as a transcription system for the Chinese language</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- can link Hanyu Pinyin with the corresponding Chinese characters if a word or text is given in both writing systems</td>
</tr>
<tr>
<td>A1.1</td>
<td>E-GOCPY-A1.1-2</td>
<td>Can write down words and short sentences in Pinyin with mostly correct tone marks</td>
</tr>
</tbody>
</table>

---

7 CFL = Chinese as a Foreign Language, in contrast to CHL = Chinese as a heritage language courses that focus on literacy of mother tongue speakers, where Pinyin is not necessarily needed.
6. Outlook

Through the growing digitalization of our world, reading of Chinese language texts, which has always been the core competence of Chinese Studies, has become significantly easier. Online dictionaries and electronic translation tools help to understand Chinese text material, a task which used to be a lot more time consuming. Unknown characters can be looked up much quicker and more easily thanks to various applications, search and input methods. Such digital reference material plays a decisive role for reading Chinese texts. Students must learn to use and apply them, and Chinese language reading instruction can no longer ignore those facilitations. Also, research competence using modern web-based communication as well as the corresponding strategies will have to be included in the respective curriculums at a rather early stage (A2).

Yet, in conclusion, it cannot be stressed enough that an independent, comprehending reading of authentic Chinese texts, be it specialised texts or literary ones, at an adequate speed, requires comprehensive exercise, experience and cultural background knowledge which by far exceeds the learning effort of reading any European language.

Finally, it is hoped that the EBCL project that was generously funded by the European Commission can be a further step in the scientific discussion of how to define language competences in Chinese as a Foreign Language. A German edition of EBCL is already on its way.
References and further readings


CEFR see Council of Europe


Hirsh, D., and P. Nation (1992), What vocabulary size is needed to read unsimplified texts for pleasure?, in: *Reading in a Foreign Language*, 8, 2, 689–696.


Laufer, Batia (1997), The lexical plight in second language reading: Words you don’t know, words you think you know, and words you can’t guess, in: James Coady and Thomas Huckin (eds), *Second language vocabulary acquisition*, Cambridge: Cambridge University Press, 20–34.


Lutjeharms, Madeline (2010), Der Leseprozess in Mutter- und Fremdsprache, in: Madeline Lutjeharms and Claudia Schmidt (eds), Lesekompetenz in Erst-, Zweit- und Fremdsprache, Tübingen: Narr, 11–26.


孙德金 Sun, Dejin (2012), 现代书面汉语中的文言语法成分研究 Xiandai shumian Hanyu zhong de wenyan yufa chengfen yanjiu, Beijing: Shangwu Yinshuguan.


阅读能力与字素能力：中文文字系统与其在 EBCL 项目对发展汉语作为外语的能力基准的影响

摘要

欧洲共同语言参考标准（CEFR2001）已是全欧洲广为使用的外语教学纲领和评量基准。为了在 CEFR 的基础上发展对外汉语教学的准则及能力说明，欧盟在 2010 年至 2012 年间推动和赞助欧洲汉语能力基准项目（EBCL）。在简单说明该以 A1 与 A2 等级为主（在网上已发表过的）项目内容后，将藉由一些例子来讨论框架等级描述为适应不同汉语水平的阅读能力而产生的不同观点与问题。此外，也将详细说明在此项目中的汉字系统以及汉语拼音的内在能力。
Lesekompetenz und Graphemkompetenz: Zu Auswirkungen des chinesischen Schriftsystems auf die Entwicklung von Deskriptoren im Rahmen des Projekts „European Benchmarks for the Chinese Language“ (EBCL)

Zusammenfassung