High-level proficiency in late L2 acquisition

Relationships between collocational production, language aptitude and personality

Fanny Forsberg Lundell & Maria Sandgren
Stockholm University / Södertörn University

The aim of the present exploratory study was twofold. The first was to investigate how indicators of high-level proficiency (collocations and grammaticality judgment) related to aptitude in late French L2 learners. Results showed a significant positive correlation between collocations and performance on the LLAMA D (Meara 2005). The second question concerned how personality relates to indicators of high-level L2 proficiency (collocations and grammaticality judgment). Two personality dimensions in the Multicultural Personality Questionnaire (Van der Zee & van Oudenhoven 2000) were significantly and positively correlated with scores for collocations and the LLAMA D. The preliminary findings suggest that collocations are a valid measure for high-level L2 proficiency and that it is necessary to consider both personality and social-psychological factors when predicting successful L2 learning.

1. Introduction

Researchers have shown that formulaic language, specifically collocations, pose problems at advanced levels of acquisition (e.g. Siyanova & Schmitt 2008; Bolly 2008; Durrant & Schmitt 2009; Mizrahi & Laufer 2010). In a study by Forsberg Lundell and Lindqvist (submitted), different aspects of lexical knowledge were tested in a population of very advanced L2 users of French (L1 Swedish) whose length of residence in France (LOR) ranged from five to 35 years. The participants performed in a nativelike way on a C-test and on a social routine test, but performed significantly poorer than a native speaker control group on a receptive deep knowledge test (mainly including word association) and a collocation test. It was shown that individual variation was particularly high on the collocation test. Interestingly, LOR did not explain the variation in any of the lexical dimensions investigated.
It was thus deemed relevant to go further and reflect on factors contributing to high-level proficiency among late L2 learners from an interdisciplinary point of view, namely linguistics and psychology. Two lines of thought developed. One concerned measurements that would accurately assess high-level L2 proficiency and its relationship with language aptitude. A second concerned personality characteristics that would be positively associated with high L2 proficiency among adult learners.

2. Collocations and high-level SLA

There is increasing evidence in the literature of the specific difficulty that collocations pose for L2 learners (e.g. Mizrahi & Laufer 2010). Findings suggest that these difficulties are both psycholinguistic and cultural. It is important to situate collocations within the theoretical framework of formulaic language. As many researchers have now stated (e.g. Schmitt & Carter 2004; Eyckmans 2009; Forsberg 2010), formulaic language is subdivided into a plethora of subcategories with different forms and functions. Collocations, especially verb-noun combinations (make a decision) and adjective-noun combinations (black coffee), have been a privileged category in L2 research, probably because it constitutes a relatively well-defined category that can be measured both in spontaneous production and, perhaps more importantly, in tests.

In Wray’s (2002) view, formulaic language is useful mainly for two reasons. It provides production relief for the speaker and decoding relief for the hearer (as evidenced by, e.g. Underwood, Schmitt & Galpin 2004; Conklin & Schmitt 2008), which, thus, makes it cognitively economical for both parties. It also implies social advantages, since ready-made sequences are community-wide in use and convey a common identity. The cultural dimension of formulaic language has also been highlighted (e.g. Skandera 2007), implying that formulaic language learning, which collocations are part of, is also part of a process of acculturation. Of particular interest for the present study is the suggestion by Dörnyei, Durow and Zahran (2004) and Adolphs and Durow (2004) that formulaic sequence learning is a socially loaded process.

However, the cultural dimension of formulaic language acquisition has generally been ignored in SLA studies, which tend to concentrate more on pure linguistic mastery of collocations. The review below will focus on findings regarding advanced/high-level learners’ mastery of collocations. This concerns receptive knowledge, productive knowledge, and use in writing and speech. Advanced and high-level are subjective terms. The review is based on studies that report studying advanced or high-level learners/users.
There seems to be a common misconception that recognition of collocations at advanced levels of L2 learning is less problematic than productive knowledge and use. Both Bonk (2000) and Gyllstad (2007) found strong correlations between a vocabulary size test and a receptive collocation test, which suggests that recognizing collocations is not more problematic than recognizing single words. This suggests a less problematic role for collocations than reported elsewhere in the L2 literature. However, as already hinted, producing collocations in tests, writing and speech is more challenging for L2 learners. Several early studies found that advanced university students performed significantly worse on cloze tests involving collocations and translations tasks (Bahns & Eldaw 1993; Farghal & Obiedat 1995). Two recent studies have tested productive collocational knowledge in highly advanced L2 speakers who, according to self-identification or researcher judgment, approach near-native levels. Mizrahi and Laufer (2010) found that speakers who self-identified as near-native in L2 English and had had extensive exposure to English performed in a nativelike way on a productive vocabulary size test. Only a minority of them performed in a nativelike way on a productive collocation test. Forsberg Lundell and Lindqvist (submitted) tested highly advanced L2 French speakers on four different vocabulary aspects. Just as in Mizrahi and Laufer (2010), significant differences were found between native and non-native speakers on a similar productive collocation test.

In written and spoken production, several tendencies emerge. Some studies show that, in the quantity of collocations produced, no significant differences are found between advanced learners and native speakers (Nesselhauf 2005; Siyanova & Schmitt 2008; Forsberg 2010) (for the most advanced group of L2 users). Nonetheless, qualitative differences are often found in such areas as L1 transfer (Granger 1998; Nesselhauf 2005; Laufer & Waldman 2011), degree of restriction of the collocation (Howarth 1998; Nesselhauf 2005; Bolly 2008), frequency of collocations used (Durrant & Schmitt 2009), and type of category used (Forsberg 2010). Another trend emerges in studies by Laufer and Waldman (2011) and Mizrahi and Laufer (2011), which both conclude that advanced L2 English learners fail to produce even similar quantities of collocations. Laufer and Waldman (2011) also report persistent collocational errors in written production.

