CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Bilingual language experience as a multidimensional spectrum: Associations with objective and subjective language proficiency

Jason W. Gullifer[1] , Shanna Kousaie[2], Annie C. Gilbert[3], Angela Grant[4], Nathalie Giroud[5], Kristina Coulter[6], Denise Klein[7], Shari Baum[7], Natalie Phillips[8] and Debra Titone[9,*]

[1]McGill University and Centre for Research on Brain, Language and Music, [2]University of Ottawa and Centre for Research on Brain, Language and Music, [3]McGill University and Centre for Research on Brain, Language and Music, [4]Missouri Western State University and Centre for Research on Brain, Language and Music, [5]University of Zurich and Centre for Research on Brain, Language and Music, [6]Concordia University and Centre for Research on Brain, Language and Music, [7]McGill University and Centre for Research on Brain, Language and Music, [8]Concordia University and Centre for Research on Brain, Language and Music and [9]McGill University and Centre for Research on Brain, Language and Music
*Corresponding author. E-mail: debra.titone@mcgill.ca.

Abstract

Despite the multifactorial space of language experience in which people continuously vary, bilinguals are often dichotomized into ostensibly homogeneous groups. The timing of language exposure (age of acquisition) to a second language (L2) is one well-studied construct that is known to impact language processing, cognitive processing, and brain organization, but recent work shows that current language exposure is also a crucial determinant in these domains. Critically, many indices of bilingual experience are inherently subjective and based on self-report questionnaires. Such measures have been criticized in favor of objective measures of language ability (e.g., naming ability or verbal fluency). Here, we estimate the bilingual experience jointly as a function of multiple continuous aspects of experience, including the timing of language exposure, the amount of L2 exposure across communicative contexts, and language entropy (a flexible measure of language balance) across communicative contexts. The results suggest that current language exposure exhibits distinct but interrelated patterns depending on the socio-experiential context of language usage. They also suggest that, counterintuitively, our sample more accurately self-assesses L2 proficiency than native language proficiency. A precise quantification of the multidimensional nature of bilingualism will enhance the ability of future research to assess language processing, acquisition, and control.

Keywords: bilingualism; language entropy; language proficiency

Bilingualism is a complex and multifaceted life experience. The onset of bilingualism can occur at any stage in life, and it can extend into old age. Some bilinguals learn both languages at a young age in the home, and others learn their languages in

CrossMark

adulthood when they immigrate to a new country. Language and neurocognitive processing are intimately related to the timing of bilingual language experience, such as age of second language (L2) acquisition (Berken, Chai, Chen, Gracco, & Klein, 2016; Berken, Gracco, Chen, & Klein, 2015; Flege, Munro, & Mackay, 1995; Gullifer et al., 2018; Johnson & Newport, 1989; Klein, Mok, Chen, & Watkins, 2014; Kousaie, Chai, Sander, & Klein, 2017; Kousaie & Phillips, 2011). However, bilinguals also use their languages to various degrees and for a multitude of reasons, independent of the age at which they were acquired.

Sometimes, the communicative context or geographic location may dictate the use of one language or the other. For example, bilinguals living in English monolingual areas of the United States may use English primarily in the workplace. Bilinguals living in the French monolingual province of Quebec (Canada) are regulated to use French, at least initially, with customers in the workplace. At the same time, bilinguals in each of these areas may use both of their languages jointly in other settings, such as when speaking with bilingual friends or family. Theories, past and present, together with recent psycholinguistic research suggest that it is crucial to quantify these types of language experience and usage patterns in order to provide an adequate description of the core phenomena of interest, which include language processing, language learning, and language control (Abutalebi & Green, 2016; Anderson, Hawrylewicz, & Bialystok, 2018; Anderson, Mak, Keyvani Chahi, & Bialystok, 2018; Beatty-Martinez & Dussias, 2017; Beatty-Martinez et al., 2020; Bice & Kroll, 2019; DeLuca, Rothman, Bialystok, & Pliatsikas, 2019; Green & Abutalebi, 2013; Green & Wei, 2014; Grosjean, 1985, 2016; Gullifer et al., 2018; Gullifer & Titone, 2020a, 2020b; Sulpizio, Del Maschio, Del Mauro, Fedeli, & Abutalebi, 2019).

It is now becoming clear that language experience is not a categorical phenomenon. Instead, it exists on a multidimensional spectrum, and the social factors surrounding language experience have consequences for language processing, cognitive processing, and brain organization (e.g., Anderson, Hawrylewicz, et al., 2018; Baum & Titone, 2014; Dash, Berroir, Joanette, & Ansaldo, 2019; DeLuca, Rothman, et al., 2019; DeLuca, Rothman, Bialystok, & Pliatsikas, 2020; Gullifer et al., 2018; Gullifer & Titone, 2020a; Li, Legault, & Litcofsky, 2014; Luk & Bialystok, 2013; Sulpizio et al., 2019; Tiv, Gullifer, Feng, & Titone, 2020). Given that these phenomena are studied in different locations around the world, a common quantification of these experiences is necessary to ensure comparability between studies.

Of relevance here, recent theoretical efforts to describe neurocognitive aspects of bi- and multilingualism invoke the idea of relative balance of use for each language (i.e., the distribution of first language [L1] vs. L2+ language experience), which may be assessed in addition to sheer language exposure. Thus, a key factor in assessing language experience would involve understanding whether people use some or all of the languages they know jointly in communicative contexts or separately. In line with this idea, various theoretical perspectives have converged upon similar ideas, but using different terminology. For example, the adaptive control hypothesis (Abutalebi & Green, 2016; Green & Abutalebi, 2013; Green & Wei, 2014) makes the distinction between *single language contexts* (where only one language is used), *dual language contexts* (where two languages are used, typically with different speakers), and *dense code-switching contexts* (where two languages are mixed within a single utterance). Dual language contexts are thought to involve greater competition

between the two languages, requiring greater need for the application of control mechanisms. A related theory, Grosjean's "language mode hypothesis" (Grosjean, 1997, 2001), conceptualized the idea of a *bilingual mode* and a *monolingual mode*, which, when entered, can up- or downregulate the activation of the unintended language. Grosjean states that these modes may not be completely independent and instead exist along a continuum.

Crucially, bilinguals differentially distribute their use of particular languages depending on the communicative context in question and even the topics of conversation that occur in that context (e.g., Grosjean, 1998, 2016; Schrauf, 2002, 2009; Tiv, Gullifer, et al., 2020; Vaid & Menon, 2000). Grosjean refers to this observation as the complementarity principle, and he notes that each individual may exhibit an entirely unique pattern of language distribution. For example, in Montréal, Quebec, a participant may report using only French in the home, and report using both French and English in social situations. In contrast, another participant may report using both languages throughout all of their communicative contexts.

In some cases, the communicative context may place specific demands on the speaker. Provincial regulations in Quebec stipulate that French is the prominent language of business and commerce, and thus most work environments in the bilingual city of Montréal require knowledge and use of French. In the Montréal context, we consistently find distinct patterns of language distribution between home and work contexts (Gullifer & Titone, 2020a; Tiv, Gullifer, et al., 2020), though we note that sometimes work contexts are reported to be quite bilingual, counter to official regulations. In other cases, language choice may be under complete control of the speaker, such as the use of language for purposes of inner thought and mental computations (e.g., dreaming, thinking, counting, and arithmetic). Although bilinguals differ in the extent to which they report using each of their languages for various purposes, individual differences in language usage in various communicative contexts are differentially associated with several aspects of bilingual language experience, including length of residence in a country, L2 age of acquisition (AoA), self-reported language proficiency, objective language proficiency, and language dominance (e.g., Gullifer & Titone, 2020a; Schrauf, 2009; Tiv, Gullifer, et al., 2020; Vaid & Menon, 2000). Of importance, these observations have profound implications for assessing language proficiency.