To conclude, findings as to the extent to which advanced learners are capable of producing formulaic language at nativelike levels are mixed. Some studies suggest that, at a quantitative level, nativelike attainment is possible. Others do not report nativelike attainment for either quantity or quality. One explanation could be that there is a lack of consistency in proficiency measurements, i.e. we cannot be certain that the highly advanced learners in all of these studies are really at the same level of L2 proficiency. Another possible explanation is that typological
similarities between L1 and L2 will have a positive effect on L2 speakers’ use of collocations. However, all studies seem to agree that non-native and native speakers differ in qualitative aspects of collocation production. Since it has been argued that collocations and formulaic language are both cognitively and culturally important, it seems plausible to turn to psychological and social factors to explain this variation in high-level L2 proficiency.

3. Social-psychological challenges for the adult L2 learner

To attain high-level L2 proficiency, the learner has to make a major, extended effort. This constitutes a high-stakes project for the individual and may even exhaust cognitive and emotional resources. The body of literature on L2 learners inspired us to take a closer look at factors that will constitute challenges for acquiring high-level proficiency.

Judging from the reviews of findings on individual differences (e.g. Dörnyei 2005), it appears that language aptitude has been one of the most successful individual factors in predicting second language success. Lately, the role of aptitude for naturalistic high-level achievement has been highlighted (e.g. Abrahamsson & Hyltenstam 2008; Granena & Long 2013). In particular, non-traditional measures of aptitude, stemming from cognitive psychology, such as phonological short-term memory, have been linked to high-level L2 achievement and to the learning of formulaic language (Skrzypek 2009; Bolibaugh & Foster this volume). If we wish to investigate how individual factors influence the learning of collocations in late L2 acquisition, it is necessary to investigate the role of language aptitude.

However, it is believed that aptitude alone cannot explain high-level L2 proficiency. Second language acquisition is not only a cognitive challenge, but also challenges an individual’s social identity. This is the case for the type of L2 users included in our study i.e. long-term residents in the TL community. While learning an L2, the individual will go through a process of re-socialisation and consistently negotiate his or her social identity and varying degrees of affiliation to different social groups. As a result, social-psychological factors such as integrativity (Gardner 2001) and ethnic group identification (Gatbonton & Trofimovich 2008) may affect the individual’s development of his or her potential.

3.1 Aptitude and SLA

Language aptitude has been one of the most successful individual factors in explaining L2 research findings. Traditionally, it is defined as the “strengths
individual learners have – relative to their population – in the cognitive abilities information processing draws on during L2 learning and performance in various contexts and at different stages" (Robinson 2005:46). In addition, aptitude tests have, during the past 50 years or so, conceived of aptitude in terms of the rate at which learning takes place (Carroll 1981). According to Robinson (2005), the creation of aptitude tests in the 1950s–1970’s was mostly driven by an eagerness to find tests that would predict rate of individual learning for selection purposes. For a variety of these tests, the MLAT (Carroll 1981), PLAB (Pimsleur 1966) and DLAB (Peterson & Al Haik 1976), studies have yielded strong correlations between success in instructed language learning and scores on the aptitude test in question (Sawyer & Ranta 2001; Dörnyei & Skehan 2003). A newer test, based on the same theoretical foundations as the MLAT, is Meara’s (2005) LLAMA test. The standard aptitude components of these tests have been phonetic coding ability, grammatical sensitivity, rote learning ability and inductive learning ability. For an overview of research on aptitude, see Granena (this volume).

The literature often stresses that these tests were not intended to predict high-level L2 use which involves pragmatic abilities that they do not measure. It has also been argued that these tests do not measure the capacity to learn a language from mere exposure, which would require more implicit learning, and that, rather, they tap into language analytic ability. Since the 1990’s, new reconceptualizations of aptitude have included working memory and phonological short-term memory as relevant subcomponents of the construct (Miyake & Friedman 1998; Sawyer & Ranta 2001; see also Kormos this volume, and Bolibaugh & Foster this volume).

Convincing results have been reported concerning the effects of these components on specific aspects of language proficiency, such as a strong correlation between phonological sequence learning and vocabulary learning (Speciale, Ellis & Bywater 2004), phonological short term memory and collocation learning (Skrzypek 2009), and more general language skills, such as reading, writing, listening and speaking, and phonological short term memory in Kormos and Safir (2008). Rota and Reiterer (2009) also investigated the relationship between cognitive variables and pronunciation talent and performance. Apart from measuring working memory (which yielded significant correlations with pronunciation talent), they also included empathy as an independent variable, drawing on Guiora, Brannon and Dull’s (1972) early work. Interestingly, significant correlations were observed between empathic skills (as measured by the E-scale, Leibetseder, Lairreiter, Ripler & Köller 2001) and different pronunciation-related measures, including pronunciation performance ($r = 0.29, p = 0.013$) and phonetic coding ability as measured by the MLAT ($r = 0.20, p = 0.04$).
Another strand of research, which has gained more ground recently and may be linked to phonological capacities, is that of relationships between musical aptitude and language proficiency. Once considered insignificant, there is now increasing evidence of such a relationship. Gilleece (2006) found low to moderate correlations, which have been corroborated in neuroscience studies (e.g. Milovanov, Huotilainen, Välimäki, Esquef & Tervaniemi 2008). These studies have found a significant relationship between musical aptitude, English L2 pronunciation, chord discrimination ability and sound-change evoked brain activation. Nardo and Reiterer (2009) also found significant relationships between several measurements of musical ability, L2 pronunciation skills and scores on subcomponents of the MLAT.

For the present study, the most relevant findings in the L2 aptitude literature are those of DeKeyser (2000), Abrahamsson and Hyltenstam (2008) and Granena and Long (2013). All these studies found significant relationships between L2 attainment and language aptitude, especially in late learners. In DeKeyser’s study, both early and late Hungarian L1 – English L2 learners completed a grammaticality judgment test (GJT). They also completed a translated version of the words-in-sentences subtest of the MLAT. The study showed that, in the group of adult arrivals, aptitude scores were significantly correlated with GJT results, but this was not the case among learners who had arrived in the L2 environment before closure of the critical period. Abrahamsson and Hyltenstam (2008), who conceived their study as a follow-up to DeKeyser’s (2000), investigated the relationship between aptitude, as measured by Meara’s (2005) LLAMA test, and written and auditory GJT scores in early and late Spanish L1 – Swedish L2 learners. They found strong correlations in the late group, but also trends toward significant correlations in the early group, suggesting a more important role for aptitude in child SLA than the results reported by DeKeyser (2000). Granena and Long (2013) continued this line of inquiry while adding two more linguistic domains, namely pronunciation and lexis/collocations. Participants were 65 early and late Chinese L1 – Spanish L2 learners, and 12 native speaker controls. Unlike the two preceding studies, no significant correlation was found between auditory GJT scores and aptitude, as measured by the LLAMA test, in the late group, whereas statistically significant correlations were obtained in both of the other linguistic domains in the late, but not in the early, group. Interestingly, scores on the lexis/collocation test correlated significantly with the two auditory subtests in the LLAMA: LLAMA D (sound recognition) \( r = .46 \) and LLAMA E (sound-symbol correspondence) \( r = .36 \). These are two aptitude subtests where participants’ phonological capacities (phonetic memory) could have played the greatest role.
3.2 Personality and SLA

The next section provides a brief review of personality characteristics, identifying cognitive styles and personality dimensions that may be favourable for engaging in learning and interacting with an unfamiliar culture, i.e. the dominant L2 community. The selection of references is diverse, as the field is rather scattered and results are mixed. Additionally, personality traits of importance for intercultural adjustment effectiveness are included. Behavioural tendencies, reflected in personality traits, influence habits that may have an impact on intercultural effectiveness. Personality traits mirror what a person will do, whereas cognitive styles reflect what an individual can do (Chamorro-Premuzic, Furnham & Moutafi 2004). Even if cognitive styles are primarily defined as innate aptitudes in psychology, they may relate to personality traits, as in the Five Factor Model (FFM; Costa & McCrae 1992). For example, openness to experience is positively related, and conscientiousness negatively related, to IQ measures.

3.3 Cognitive styles

Two cognitive styles, tolerance for ambiguity (TA) and field-dependence-independence (FDI), have been investigated for their importance for language learning and proficiency. TA and FDI represent two different modes of cognitive functioning with respect to how an individual perceives, acquires and processes information. In an early study of ‘The good language learner’, Naiman, Frölich, Stern and Todesco (1978) investigated TA, defined as the tendency to perceive and interpret ambiguous stimuli as sources of threat, using a test by Budner (1962). Ambiguous situations are characterized by complexity, novelty and insoluble issues. Thus, an individual with a high level of TA would experience discomfort and anxiety when confronted with threatening stimuli, and react by avoiding or rejecting the situation. Naiman et al. (1978) found that TA was related to language proficiency.

Although L2 research has not pursued the exploration of TA, TA could be of importance in the individual's motivation to create a positive bond with the L2 community, due to its link to ethnocentrism (see Giles & Johnson 1987). More recently, researchers have drawn attention to how TA can be used to assess adjustment in cross-cultural settings (Herman, Stevens, Bird, Mendenhall & Oddou 2010). Herman et al. further developed Budner’s scale and suggested four dimensions: (a) valuing diverse others, (b) coping with change, (c) dealing with unfamiliar situations, and (d) managing conflicting perspectives. It is likely that a person will adapt in a cross-cultural context if he/she has a positive attitude towards...
unfamiliar people, is able to cope with change and deal with strange situations (i.e. manage the unexpected), and, not least, if he/she has the ability to encompass various perspectives (i.e. understand that there are various ways to grasp what is happening).

3.4 Intuition and openness

Ehrman (2008) chose to categorize her sample as a group of "learners," and not as talents, implying that the role of language aptitude is reduced in favour of other factors influential in engagement in learning. At her disposal was a large sample of foreign service workers ($n = 3,145$), of whom only 2% had attained the highest proficiency level (Level 4) assessed by an Oral Proficiency Interview conducted at the Foreign Service Institute. This group was labelled "best language learners," while another group was labelled "good language learners." Using the Myers-Briggs Type Indicator (MBTI; Myers 1976), the individuals were assessed on the basis of four of Jung's personality types: extraversion-introversion, sensing-intuition, thinking-feeling and judging-perceiving, which can form a total of 16 psychological types. The MBTI is assumed to tap connections between one's personal style and professional specialization.

Results indicated that the best language learners were characterized by the combination Intuition-Thinking, although only the combination of Introversion-Intuition-Thinking-Judging was found significant. The Intuition-Thinking type is interested in mastering intellectual matters. He/she prefers analytical and strategic thinking, as shown in the drive to search for precision, especially lexical precision, including idioms. Regarding extroversion, individuals represented by the combination Extraversion-Intuition were also found in Level 4 (Ehrman 2008). In contrast, the Sensing type was most frequently found outside the best-learner-domain. These individuals rely on the concreteness of the world, i.e. observing and embracing present, physical and factual aspects.