Minimally, language proficiency reflects one's ability to apply language knowledge for the purposes of comprehension and production. Although many measures of language proficiency probe various linguistic domains (e.g., lexical and syntactic domains), within-domain proficiency is often treated in a global sense, without regard for the communicative context in which the language is used. However, because bilinguals distribute their use of particular languages differentially throughout various social and communicative contexts, proficiency for a given language may develop uniquely in some contexts. As a simple example, consider a French–English bilingual who uses French but not English in the home. This individual may have greater difficulty accessing English vocabulary knowledge for concepts that typically occur in the home (e.g., terms for cooking appliances), which could manifest as slower reaction times on a picture naming test for those items or fewer exemplars produced in a verbal fluency task. Similar difficulties could extend to French vocabulary knowledge for concepts that occur outside the home. In other words, "people become bilingual by domain and not globally" (Schrauf, 2009).

It thus follows from these theories that fluent language production, comprehension, and associated control processes may become adapted to the social context of language usage. In order to properly assess how bilingual (or multilingual) experiences contribute to these adaptive processes, it is necessary to assess multiple aspects of language experience, including age of exposure, amount of sheer language usage, and language balance, across several social and communicative contexts. Despite the existence of highly sophisticated methods for quantifying behavior and brain function, and detailed questionnaires that probe various experiences with language, there is a lack of validated analysis routines that allow researchers to precisely quantify the rich and diverse nature of people's experiences using language.

To address these issues, which have led to persistent gaps in the literature, we proposed a method for assessing language balance in various communicative contexts among multilingual participants called *language entropy* (Gullifer et al., 2018; Gullifer & Titone, 2018, 2020a, 2020b).[1] The general concept of entropy comes from physics and information theory, and it provides a measure of diversity and uncertainty when the relative proportion of occurrences for a set of "states" is known (e.g., Shannon, 1948). Here, we take "states" to be the usage of a particular language (e.g., English or French) within a communicative context (e.g., at home or at work). This information is frequently elicited directly by language background questionnaires, or it can be otherwise computed from questions probing language usage, and it can be used to compute an entropy score directly for each probed context. To reduce the number of variables and identify the core latent constructs that potentially drive language entropy variables, we use latent variable analysis.

We argue that language entropy can continuously estimate the extent to which individuals immerse themselves in dual versus single language contexts (or language mode; Gullifer et al., 2018; Gullifer & Titone, 2018, 2020a, 2020b). To illustrate language entropy, imagine two hypothetical people, William and Zoe. William reports hearing only English in the home and never French (100% English, 0% French; a single language context that may induce a monolingual mode). William would receive a low language entropy score in the home (i.e., language entropy of 0), reflecting low language-related diversity and low language-related uncertainty in the home. Thus, at any given moment in the home, William is quite certain that he would hear English (and not other languages). In contrast, Zoe reports equal exposure to English and French in the home (50% French, 50% English; a dual language context that may induce a bilingual mode). Zoe would receive a higher score (i.e., language entropy of 1), reflecting higher language-related diversity and uncertainty in the home. In other words, at any given moment in the home, Zoe may be highly uncertain about whether French or English will be used next (in the absence of other information that could provide a cue about language, such as speaker identity) due to higher diversity in language exposure compared to William.

Our initial work shows that for a sample of over 400 Montréalers, language entropy is composed of two core latent constructs (Gullifer & Titone, 2020a, 2020b): general entropy (composed of variables related to reading, speaking, entropy at home, and entropy for social purposes) and work-related entropy (composed of variables related to entropy at work with some cross-loading from social entropy). We then show that individual scores on these constructs are associated with self-reported language proficiency (such as L2 abilities and L2 foreign accentedness; Gullifer & Titone, 2020a)

and with proactive executive control abilities and associated brain networks (Gullifer et al., 2018; Gullifer & Titone, 2020b), in line with predictions of theoretical models outlined above. Work by others shows that language entropy is related to patterns of dual language use such as engagement in language mixing (Kałamała, Szewczyk, Chuderski, Senderecka, & Wodniecka, 2020), the ability to mentalize (or engage in social–cognitive processing) in the first and second languages (Tiv, O'Regan, & Titone, 2020), and the organization of brain networks responsible for language and executive control (Sulpizio et al., 2019).

Importantly, as raised earlier, bilingualism is a multifaceted language experience, and thus several other constructs must be assessed besides language entropy, such as historical language exposure (e.g., L2 AoA) and sheer exposure to the second language. The relative import of each of these variables may depend on the domain of inquiry (e.g., language proficiency vs. language/executive control), the bilingual or multilingual population in the sample, and the communicative context of language usage. Thus, in our view, these constructs should be investigated with continuous measures that are assessed jointly. This allows for the estimation of the unique explanatory power of each construct that indexes bilingual language experience. An emerging picture from recent work on joint contributions suggests that several factors (i.e., L2 AoA, sheer L2 exposure, and language entropy) predict self-reported language proficiency (Gullifer & Titone, 2020a), but language entropy best predicts proactive executive control (Gullifer & Titone, 2020b).

Critically, our prior work has limitations that we identify here, some of which we are able to directly address. First, Gullifer and Titone (2020a) relied primarily upon self-reported measures of language abilities as a proficiency estimate. Although there are some demonstrations that self-reported measures correlate well with objective measures of proficiency (e.g., Blanche & Merino, 1989; Jia, Aaronson, & Wu, 2002; Schrauf, 2009), there are concerns that self-ratings may be poor indicators of proficiency. Participants may not be able to accurately assess their own language abilities, or they may up- or down-weight their abilities to adhere to cultural norms and expectations (Tomoschuk, Ferreira, & Gollan, 2019; Zell & Krizan, 2014). Thus, there may be inherent issues using self-ratings as indicators of language proficiency when comparing across populations. However, as described earlier, objective proficiency measures may also be fundamental indicators of social and cultural patterns of language use. Similar critiques about validity can be made for other aspects of language experience that rely on self-report, including L2 AoA, sheer L2 exposure, and language entropy. In our view, in the absence of explicit validation of these constructs, the best approach is to continue assessing the interrelationships between these constructs, including objective measures.

Second, in our previous work, the assessment of the communicative contexts in which language is used was highly coarse. For instance, we assessed language entropy in a small number of communicative contexts (e.g., at home, at work, in social settings, for reading, and for speaking). Yet, individuals engage in many more contexts of language use throughout their day. This includes the use of language for mental processes and computations such as thinking, dreaming, arithmetic, and counting (e.g., Schrauf, 2009; Vaid & Menon, 2000); language use in other communicative contexts, such as in the university environment; and language use for consumption of media (e.g., television, radio, and internet). The importance of assessing

language usage across several communicative contexts, and in particular the ways bilinguals and multilinguals use their languages flexibly and integratively within these contexts (see e.g., translanguaging; García & Wei, 2012), has long been known from work within and at the intersection of sociolinguistics, applied linguistics, and bilingualism (e.g., Commins, 1989; García & Wei, 2012; Schrauf, 2009; Vaid & Menon, 2000). These approaches are gaining recognition in the cognitive sciences more generally (e.g., Anderson, Hawrylewicz, et al., 2018; Anderson, Mak, et al., 2018; DeLuca et al., 2019; Grosjean, 2016).

Thus, our purpose here is to further assess the structural properties of and inter-relationships between several constructs related to bilingual language experience, including the timing of language exposure, language entropy, sheer amount of exposure to the L2, and language proficiency. We extracted this information from several measures on a health and language background questionnaire. To assess the timing of L2 exposure, we extracted information from an item that probed L2 AoA. To assess language entropy and sheer L2 exposure, we extracted information about language usage from multiple items that tap into different communicative contexts of language use. To assess language proficiency, we extracted verbal fluency production data (objective proficiency) and self-rated language abilities (subjective proficiency).