The role of Intuition in high language achievement is underlined by Ehrman (2008). The explanation might be that intuition, as measured by the MBTI, implies a disposition to perceive patterns and relationships in information. Individuals scoring high on Intuition are inclined towards an open and tolerant attitude and prepared for constant change. They are future-oriented, seek hidden patterns, and make associations. In the case of the highly successful language learner, this would mean that "when it is necessary to adapt to unfamiliar ways of speaking or to pick up native-like ways of self-expression […] a tendency to perceive the world in intuitive ways is likely to be helpful" (2008:6). Ehrman also pinpointed the link between Intuition and Openness to experience in the Five Factor Model. The intuitive individual has an open-minded character and shares characteristics with individuals who rate
high on the scale of Openness to experience. These characteristics include curiosity, broad interest, creativity, originality and imagination. Later studies indicate that Openness correlates not only with Intuition, but also with other dimensions of the MBTI (see Furnham, Dissou, Sloan & Chamorro-Premuzic 2007).

3.5 Multicultural personality dimensions

As previously stated, research on dimensions of personality and L2 performance has focused on how individuals perceive and process information from external sources assessed by TA and MBTI. This line of research has been followed up in studies on intercultural adjustment with an important addition: indicators have been created of how individuals feel and act in a multicultural setting. There have been several attempts to identify psychological and behavioral correlates of intercultural adjustment and adaptation.

One of the more validated measures is the Multicultural Personality Questionnaire (MPQ), created by Van der Zee and van Oudenhoven (2000, 2001). They define multicultural competence in broad terms as "success in the fields of professional effectiveness, personal adjustment and intercultural interactions" (2000:293) and conclude that the Five Factor Model (FFM) was too broad to cover traits relevant for multicultural competence. The MPQ is designed for predictions of an individual's cross-cultural competencies on the basis of three criteria for successfully operating within new cultural environments: "the capacity to make things work," "a feeling of well-being in that environment," and "an interest and ability to deal with individuals from different cultural backgrounds" (2000:293). The instrument has, like the FFM, five scales, but those of the MPQ are intended to focus more narrowly on cultural empathy, open-mindedness, flexibility, social initiative and emotional stability. (For more detailed descriptions of the scales, see personality measures.)

The MPQ has demonstrated its predictive value for multicultural effectiveness among expatriates Peltokorpi (2008), business professionals (Peltokorpi & Froese 2011) and international students (Van der Zee & van Oudenhoven 2000, 2001; Leone, Van der Zee, van Oudenhoven, Perugini & Ercolani 2005). Cultural empathy, open-mindedness and flexibility are linked to the domain of intercultural interaction, whereas social initiative and emotional stability are associated with the domain of personal adjustment.

Leone et al. (2005) examined the relationship between the MPQ data and the FFM in terms of the usual five traits, and found a range of significant correlations. The strongest positive correlations were between cultural empathy and openness to experience, open-mindedness and openness to experience; and social initiative and extroversion. The strongest negative and correlations were between flexibility
and conscientiousness, and emotional stability and neuroticism. The findings, nonetheless, contradict the assumed specificity of the FFM, but the predictive value of the MPQ is still superior to that of the FFM for intercultural orientation and across cultures.

All five scales are expected to be unrelated to cognitive abilities, such as intelligence. Yet Van der Zee, Zaal and Pickstra (2003) found that cultural empathy, open-mindedness and flexibility showed positive but weak associations with verbal abilities assessed as abstract-verbal-logical (analytical and logical reasoning with verbal materials), verbal fluency (vocabulary), and language comprehension and productive language skills (text completion and jumbled sentences) (GITP; Tjoa 1965). In the same study, openness to experience in the FFM was also related to verbal abilities. It seems that both open-mindedness and openness demonstrate a more intellectual orientation than mere openness to cultural phenomena. Less is known about what cultural empathy and flexibility have in common with respect to verbal abilities.

In the present study, the MPQ emerged as the first choice measure, due to its high predictive value on a range of criterion variables related to multicultural effectiveness. The chosen sample was assumed to be socially and professionally well-integrated, based on their answers to the background questionnaire. To include measures of cognitive styles appeared less appropriate, as L2 studies tend to use various criterion variables, such as learning strategies, speech production and examination grades. Consequently, most studies have yielded inconsistent outcomes. Furthermore, it is difficult to draw any conclusions regarding the influence of personality variables on L2 performance, as the linguistic variables are often too general to be affected only by individual differences, such as grades, fluency and reading comprehension.

4. Aims of the study

Following the preceding argumentation, the present exploratory study had two major aims. The study aimed to investigate relationships among collocations, grammaticality judgments and language aptitude, as measured by the LLAMA test (Meara 2005; see also Granena, this volume) in a sample of highly proficient late L2 learners. Another aim was to investigate relationships between personality as assessed by the Multicultural Personality Questionnaire (MPQ; Van der Zee & van Oudenhoven 2000) and linguistic abilities. Further exploration of a sample of highly proficient late learners of French (see Forsberg Lundell & Lindqvist submitted) was undertaken. The individuals were contacted for an extended investigation of aptitude and personality measurements. Data on background
variables, interests, informal and formal learning of foreign languages were also collected, although only demographic data are presented here.

5. The empirical study

5.1 Research questions

In view of the study’s aims, the following research questions were posed:

1. How are indicators of high-level proficiency (collocations and grammaticality judgments) related to aptitude in L2 French among late learners?

2. How are personality dimensions related to indicators of high-level proficiency (collocations and grammaticality judgments) in L2 French among late learners?

6. Method

6.1 Participants

Participants were 13 late Swedish L1 – French L2 speakers, with 12 women and one man. Their average age was 38 years, ranging from 28 to 58. A majority were living with a French partner and had children at home. All of them had started learning French after the age of 12. Their average LOR in France was 14.5 years, ranging five to 35 years. The average age of arrival was at 20. Both LOR and participants’ age varied significantly within the group. In Forsberg Lundell & Lindqvist (submitted), LOR did not correlate with scores on any of the linguistic tests, one of which was the collocation test. This was probably due to the fact that all participants had spent at least five years in France, and it has been suggested that the effect of LOR tends to diminish after five years in the TL community (Cummins 1981; Piske, MacKay & Flege 2001). Abrahamsson and Hyltenstam (2009) prefer the figure of 10 years, after which they suggest that an interlanguage tends to stabilize.

Most participants had studied French formally, although length and quality of instruction varied. Some had studied French as a foreign language for at least six years in school, and some had also taken language courses in France. Others had studied French in school for only a few years. All reported having mainly learnt French through language use while living in France. Some worked in bilingual settings (French/Swedish) whereas others in entirely French settings.
All of them had completed at least upper secondary education and a majority had university degrees.

In sum, this group of L2 late learners can be described as well educated, socially and professionally integrated individuals who had decided to move to France of their own free will. Most claimed they had an interest in French language and culture even before moving. For these reasons, it seems we were dealing with a group of immigrants rarely investigated in the SLA literature, i.e. highly motivated L2 users living in favourable social circumstances.

The participants were recruited mainly from lists of participants in earlier projects on high-level L2 French by one of the authors, chiefly through advertisements at the Swedish Institute and the Swedish Church in Paris. A few of them were recruited through snowballing, that is, new participants were recruited through earlier recruited participants. The linguistic tests were completed in Paris in December, 2010. The collocation test and GJT were part of a larger battery of linguistic tests, both oral and written, lasting approximately two hours. None of the individual tests was timed, but participants were told that they were to finish all of them in two hours. They were paid 20 euros for their participation. The LLAMA test and the MPQ, along with a questionnaire on background variables and language use, were administered in June and July, 2011. A French research assistant met all of the participants again for another two-hour session. The MPQ is a paper-and-pencil test, whereas the LLAMA is computerized. Participants received a gift card of 20 euros the second time around. They were informed of the overall aim of the study and assured confidentiality.

6.2 The aptitude tests
The aptitude test was the LLAMA (Meara 2005), thoroughly described in Granena (this volume) and used by Granena (this volume) and Granena and Long (2013), as well as by Abrahamsson and Hyltenstam (2008). Theoretically, it builds on the same constructs as the MLAT. The most notable difference, according to Meara, is its user-friendliness. The test is intended to measure four different components of language aptitude (to date), which also explains why the results of each subtest are accounted for individually. The four components are: LLAMA B (vocabulary learning), LLAMA D (sound recognition), LLAMA E (sound-symbol correspondence), and LLAMA F (grammatical inferencing). All the subtests are based on artificial or rare languages that are most likely unknown to the participants. The tests are downloadable and supplied by Meara and associates. It should be noted that Meara (2005) insists on the preliminary quality of the test, and that it should not be used in high-stakes situations.
6.3 Personality measurement

The intercultural personality traits were measured with a Swedish version of the MPQ (Van der Zee & van Oudenhoven 2000, 2001). The translation from English to Swedish was made by the authors of the present study. The MPQ is broadly based on Costa and McCrae’s (1992) FFM, but narrowed to predict traits relevant to multicultural effectiveness. Five dimensions of multicultural orientation are assessed using 91 five-point scales, from 1 (not at all applicable) to 5 (completely applicable). Higher scores indicate increased loadings on each of the five dimensions. The scales are Cultural Empathy (18 items, high scorers have an interest in other people and are sensitive to their feelings and beliefs); Open-mindedness (18 items, high scorers have an absence of prejudice); Social Initiative (17 items, high scorers tend actively to approach social situations and take the initiative); Emotional Stability (20 items, high scorers remain calm in stressful situations); Flexibility (18 items, high scorers adapt to new situations). In a review by Ponterotto (2008), criterion-related validity was summarized. Scores succeeded in predicting variance in criterion variables, such as well-being, health, coping, job satisfaction and multicultural activity, in a range of samples, such as university students and adult professionals in various countries and continents (Europe, North America).

In addition, participants answered questions on age, gender, socioeconomic status, civil status and educational level.

6.4 The collocation test

In the present study, the overarching aim is to gauge late L2 learners’ possibilities for attaining nativelikeness. It is therefore important to design the elicitation measures carefully, so as to avoid ceiling effects. It is also well-known that productive tests are more difficult than receptive tests (cf. Schmitt 2010). It was therefore decided to create a productive collocation test in French. Collocations can be conceived within a phraseological or a frequency-based framework. In the present study, as opposed to our earlier studies of free speech (Forsberg 2010; Forsberg Lundell & Lindqvist 2012), it was decided to use a frequency-based definition, since it offers a more objective identification procedure and is easily applicable to the selection of items for a test.