We then computed factor analyses and resultant factor scores for each of the constructs with multiple items: language entropy, sheer L2 exposure, and language proficiency. The factor analyses provide information about the latent structure of each facet of bilingual language experience. Next, we extract individual scores for each latent factor and model the extent to which the scores on these factors can jointly classify the timing of language exposure in a grouped manner (simultaneous acquisition, sequential L1 French, or sequential L1 English). Finally, we model whether combinations of factor scores for language entropy, sheer L2 exposure, and L2 AoA (as a continuous measure) relate to factor scores for language proficiency. This approach addresses some of the limitations of previous work; namely, we included objective measures of language proficiency together with subjective measures. Moreover, our questionnaire probed estimated language usage in 16 communicative contexts, including for internal purposes (for things like thinking and dreaming), use in various social settings, and use for media consumption.

To preview the findings reported below, we show that language entropy and sheer L2 exposure comprise similar latent structures, involving two core latent variables: (a) a factor that includes measures of internalized language use (such as thinking, dreaming, and counting) and social language use; and (b) a factor that includes measures related to externally directed or professional use. Language entropy (but not sheer L2 exposure) also includes a third factor that reflects measures of media usage. Crucially, we then showed that factor scores for sheer L2 exposure (but not language entropy) discriminate whether individuals acquired their languages simultaneously or sequentially.

Next, we assessed relationships with language proficiency. We show that language proficiency comprises three latent variables: (a) a factor that includes measures of objective *and* subjective L2 proficiency, (b) a factor that includes measures of subjective L1 proficiency, and (c) a factor that includes measures of objective L1 proficiency. Factor scores for both sheer L2 exposure and language entropy were

important in predicting factor scores for L2 proficiency. In contrast, only factor scores for sheer L2 exposure predicted factor scores for subjective L1 proficiency. There were no significant predictors of factor scores for objective L1 proficiency. Together, these results demonstrate how bilingual language experience can be quantified jointly and continuously to predict other constructs of interest. Moreover, they also show the differential contributions of each construct related to bilingual language experience, depending on the outcome measure and also the communicative context in which language is used.

Method

Participants

We tested 87 bilingual speakers of French and English who completed a health and language background questionnaire as part of their participation in a larger project. Participants were young adults ($M_{age} = 23.8$ years, $SD_{age} = 4.1$ years) from the Montréal area, and many were university students. The majority of participants were born in Canada ($n_{Canada} = 75$). Some of the participants were born outside of Canada and subsequently moved to Canada at an early age ($n_{France} = 3$; $n_{USA} = 2$; $n_{China} = 1$; $n_{Zimbabwe} = 1$; $n_{Australia} = 1$; $n_{unknown} = 4$). The speakers differed with respect to the languages that they first acquired: 27 acquired French and English simultaneously, 34 English followed by French sequentially, and 26 acquired French followed by English sequentially. On average, participants acquired their L2 early in life ($M_{L2AoA} = 3.8$ years $SD_{L2AoA} = 3.2$ years). Participants' first language was determined through inspection of language history questions that probed age of language onset. Some participants ($n = 15$) reported current exposure to a third language (L3) besides English or French. On average, for conversations, participants reported using the L1 64% of the time and the L2 35% of the time. Among participants with and L3, they reported using it 5% of the time. Of note, average proportions do not sum to 100% because not all participants spoke an L3. Participants came from generally high socioeconomic backgrounds as measured by highest parental education. Among the available data on parental education, 73 participants had parents with at least some college education. See Table 1 for a summary of demographic and language history data. All participants gave informed consent, and the McGill Research Ethics Board approved data collection for any linked projects.

Materials

Assessing self-reported language exposure

All participants included in this sample completed a health and language background questionnaire. This questionnaire has sections that probe basic demographic information (e.g., age, gender, and education), language background and experience (including language use and exposure for various purposes and in various communicative contexts, self-reported language abilities, etc.), and basic health information (e.g., data collected to assess eligibility for magnetic resonance imaging studies). This questionnaire was adapted from various questionnaires used in the field, including the Language Experience and Proficiency Questionnaire (Marian,

Table 1. Summary of participant characteristics, including demographic, language history, and verbal fluency data

Measure	N	M	SD	Min	Max
Age (years)	87	23.79	4.06	18	36
L2 AoA (years)	87	3.76	3.24	0	15
L1 conversations (%)	69	0.64	0.24	0.1	1
L2 conversations (%)	69	0.35	0.24	0	0.9
L3 conversations (%)	15	0.05	0.13	0	0.5
Mixing with family (1[*never*]–7[*always*])	66	3.77	2.01	1	7
Mixing with friends (1[*never*] –7[*always*])	74	4.61	1.55	1	7
Mixing with classmates (1[*never*]–7[*always*])	54	3.00	1.44	1	6
Mixing with coworkers (1[*never*]–7[*always*])	56	3.77	1.72	1	7
L1 accent (1[*none*]–7[*strong*])	73	1.40	0.88	1	5
L2 accent (1[*none*]–7[*strong*])	86	2.79	1.24	1	6
L1 reading ability (1[*very poor*] –7[*nativelike*])	86	6.83	0.49	5	7
L1 writing ability (1[*very poor*]–7[*nativelike*])	86	6.54	0.82	4	7
L1 speaking ability (1[*very poor*] –7[*nativelike*])	86	6.85	0.45	5	7
L1 listening ability (1[*very poor*]–7[*nativelike*])	85	6.92	0.35	5	7
L2 reading ability (1[*very poor*]–7[*nativelike*])	87	5.95	1.00	4	7
L2 writing ability (1[*very poor*]–7[*nativelike*])	87	5.29	1.25	2	7
L2 speaking ability (1[*very poor*] 7[*nativelike*])	87	5.64	1.04	3	7
L2 listening ability (1[*very poor*]–7[*nativelike*])	86	6.12	1.01	3	7
L1 category fluency (mean exemplars produced)	86	20.56	6.36	9	38
L1 letter fluency (mean exemplars produced)	87	12.96	3.47	5.3	24
L2 category fluency (mean exemplars produced)	87	15.82	5.36	6	30
L2 letter fluency (mean exemplars produced)	87	9.72	2.86	4	16

Blumenfeld, & Kaushanskaya, 2007) and the Language History Questionnaire 2.0 (Li, Zhang, Tsai, & Puls, 2014), and questions were adapted to apply to the Montréal and Canadian contexts. We detail the extraction of these measures below.

Assessing self-reported age of language acquisition
Participants answered several questions about the AoA for all of the languages that they knew. Participants were asked: "At what age did you first start to learn each language in terms of speaking (at what age did you speak your first words?), reading, and writing, and the number of years you have spent learning each language." They were also asked: "Please indicate the age at which you started to learn each language in the following situations—indicate the age in the boxes only for situations that are relevant: at home,

at school, after immigrating to the country where spoken, informal settings (e.g., nannies or friends), software (e.g., Rosetta Stone), other." Through visual inspection of these data, we determined AoA for each language based on the earliest age that the participant entered. If the AoA was the same for multiple languages, we determined that they acquired the languages simultaneously. In this case, L1/L2/L3 were determined on the basis of self-reported proficiency. If the AoA was different for multiple languages, then the L1/L2/L3 were determined on the basis of AoA.

Assessing self-reported language usage among communicative contexts
Participants reported the extent to which they used their languages in 16 different communicative contexts with one of two measurement types. The communicative contexts together with their measurement types are listed in Table 2. Measurement types included either a percentage of language use (0%–100%) or 7-point Likert scale where 1 indicated *no use* and 7 indicated use *always*. Language use related to inner thoughts and for the purposes of emotion or anger was elicited with the following prompt: "How often do you use your languages for the following activities? Use the following scale and fill in the number in the table. [1 (*never*)–7(*always*)]." Language use for a set of questions that generally probed passive language use (e.g., listening, reading, and writing) was elicited via the following prompt: "Please estimate the total number of hours each day that you spend engaged in the following activities, and indicate what percentage of that time you spend engaged in that activity in each of the languages that you know (please write down the languages). If you are not currently engaged in an activity using that language write '0'; the total percentage for each activity should equal 100%." Language use related to conversations in different social contexts was elicited via the following prompt: "Please estimate the percent of conversations that take place in each of your languages, and what percentage of that is with the following people. The total across languages should equal 100% and the total within each language should equal 100%."

The computation of language entropy requires the proportion of language usage within each communicative context to be known. Percentages are easily converted to proportions by dividing by 100. Likert scales were converted to proportions with the following procedure. We first subtracted 1 from each elicited score, so that the value for "never" would be adjusted to 0. We then divided a language's score within a context by the sum total of the scores across languages within that communicative context. For example, a participant who reported (after adjustment) the following data for dreaming "L1: 6" and "L2: 5" would receive the following proportions for dreaming L1: "6/11" and "L2: 5/11."

A careful reader will notice that the wording of questions about passive language use differed from the wording of questions about conversations in social contexts. Specifically, for passive language use, participants were told to sum language use for each activity to 100% (e.g., Reading for fun––French: 50%, English: 50%, total: 100%). This wording is ideal for the computation of language entropy for each activity. In contrast, for conversations in various social contexts, participants were asked to sum use with each interlocuter type to 100% within a language (e.g., English use––with family: 10%, with friends: 40%, with classmates: 20%, with coworkers:

Table 2. Communicative contexts and measurement types elicited by the language history questionnaire. Information was elicited for the L1, L2, and L3

Communicative context	Original measurement type
Listening to radio/watching TV	Percentage (0%–100%)
Reading for fun	
Reading for work	
Reading on the internet	
Writing e-mails to friends	
Writing articles/papers	
Doing arithmetic (e.g., counting, numerical operations)	Likert scale (1–7)
Remembering numbers (e.g., student ID, telephone)	
Dreaming	
Thinking	
Talking to yourself	
Expressing anger or affection	
Speaking with family members	Percentage (0%–100%)
Speaking with friends	
Speaking with classmates	
Speaking with coworkers	

30%, total: 100%). This wording is less ideal for the computation of language entropy for each interlocuter type (though it could, in principle, be used to compute a type of interlocuter entropy within language). Crucially, because participants reported the percentage of all conversations that occur in each language across interlocuter types (e.g., French: 80% and English: 20%), we could derive the percentage of language use within each interlocuter type.

To derive the percent of language use within interlocuter type, we conducted the following procedure. We multiplied the proportion of usage for each interlocuter type within a language (e.g., English use with family) by the proportion of overall usage of that language (e.g., English use overall), yielding proportions of use that sum to 1 across all languages and interlocuter types. Then, within each interlocuter type, we divided each proportion by the sum total proportion within the interlocuter type, yielding proportions of usage for each language that sum to 1 within an interlocuter type.

Once all of the data were in common proportions, tracking language use in each context, we computed language entropy for each communicative context/interlocuter type. We also retained proportions of language use specific to the L2 for each context, and we used these data to index sheer L2 exposure.

Computing language entropy

For each of the 16 communicative contexts, we computed language entropy (H) with the following equation $H = -\sum_{i=1}^{n} P_i log_2(P_i)$ using the methods available in the languageEntropy R package (Gullifer & Titone, 2018). In this equation, n represents the total possible languages within a context and P_i is the proportion that language$_i$ is used within a context. Thus, if a bilingual reported using French 80% of the time and English 20% of the time within the work context, one computes language entropy by summing $0.80 \times log_2(0.80)$ and $0.20 \times lo_2(0.20)$, and then multiplying by –1 to yield a positive language entropy value. Language entropy for this hypothetical individual's work context is 0.72.

Theoretically, the entropy distribution has a minimum value of 0 when the proportion of usage for one language is 1.0 (and the other is 0), representing completely compartmentalized language usage within a communicative context. The distribution has a maximum value equal to log n (1 for two languages and approximately 1.585 for three languages) when the proportion of use for each language is equal, representing completely integrated language usage within a communicative context. This procedure resulted in 16 entropy scores for each participant that pertained to language entropy in each of the 16 communicative contexts.

Assessing objective language proficiency

Objective language proficiency was assessed through performance on verbal fluency tasks. Participants completed category and letter fluency tasks in French (category: fruits, letters: *F*, *L*, and *P*) and English (category: animals, letters: *A*, *F*, and *S*). In each portion of a task, they were asked to name as many exemplars as possible in 60 s, drawn from each category or letter. We computed the mean number of exemplars produced in each language across letter prompts, and because participants only completed one category in each language, we used the number of exemplars produced.

Assessing subjective language proficiency

Subjective language proficiency was extracted from the health and language background questionnaire (i.e., items probing language abilities). Participants reported their language abilities in the L1 and L2 for reading, writing, speaking, and listening. These data were elicited with the following prompt: "Please rate your current ability in reading, writing, speaking, and listening for all languages that you know according to the following scale [1-*very poor*—7-*native-like*]." They also reported the extent to which they believe they have an accent when speaking via the following prompt: "Do you have a foreign accent in the languages that you speak? Please rate how strong you think your accent is according to the following scale [1-*none*—7-*extremely strong*]." Of note, participants did not regularly report accentedness in the L1, and as such, we extracted only accentedness in the L2.

Table 3. The number of missing cases for each communicative context

Communicative context	Number of missing entropy scores (/89)
Listening to radio/watching TV	13
Reading for fun	25
Reading for work	20
Reading on the internet	4
Writing e-mails to friends	21
Writing articles/papers	37
Doing arithmetic (e.g., counting, numerical operations)	2
Remembering numbers (e.g., student ID, telephone)	1
Dreaming	4
Thinking	1
Talking to yourself	2
Expressing anger or affection	4
Speaking with family members	8
Speaking with friends	7
Speaking with classmates	28
Speaking with coworkers	25

Results

After extracting and computing language entropy, sheer L2 exposure, and language proficiency (verbal fluency performance and self-rated abilities), we prepared each data set for factor analysis. A primary concern was how to handle missing observations. All of the 16 communicative contexts measuring language usage and exposure (used to compute language entropy and sheer L2 exposure) had some missing observations (there were only 17 complete cases). Table 3 illustrates the number of missing cases for each communicative context. Missing data could have been due to many reasons, such as participants not engaging in a particular context (e.g., not listening to radio or TV) or reporting exposure with a specific language in that context (e.g., only listening to the radio/TV in their first language). We could not omit missing values, as the sample size would not be adequate for analysis, and omitting outliers can introduce statistical bias. Language entropy can be successfully computed when particular languages are not used within a communicative context, as the missing language(s) will be assumed to have exposure of 0. However, for instances where participants did not engage in a certain context at all with any language, language entropy would be unknown. Thus, we imputed language entropy data by partial mean matching using the methods available in the *mice* package in *R* (van Buuren & Groothuis-Oudshoorn, 2010). We did not impute values for sheer L2 exposure, and instead assumed that an unknown value was 0 in that context (similar to the assumptions of the language entropy method).
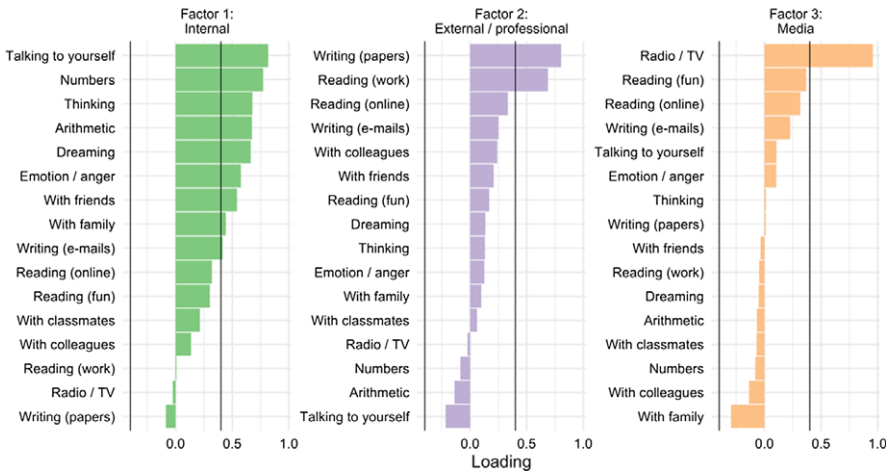
Figure 1. Illustration of latent structure for language entropy. The vertical axis depicts each communicative context for which language entropy was computed. The horizontal axis depicts the factor loading. Each factor (entropy F1, entropy F2, and entropy F3) is displayed as a separate panel. The entropy F1 factor appears to index language entropy for internal purposes or with friends and family. The entropy F2 factor appears to index language entropy in professional settings or for externally directed purposes. The entropy F3 factor appears to index language entropy for media consumption.