For the selection of items, Gyllstad’s (2007) procedure was followed quite closely. With respect to the format, we were inspired by Mizrahi and Laufer (2010), since they had also designed a productive test, whereas Gyllstad designed a receptive one. There follow the different steps taken to create the collocation test.
In conformity with Gyllstad (2007), it was decided to focus on verb–noun collocates, such as rendre justice (Eng. “do justice”). It was also decided to work with data from written French, since the test was to be taken in the written mode:

1. 150 frequent nouns were selected randomly from Tom Cobb’s frequency bands for written French. Gyllstad used words from low-frequency bands, such as 10K–14K, but since no such low-frequency bands exist for French, the 1K, 2K and 3K frequency bands from the Lexical tutor were used instead (http://www.lectutor.ca/freq/lists_download/). These frequency bands are based on the French language newspapers Le Soir (Belgium) and Le Monde (France).

2. The second step was to search for collocating verbs for these 150 nouns. This was in order to use a corpus large enough to allow strong collocational patterns to emerge. To stay within the journalistic genre, we used the database Les Voisins de Le Monde (http://redac.univ-tlse2.fr/voisinsdelemonde/). This database contains 200,000,000 words and is based on every edition of the newspaper Le Monde from 1991–2000. The database provides both frequencies and a mutual information (MI) score, which is a statistical measure used in corpus linguistics to calculate a collocation’s degree of cohesiveness.

3. Several selection criteria were applied in the search for relevant test items: the MI-score threshold was set at > 3, in accordance with many other studies in the field (Ellis, Simpson-Vlach & Maynard 2008; Siyanova & Schmitt 2008). The frequency threshold was set at 200. Given the size of the database, this yields a frequency of 1/1,000,000 words. That may seem like a small number, but Ellis et al.’s study also served as a reference here, in which the frequency threshold was set at 10/1,000,000 words. Their study was concerned with lexical bundles in academic English, such as at the end of the, which are more frequent in the input than the content-word-based collocations in our study (see Forsberg 2010, for differences in frequency between different types of formulaic sequences). After applying these two frequency thresholds, another selection criterion was employed. As this test was supposed to target the most advanced levels, it was deemed important to eliminate overly easy items. Recent L2 collocation research has shown that collocations that are equivalent in the L1 and the L2 have, not surprisingly, been shown to convey acquisition advantages (Wolter & Gyllstad 2011). It was therefore decided to eliminate all translation equivalents, or as some would call them, cognates. Examples of such collocations are commettre un crime ‘begå ett brott’ (Eng. “commit a crime”) and signer un accord ‘skriva under ett avtal’ (Eng. “to sign an agreement”). Having applied all the above criteria, 70 verb-noun combinations remained.
4. Example sentences for these 70 combinations were searched for on www.google.fr.

60 items provided good sentences, which were piloted with 10 NSs of French and nine NNSs in Stockholm. The results of this pilot were used to determine the final selection of items by establishing a minimum accuracy threshold. Only if 50% or more of the NSs provided a correct answer to an item, was the item included in the final test. This procedure yielded a total of 46 items in the final test.

As stated earlier, the test format was inspired by that of Mizrahi and Laufer (2010) and looks like this:

1. L'ONU est fermement résolue à r _____ justice aux victimes du génocide.
2. Si les petites entreprises é _____ des difficultés à embaucher, c'est aussi parce qu'elles disposent de moyens qui sont moindres.
3. L'industrie porno entend m _________ un terme au piratage d'ici 2012.

In contrast to Mizrahi and Laufer (2010), only one letter was given in the gap to avoid ceiling effects. Even though this meant that a possible alternative collocating verb was occasionally provided, it was decided, for practical reasons, only to judge as correct the answers with the most frequently collocating verbs, i.e. only one correct answer for each sentence.

6.5 The Written Grammatical Judgment test (GJT)

As already suggested, the GJT was included as a point of comparison for the collocation test and was therefore accorded a minor role in the present study. The written GJT was developed in collaboration with Inge Bartning (see Bartning 2012). The test was originally based on Bylund's (2008) GJT for Spanish. Grammatical features tested included subject-verb agreement in complex syntax, noun-phrase (NP) agreement, anaphors, verb constructions and time, mode, aspect (TMA). The test included a total of 46 items. It should be noted that both verb-phrase (VP) and NP agreement sentences include ungrammaticalities in silent morphology, which would not be audible in an auditory GJT; these probably tap into metalinguistic awareness more than any mental grammar. Here are two example sentences:

1. Les élèves croyaient que le sapin *avaient/avait comme origine la mythologie allemande (The students thought that the Christmas tree *were/ was of German origin).
2. L'inspecteur pensait que l'avion n'était plus suffisamment performant (The inspector thought that the airplane was not able to perform sufficiently well anymore).
7. Results

The first research question involved exploration of relationships among linguistic measures, collocations, grammatical judgment and aptitude tests in highly proficient late L2 French learners. Means and standard deviations for scores on the linguistic tests are presented in Table 1.

Table 1. Means, standard deviations of linguistic tests among highly proficient adult French L2 learners (n = 13)

<table>
<thead>
<tr>
<th>Test</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collocations</td>
<td>29.70</td>
<td>8.13</td>
</tr>
<tr>
<td>Grammatical judgement</td>
<td>38.31</td>
<td>3.99</td>
</tr>
<tr>
<td>LLAMA B</td>
<td>54.23</td>
<td>24.65</td>
</tr>
<tr>
<td>LLAMA D</td>
<td>37.70</td>
<td>15.22</td>
</tr>
<tr>
<td>LLAMA E</td>
<td>78.46</td>
<td>33.63</td>
</tr>
<tr>
<td>LLAMA F</td>
<td>64.62</td>
<td>24.70</td>
</tr>
</tbody>
</table>

The present sample of French L2 learners can be compared on LLAMA test scores to another group of late L2 learners (Granena & Long 2013). Our sample has an age of onset of 13–17 years, and age of arrival in France of 17–21. In the Granena and Long sample, age of onset was 16–29 years, which coincided with age of arrival 16–29 years. In our sample, the participants’ mean age of testing was 38 years, whereas mean age of testing was 31.8 in the Granena and Long sample. Our sample had almost the same means as the late learners in Granena and Long (2013) on LLAMA B (54.23 and 50.00, respectively). Differences between samples were greater for the other LLAMA tests. Our sample had descriptively higher means on LLAMA D (37.70 vs. 24.17), LLAMA E (78.46 vs. 62.22), and LLAMA F (64.62 vs. 45.56).