After preprocessing the data, we conducted factor analyses separately for language entropy, sheer L2 exposure, and language outcome data sets using methods available in the *psych* package (Revelle, 2017). The number of factors was determined through visual inspection of a parallel plot (via *fa.parallel*). For each analysis, we computed two goodness of fit metrics: the Kaiser–Meyer–Olin measure (KMO) and Bartlett's test. We assumed that factors were likely correlated, and thus we used oblique, *oblimin*, rotations.

What is the latent structure for language entropy?

We conducted a factor analysis on language entropy data for the 16 communicative contexts using the procedures outlined above. Goodness of fit metrics indicated that the language entropy data were factorable, Bartlett's test: $\chi^2 (120) = 607.785, p < .001$; KMO: 0.70. Of note, three items dropped below the 0.50 KMO threshold for sampling accuracy (language entropy for radio/TV, language entropy for writing papers, language entropy with colleagues). We determined that a three-factor solution was ideal to model language entropy data, and we thus computed an oblique factor analysis with three factors that, together, explained 42% of the variance. Factors were correlated at $|r| <= .18$.

Factor loadings for language entropy are illustrated in Figure 1. The first factor (entropyF1: 24.3% of the variance) appears to be related to language entropy for speaker internal purposes (e.g., counting, inner speech, dreaming, and thinking) and when communicating with friends and family. The second factor (entropyF2: 9.7% of the variance) appears to be related to language entropy for
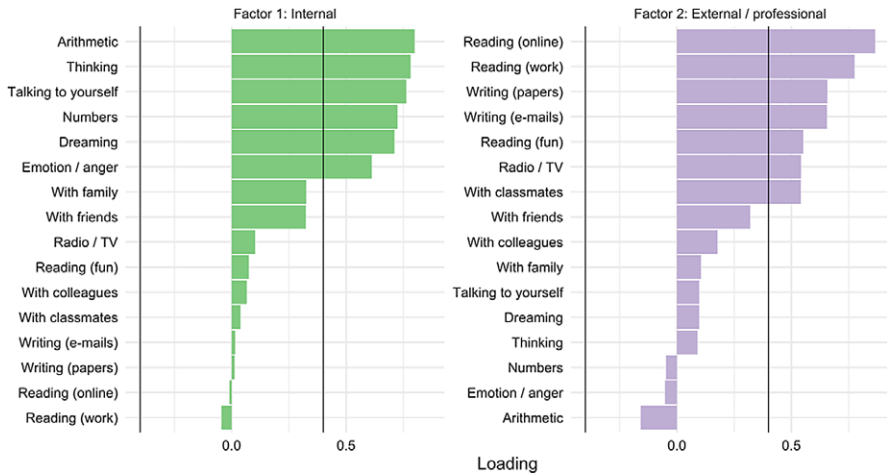
Figure 2. Illustration of latent structure for sheer L2 exposure. The vertical axis depicts each communicative context for which L2 exposure was extracted. The horizontal axis depicts the factor loading. Each factor (exposure F1 and exposure F2) is displayed as a separate panel. The exposure F1 factor appears to index L2 usage for internal purposes. The exposure F2 factor appears to index L2 usage in professional settings or for externally directed purposes.

externally directed or professional purposes. The third factor (entropyF3: 8.4% of the variance) appears to be related to language entropy for consumption of media.

What is the latent structure for sheer L2 exposure?

We next conducted a factor analysis on language history data that probed sheer exposure to the L2 in the 16 contexts above (percentage data and Likert data were both transformed to proportions of L2 usage), using the procedures outlined above. Tests indicated that the correlation matrix was factorable: KMO measure for the correlation matrix was 0.82, and no items dropped below the 0.50 threshold for sampling accuracy; Bartlett's test χ^2 (120) = 733.848, $p < .001$. We determined that a two-factor solution was ideal to model sheer L2 exposure data, and thus we computed an oblique factor analysis with *oblimin* rotation with two factors that explained 42% of the variance. The two factors were highly correlated at $r = .48$.

Factor loadings for L2 exposure are illustrated in Figure 2. The first factor (exposure F1: 21.7% of the variance) appears to be related to L2 use for speaker internal purposes (e.g., counting, inner speech, dreaming, and thinking). The second factor (exposure F2: 20.6% of the variance) appears to be related to L2 use for externally directed or professional purposes.

What is the latent structure for language proficiency?

We then conducted a factor analysis on language proficiency data (objective: category and letter fluency in L1 and L2; subjective: accentedness in L2, writing ability in L1 and L2, reading ability in L1 and L2, speaking ability in L1 and L2, and listening
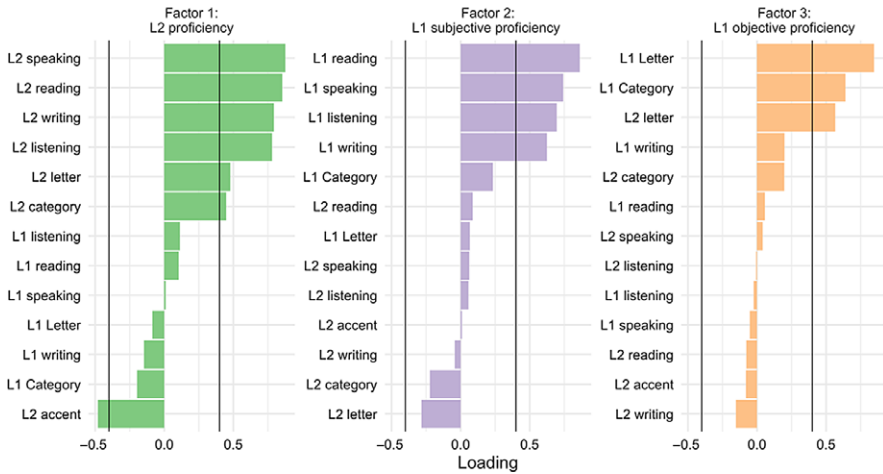
Figure 3. Illustration of latent structure for language outcomes. The vertical axis depicts each language outcome measure (e.g., letter fluency, category fluency, or self-rated ability item). The horizontal axis depicts the factor loading. Each factor (proficiency F1, proficiency F2, and proficiency F3) is displayed as a separate panel. The proficiency F1 factor appears to index subjective and objective L2 proficiency. The proficiency F2 factor appears to subjective L1 proficiency. The proficiency F3 factor appears to index objective L1 proficiency.

ability in L1 and L2), using the procedures outlined above. Tests indicated that the correlation matrix was factorable: KMO measure for the correlation matrix was 0.71, and no items dropped below the 0.50 threshold for sampling accuracy; Bartlett's test χ^2 (120) = 607.785, $p < .001$. We determined that a three-factor solution was ideal to model proficiency data, and thus we computed an oblique factor analysis with *oblimin* rotation with three factors that explained 57% of the variance. Factors were correlated at $|r| <= .13$.

Factor loadings for objective and subjective proficiency are illustrated in Figure 3. The first factor (proficiency F1: 26.8% of the variance), appears to be related to a combination of objective and subjective proficiency in the L2. The second factor (proficiency F2: 18.2% of the variance) appears to be related to subjective proficiency in the L1. The third factor (proficiency F3: 12.1% of the variance) appears to be related primarily to objective proficiency in the L1 with cross-loading from L2 letter fluency.

Does information about L2 exposure, language entropy, or language proficiency predict timing of language acquisition (simultaneous vs. sequential bilingualism)?

Once we had assessed the latent structures for sheer L2 exposure, language entropy, and language proficiency, our next goal was to determine whether individual scores on these factors jointly predict acquisition history (i.e., simultaneous, L1 English, or L1 French). Figures 4–6 illustrate the distributions of factor scores (Figure 4: L2 exposure, Figure 5: language entropy, and Figure 6: language proficiency) across the three
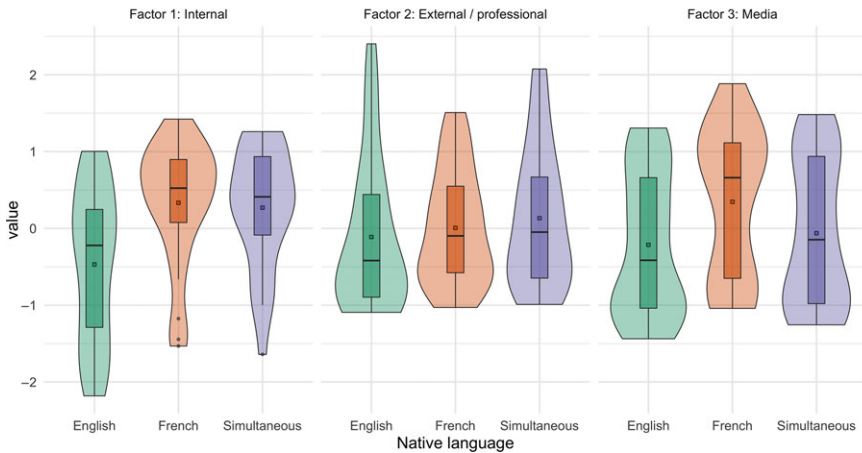
Figure 4. Distribution of language entropy factor scores for each acquisition group (simultaneous bilinguals; L1 English–L2 French bilinguals; L1 French–L2 English bilinguals). Boxplots and violin plots are displayed; square points illustrate group means.
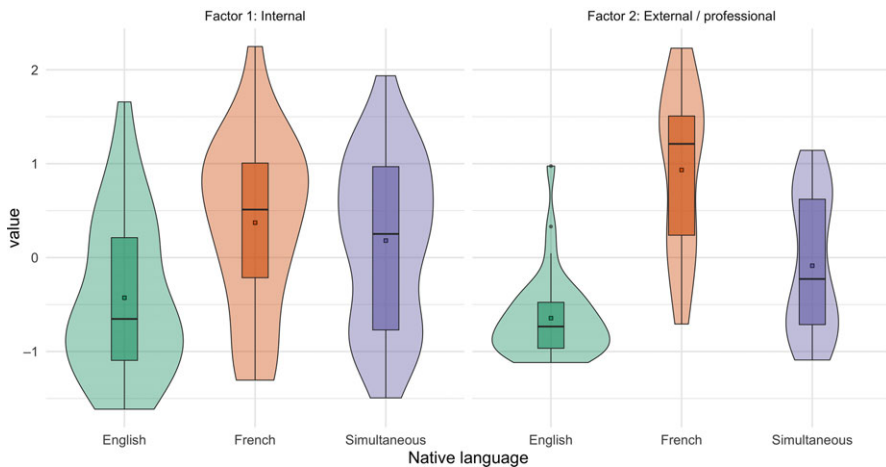

Figure 5. Distribution of sheer L2 exposure factor scores for each acquisition group (simultaneous bilinguals; L1 English–L2 French bilinguals; L1 French–L2 English bilinguals). Boxplots and violin plots are displayed; square points illustrate group means.

acquisition groups (simultaneous, L1 English–L2 French, and L1 French–L2 English). To assess whether factor scores could discriminate acquisition group membership, we conducted multinomial regression (multinom function in the nnet package; Ripley, 2002). We constructed seven models to predict acquisition group (three levels: simultaneous, English L1, and French L1) as a function of various combinations of the continuous factor scores. We selected the model that minimized the Akaike information criterion (AIC) as the best fitting model (see Table 4).

Table 4. Model comparisons for multinomial regressions predicting acquisition history (L1: simultaneous, L1 French, or L1 English) as a function of factor scores related to individual differences in sheer L2 exposure, language entropy, and language proficiency

df	Model	AIC	$AIC_i - AIC_5$
1	L1 ~ entropy F1 + entropy F2 + entropy F3	187.0	41.7
2	L1 ~ exposure F1 + exposure F2	148.1	2.8
3	L1 ~ proficiency F1 + proficiency F2 + proficiency F3	167.0	21.8
4	L1 ~ entropy F1 + entropy F2 + entropy F3 + proficiency F1 + proficiency F2 + perf F3	167.3	22.0
5	**L1 ~ exposure F1 + exposure F2 + proficiency F1 + proficiency F2 + proficiency F3**	**145.3**	
6	L1 ~ exposure F1 + exposure F2 + entropy F1 + entropy F2 + entropy F3	152.8	7.5
7	L1 ~ entropy F1 + entropy F2 + entropy F3 + exposure F1 + exposure F2 + proficiency F1 + proficiency F2 + proficiency F3	147.0	1.7

Note: AIC is depicted for each model. The best model with the lowest AIC is bolded. The difference between AIC for each model and the AIC for the best model is also depicted.
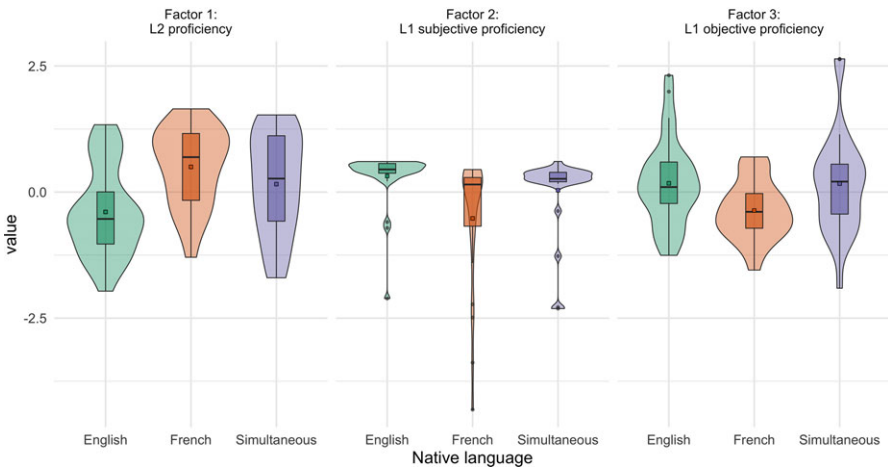

Figure 6. Distribution of language proficiency factor scores for each acquisition group (simultaneous bilinguals; L1 English–L2 French bilinguals; L1 French–L2 English bilinguals). Boxplots and violin plots are displayed; square points illustrate group means.

The best model included factor scores for sheer L2 exposure and language proficiency but no scores for language entropy ($AIC_{best} = 145.3$). However, comparing the AIC of the best model to the other models indicated that two other models may also be considered viable (see, e.g., Burnham & Anderson, 2004): there was "substantial support" (i.e., $AIC_{diff} < 2$) for the full model that included all factor scores (including language entropy; $AIC_{diff} = 1.7$) and "strong support" ($2 < AIC_{diff} < 4$)

Table 5. Summary table for the best model (Model 5)

	Dependent variable	
	English	French
	(1)	(2)
Constant	−0.543	−0.750
	(−1.394, 0.308)	(−1.568, 0.068)
Exposure F1	−0.586	−0.706
	(−1.494, 0.322)	(−1.778, 0.366)
Exposure F2	−1.265*	1.538**
	(−2.485, −0.044)	(0.492, 2.584)
Proficiency F1	−0.047	0.368
	(−0.830, 0.736)	(−0.578, 1.315)
Proficiency F2	0.869	−0.049
	(−0.235, 1.973)	(−0.767, 0.668)
Proficiency F3	−0.156	−1.079*
	(−0.814, 0.502)	(−2.118, −0.041)
AIC	145.3	
BIC	174.3	

$*p < .05. **p < .01.$

for the model that contained factor scores for sheer L2 exposure alone ($AIC_{diff} = 2.8$). All other models exhibited considerably less support (i.e., $AIC_{diff} > 4$).