Results from correlations of linguistic test scores indicated one significant positive correlation ($r = .58$, $p < .05$) between collocations and LLAMA D scores (Table 2). Other scales were also positively correlated, but not statistically significantly so. Trends in the direction of significance were found for collocations and grammatical judgment ($r = .50$, $p = .08$), for LLAMA F and GJT scores ($r = .47$, $p = .11$), and between LLAMA B and LLAMA F ($r = .47$, $p = .10$).

The second research question concerned relationships between personality dimensions, as measured by MPQ, and linguistic measures. Means, standard deviations and Cronbach alphas for scales in the MPQ are presented in Table 3. Analyses of Cronbach alpha for the subscales were satisfactory (all above .70) and
Table 2. Pearson's correlations between collocations, grammatical judgment and aptitude tests among highly proficient adult French L2 learners (n = 13)

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Collocations</td>
<td>.50</td>
<td>.32</td>
<td>.58*</td>
<td>.13</td>
</tr>
<tr>
<td>2</td>
<td>Grammaticality judgment</td>
<td>-</td>
<td>.35</td>
<td>.05</td>
<td>.10</td>
</tr>
<tr>
<td>3</td>
<td>LLAMA B</td>
<td>-</td>
<td>-</td>
<td>.35</td>
<td>.24</td>
</tr>
<tr>
<td>4</td>
<td>LLAMA D</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.10</td>
</tr>
<tr>
<td>5</td>
<td>LLAMA E</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>LLAMA F</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: *p < .05, two-tailed.

Table 3. Means, standard deviations and Cronbach alphas for the MPQ among highly proficient adult French L2 learners (n = 13)

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>Cronbach's α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cultural empathy</td>
<td>4.10</td>
<td>.31</td>
<td>.72</td>
</tr>
<tr>
<td>Open-mindedness</td>
<td>3.79</td>
<td>.53</td>
<td>.87</td>
</tr>
<tr>
<td>Social initiative</td>
<td>3.89</td>
<td>.56</td>
<td>.84</td>
</tr>
<tr>
<td>Emotional stability</td>
<td>3.15</td>
<td>.51</td>
<td>.87</td>
</tr>
<tr>
<td>Flexibility</td>
<td>3.05</td>
<td>.37</td>
<td>.72</td>
</tr>
</tbody>
</table>

comparable to those in previous studies (van Oudenhoven & van der Zee 2002; Kozziilius, van Hooft, Planken & Hendrix 2011). As most individuals in the present study were professionally and socially integrated female expatriates, it was possible that their personality profiles according to the MPQ would differ from those in other samples. A comparison of MPQ subscale means indicated that this group of expatriates exhibited higher means for cultural empathy and social initiative (M = 4.10 and 3.79, respectively) compared to samples in the literature, e.g. exchange students (M = 3.76 and 3.17, respectively) (Leong 2007) and business professionals with international experience (M = 3.78 and 3.58, respectively) (Korzziilius et al. 2011). The open-mindedness mean, as well, was higher for the expatriates (M = 3.79) than the means of exchange students (M = 3.65) and business professionals (M = 3.58). Means for emotional stability and flexibility were fairly similar.

Results showed significant positive correlations between two scales of the MPQ and collocation scores (Table 4). Both cultural empathy ($r = .62, p < .05$) and open-mindedness ($r = .57, p < .05$) correlated with collocation scores. Among
the aptitude tests, the same personality dimensions of cultural empathy ($r = .67$, $p < .05$) and open-mindedness ($r = .78$, $p < .05$) correlated positively with one of the aptitude tests, LLAMA D. The total MPQ score did not correlate with collocations, grammatical judgment or aptitude tests.

### 8. Discussion

The first research question concerned how indicators of high level proficiency (collocations and grammaticality judgments) relate to aptitude in L2 French among late learners. It was shown that collocations correlated positively and strongly with LLAMA D, but not with the other LLAMA sub-tests. It is interesting that collocations correlated only with the LLAMA D, since Granena and Long (2013) found a similar correlation. Some theoretical support can be found for this, as there is evidence of an association between phonological memory and the learning of vocabulary, or nativelike selection (Speciale et al. 2004; Bolibaugh & Foster this volume). The LLAMA D would also tap into such learning, since it measures phonetic memory. A more general theoretical implication of this finding would be that phonetic memory, which, according to Granena (this volume), taps into the implicit learning capacity, would be more important for high-level proficiency than grammatical inference. The latter assumption is in line with the results in the present study and shows that grammaticality judgment scores were not correlated with collocations or with the LLAMA tests.