We inspected the best model to determine the direction of the effects. We found that scores on the externally directed or professional factor (exposure F2) and the subjective L1 proficiency factor (proficiency F2) discriminated between the groups. Specifically, individuals with higher exposure F2 scores were more likely to have French as the L1 ($b = 1.54$, 95% confidence interval; CI [0.49, 2.58], $z = 3.706$, $p < 0.01$) and less likely to have English as the L1 ($b = −1.26$, 95% CI [−2.49, −0.04], $z = −3.474$, $p < 0.05$) relative to simultaneous bilinguals. Individuals with higher proficiency F2 scores were less likely to have French as the L1 ($b = −1.08$, 95% CI [−2.12, −0.04], $z = −2.179$, $p < 0.05$) relative to simultaneous bilinguals. See Table 5 for the full results of the best model.

Does information about acquisition history, L2 exposure, or language entropy predict language proficiency?

Our next goal was to determine whether individual factor scores for sheer L2 exposure, language entropy, and acquisition history (continuous effect of L2 AoA) jointly predict language proficiency (composed of the set of three-factor scores representing: L2 proficiency, subjective L1 proficiency, and objective L1

Table 6. Model comparisons for linear regressions predicting language proficiency factor scores as a function of factor scores related to individual differences in sheer L2 exposure, language entropy, and L2 AoA

df	Model	AIC DV = proficiencyF1	AIC DV = proficiencyF2	AIC DV = proficiencyF3
1	DV ~ AoA + entropy F1 + entropy F2 + entropy F3	213.6	235.1	217.9
2	DV ~ AoA + exposure F1 + exposure F2	193.2	**213.5**	**214.5**
3	DV ~ AoA + entropy F1 + entropy F2 + entropy F3 + exposure F1 + exposure F2	**190.7**	218.2	217.5
4	DV ~ AoA * (entropy F1 + entropy F2 + entropy F3 + exposure F1 + exposure F2)	196.5	224.7	225.9

Note: Separate models were run for each of the three factor scores: L2 proficiency (DV = proficiencyF1), L1 subjective proficiency (DV = proficiencyF2), and L1 objective proficiency (DV = proficiencyF3). AIC is depicted for the model of each fluency factor score. The best models (lowest AIC) are bolded.

proficiency). Thus, we conducted three sets of linear regressions for each language outcome component, with four models specified in each set. All regressions included an effect of L2 AoA. However, within each set of regressions, there were four parameterizations to assess the relative importance of entropy factor scores and exposure factor scores (separately and together), and to assess interactions between L2 AoA and the factor scores. For each of the three language proficiency factor scores, we selected the model that minimized AIC as the best fitting model. Model parameterizations and comparisons (with AIC) are illustrated in Table 6.

For models predicting L2 proficiency (proficiency F1), the best model included additive effects of sheer L2 exposure and language entropy scores ($\text{AIC}_{\text{best}} = 190.7$). Of note, another model with only sheer L2 exposure scores also received "strong support" ($\text{AICdiff} = 2.5$). See Table A.1 for the summary results of the four models. We inspected the best model to determine the direction of the effects. In this model, there was a significant effect related to factor scores for language entropy for externally directed or professional purposes (entropy F2; $b = 0.193$, 95% CI [0.011, 0.374], $t = 3.706$, $p < .05$). There was also a significant effect related to the sheer L2 exposure for internal purposes (exposure F1; $b = 0.819$, 95% CI [0.448, 1.190], $t = 4.32$, $p < .001$). These effects suggest that higher entropy F2 or exposure F1 scores were associated with higher L2 proficiency scores. No other effects were significant. Figure 7 provides an illustration of all effects in the model.

For models predicting subjective L1 proficiency (proficiency F2), the best model included additive effects of sheer L2 exposure scores ($\text{AIC}_{\text{best}} = 213.5$), and all other models exhibited considerably less support (i.e., $\text{AIC}_{\text{diff}} > 4$). See Table A.2 for the summary results of the four models. We inspected the best model to determine the direction of the effects. In this model, there was a significant effect related to the sheer L2 exposure scores for externally directed or professional purposes (exposure F2; $b = -0.514$, 95% CI [−0.746, −0.281],
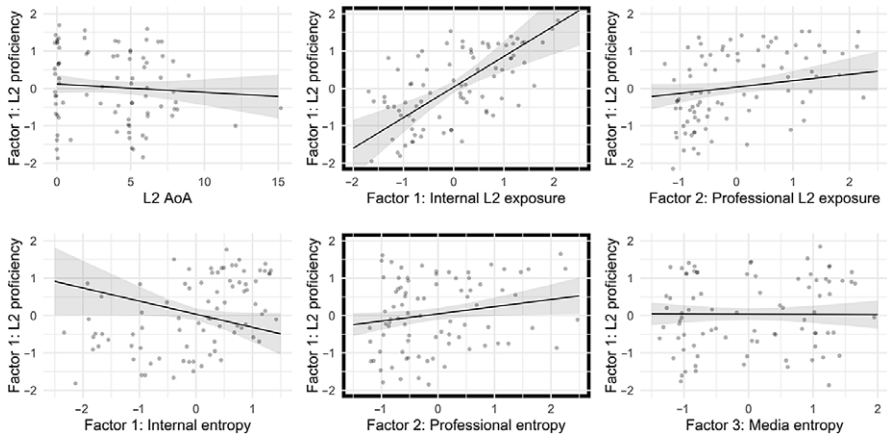
Figure 7. Illustration of model-predicted effects for proficiency F1 factor scores (i.e., L2 proficiency) as a function of L2 age of acquisition (AoA) and language exposure factor scores. Bolded outlines illustrate significant effects ($p < .05$). Increases in sheer L2 exposure for internal purposes were associated with increased L2 proficiency factor scores. Increases in professional language entropy were associated increased L2 proficiency factor scores. Confidence bands illustrate 95% confidence intervals.
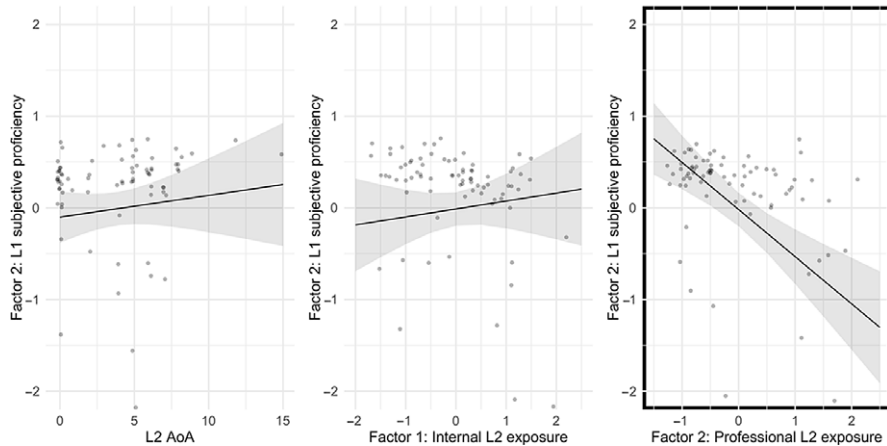

Figure 8. Illustration of model-predicted effects for proficiency F2 factor scores (i.e., subjective L1 proficiency) as a function of L2 area of acquisition (AoA) and language exposure factor scores. Bolded outlines illustrate significant effects ($p < .05$). Increases in sheer L2 exposure in professional settings were associated with reduced subjective L1 proficiency factor scores. Confidence bands illustrate 95% confidence intervals.

$t = 4.33$, $p < .001$). This effect suggests that higher exposure F2 scores are associated with lower subjective L1 proficiency. No other effects were significant. Figure 8 provides an illustration of all effects in the model.