The second research question investigated whether personality dimensions relate to indicators of high-level proficiency (collocations and grammaticality judgments) in L2 French among late learners. Findings showed that the personality dimensions of cultural empathy and open-mindedness were significantly
and positively associated with both collocations and LLAMA D scores, but not with grammaticality judgment test scores. Both personality dimensions appear to involve a cognitive capacity to overcome in-group bias, usually leading to devaluation of the outer group, in this case the L2 community, as high scorers on cultural empathy and open-mindedness are disposed to adopt other people's perspective easily, to be tolerant and not prejudiced. The present sample may consist of individuals motivated to learn the L2, who had an open mind and were able to empathize with others that led to their engagement with successful learning strategies (including exposing themselves to the L2 community), in order to integrate well professionally and socially. Learning collocations may be one means of achieving these ends. All in all, these individuals can be characterized as having achieved multicultural effectiveness. From another point of view, taking the findings on associations between cultural empathy, open-mindedness and verbal abilities (Van der Zee et al. 2003) into consideration, it may be that these individuals had originally rather advanced verbal abilities and, therefore, learned collocations more effortlessly than individuals with weaker verbal abilities.

It was intriguing to find correlations between personality dimensions and LLAMA D. These correlations were not foreseen by the design and are not easily explained by the literature. We can only speculate about why the LLAMA D test, which assesses phonetic memory, was the only LLAMA sub-test that related to personality. Rota and Reiterer (2009) found an association between empathic skills (as measured by the E-scale, Leibetseder et al. 2001) and different pronunciation-related measures, including pronunciation performance and phonetic coding ability as measured by the MLAT aptitude test, which is likely similar to the LLAMA D. It is possible that the E-scale shares variance with the personality dimensions of cultural empathy and open-mindedness with respect to sensitivity and concern with emotions and cognitions. As this individual, capable of perspective-taking (cultural empathy) and showing tolerance (open-mindedness), would also understand social situations well and show empathic emotions. He/she would also be able to put him- or herself into the emotional situations of another person, as measured by the E-scale. Empathy involves investigating reactions to real-life situations and showing concern that requires cognitive analysis. The three measures appear to share a perspective-taking and empathic approach. This line of argument is speculative, and for future research it would be interesting to investigate associations between empathy and phonetic memory/pronunciation, and how those dimensions relate to levels of mastery of collocations.

Research on cognitive styles and SLA offers arguments for the importance of perspective-taking ability and tolerance for multicultural effectiveness shown in both tolerance of ambiguity (interpersonal aspects, coping with change) and
field-dependence-independence (functioning rather autonomously from external sources of information, i.e. demonstrating independence).

It should be noted that it is indeed an open question if the elevated scores for cultural empathy and open-mindedness already characterized the sample at the beginning of their stay in France or if the scores were to some degree the result of successful intercultural adaptation (see Dewaele & Van Oudenhoven 2009).

Grammaticality judgment was included in the design for comparative reasons because high grammaticality judgment scores can indicate that an individual has high metalinguistic awareness. This kind of proficiency does not appear to be of importance for collocations, as there was no correlation between them in the present study, and did not relate to aptitude as measured by the LLAMA sub-tests. This confirms our belief that collocations are a particularly fruitful linguistic category to include in studies of high L2 proficiency.

We would also like to draw attention to relationships with trends towards statistical significance between MPQ scores and linguistics measures. One of the LLAMA sub-tests, namely, LLAMA E, had only negative correlations; the weakest correlation was -.13 for cultural empathy and the strongest for emotional stability ($r = -.43$). This means that an individual with low scores on emotional stability has high scores on LLAMA E, tapping the aptitude to detect sound-symbol correspondences. Future studies are warranted to examine this link more closely; whether these rather curious findings are only relevant for LLAMA E and certain personality dimensions, or only related in samples with highly proficient adult L2 learners with similar personality profiles. Moreover, the present sample had elevated scores on social initiative. High scorers on social initiative are less likely to feel discouraged after facing obstacles and failures, whereas low scorers tend to have difficulty finding alternative solutions to problems and maintaining commitment to goals. Research shows that social initiative in the MPQ and extraversion in the FFM are strongly correlated (see van der Zee & van Oudenhoven 2000). It is interesting to note that the proactive approach of social initiative did not relate to L2 proficiency, but may have been an advantage for the present sample’s successful adaption to their new country. The present result with high scores on social initiative is in line with the literature on its role for work adjustment (Pelto, 2008; Pelto & Froese 2011) and on the extroverts’ disposition to seek social contact, status, and power (Wilt & Revelle 2009).

The limitations of our exploratory study need to be noted. Firstly, the sample was small ($n = 13$). Secondly, it was a convenience sample, and included socially and professionally integrated individuals. The sample was also limited in that the majority of participants were female and well-educated. Additional research with larger, more diverse samples is warranted. Despite the evident limitations, we believe that the present results add to the body of research on highly proficient late
L2 learners, due to the strong positive correlations observed between personality measures intended to assess multicultural effectiveness and measures of linguistic proficiency and aptitude.

9. Concluding remarks

The present exploratory study provides some support for the use of personality measures in L2 acquisition, as two personality dimensions in the MPQ correlated strongly with a specific linguistic measure of high-level L2 proficiency, i.e. collocations, and also with one component of language aptitude, i.e. phonetic memory/sound recognition. In particular, personality dimensions encompassing perspective-taking and an open-minded attitude appear to be more important for L2 attainment than personality dimensions emphasizing a more social and outgoing individual, when high-level L2 proficiency is concerned. Moreover, the use of measures of colloquional knowledge to tap high L2 proficiency is a promising area for further investigations. Future studies are warranted to replicate the results in larger and more diverse samples, given the exploratory nature of the study and the small sample used, which could have influenced the results of the correlations. With respect to the multifaceted learning process for L2, it is suggested that both personality and social psychological factors should be taken into consideration when predicting successful L2 learning.

References


Kormos, J. (this volume). The role of working memory, phonological short-term memory and language aptitude in second language attainment.