For models predicting objective L1 proficiency (proficiency F3), the best model included additive effects of sheer L2 exposure ($AIC_{best} = 214.5$). Of note, two other models also received "strong support" (entropy alone, $AIC_{diff} = 3.4$; entropy and
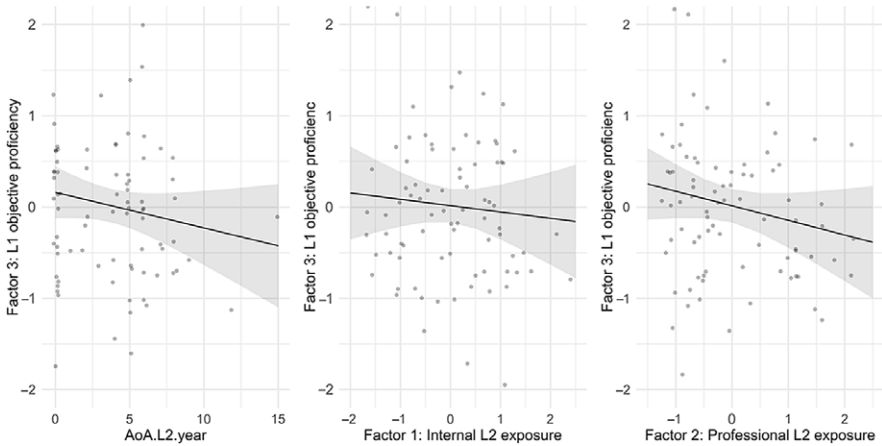
Figure 9. Illustration of model-predicted effects for proficiency F3 scores (i.e., objective L1 proficiency) as a function of L2 area of acquisition (AoA) and language exposure factor scores. Bolded outlines illustrate significant effects ($p < .05$). No effects were significant. Confidence bands illustrate 95% confidence intervals.

exposure, $AIC_{diff} = 3$). See Table A.3 for the summary results of the four models. We inspected the best model to determine the direction of the effects. There were no significant effects in this model (or any of the other "supported" models). Figure 9 provides an illustration of all effects in the model.

Discussion

Our goal was to assess the structural properties of and relationships between continuous constructs that jointly index the bilingual language experience, including the timing of language exposure, language entropy, sheer L2 exposure, and language proficiency (measured through objective and subjective measures).

As a marker of the timing of language exposure, we measured L2 AoA continuously. L2 AoA is a crucially important determinant of several aspects of linguistic, cognitive, and brain function (Berken et al., 2015, 2016; Flege et al., 1995; Gullifer et al., 2018; Johnson & Newport, 1989; Klein et al., 2014; Kousaie et al., 2017; Kousaie & Phillips, 2011). Of note, participants are often dichotomized into groups on the basis of L2 AoA (e.g., simultaneous bilinguals, L1 French bilinguals, and L1 English bilinguals). This approach was also taken here to provide a means of comparison. Of note, we also included measures of current exposure including sheer L2 exposure and language entropy. Successfully acquiring an L2 requires exposure to that L2, and measures of sheer L2 exposure are associated with several core language processing phenomena, such as the magnitude of lexical frequency effects during L2 and in L1 reading (Whitford & Titone, 2012).

Finally, we included measures of language entropy. Language entropy provides an estimate of balance or diversity in language exposure. While language entropy covaries with sheer L2 exposure (measures of L2 exposure figure into the

computation of entropy), it taps into a different theoretical construct as it provides an index of balanced exposure accounting for multiple languages.

We assessed each of these constructs continuously through data collected on a health and language background questionnaire with the additional inclusion of verbal fluency data to assess objective language proficiency. For constructs that tap into current language experience and exposure, we further assessed the variation that exists between and within communicative contexts providing granularity to the assessment. Specifically, for sheer L2 exposure and language entropy, we considered language usage for various purposes, including for internal thought (e.g., thinking and dreaming) and mental computations (e.g., counting and mental arithmetic); with friends and family, for school and work, and for the consumption of media. We also assessed L2 proficiency through both objective (verbal fluency) and subjective (self-report) measures.

We conducted factor analyses on these measures to assess the latent structures of these constructs: sheer L2 exposure, language entropy, and language proficiency. We showed that latent factor scores for sheer L2 exposure and language proficiency could discriminate between groups of bilinguals who differ in their acquisition history (i.e., simultaneous French–English bilinguals, L1 French–L2 English bilinguals, and L1 English–L2 French bilinguals). Latent factor scores for language entropy factor scores were relatively less important in discriminating the groups. We then showed that latent factor scores for sheer L2 exposure and language entropy were important in predicting language proficiency (L2 proficiency and subjective L1 proficiency but not objective L1 proficiency). We now discuss the core findings in turn.

Measuring current language exposure: Latent structure of sheer L2 exposure and language entropy

The factor analysis on measures of sheer L2 exposure revealed evidence for two latent factors indexing sheer L2 exposure (a) for internal-driven purposes (e.g., arithmetic and numerical operations, thinking talking to oneself, dreaming, and expressing emotion), and (b) for externally directed or professional purposes (e.g., reading online, reading for work, writing papers, and writing e-mails). The factor analysis on language entropy measures revealed a similar factor structure with the inclusion of an additional factor indexing language entropy for the consumption of media. Thus, there appears to be a fundamental distinction between language usage patterns for internally driven versus externally driven purposes.

These findings are compatible with other work that suggests a similar dichotomy between internal and external usage patterns (e.g., Schrauf, 2009; Vaid & Menon, 2000). For example, Schrauf showed that among older adult bilinguals who moved from Puerto Rico to the mainland United States, English (L2) usage patterns resulted in three factors: amount of English used for internally driven purposes (e.g., thinking, dreaming, counting, expressing emotion, swearing, and talking to oneself), among coworkers and family (workmates, significant other, offspring, in-laws, and family), and in social settings (friends and neighbors). The importance and structure of the internally driven measures are generally consistent across studies and suggests that most of the variance in the patterns of reported language usage can be explained by internal motivations of the participants.

Moreover, an individual's personal usage of language may be quite distinct from how that individual reports using language in the environment for communication. The loading of measures related to the social usage of language (e.g., patterns of usage among coworkers, or family and friends) tends to shift across studies conducted in different environments. Thus, these patterns of social language usage are fundamentally related to the broader environment in which language is used (e.g., French and English in Montréal vs. English and Spanish in New York).

Our findings extend past work on the latent structure of language entropy for French–English bilinguals living in Montréal (Gullifer & Titone, 2020a). Past work shows that language entropy is composed of two principal factors related to general language entropy (reading, speaking, home, and social entropy) and language entropy for professional settings (at work, with some cross-loading from social language entropy). The structural separation between work contexts and other contexts is generally supported in the present data set because measures indexing use in the work environment (together with other externally driven factors) loaded separately from language use for internal purposes, with family members, and with friends. The results also support the existence of another distinct component related to distributions of language use for the purposes of media consumption. However, this factor was generally weak, explaining only a small portion of variance and included only one major measure (usage for TV and radio). Future questionnaires should probe more specific types of media in order to more accurately measure this component (e.g., movies, online videos, podcasts, etc.).

Crucially, our detailed assessment of current bilingual language experience, including sheer L2 exposure and language entropy, represents an expansion on past work. When prior studies assess sheer L2 exposure, they tend to extract responses on a single questionnaire item (e.g., self-reported L2 exposure globally). However, the use of single, "global" items can be problematic, as participants may respond to these items with different contexts in mind, or they may be unable to accurately compute a mental average of language usage. Global items ultimately ignore potential variation within and across communicative contexts. Here we assessed sheer L2 exposure or use within several communicative contexts, including for internal thoughts and computations, with friends and family, for school and work, and for the consumption of media. This approach mirrors that taken for our assessment of language entropy in several communicative contexts. Of note, here we have expanded the number of communicative contexts for which we assessed language entropy relative to prior work. In the past, we assessed language entropy for the purposes of reading, speaking, talking with friends, talking with coworkers, and talking with family. That approach either leaves out or collapses across several crucial dimensions of bilingual language usage.

Although this study extends past work on language entropy, there are notable divergences that we point out here concerning the first latent variable. In past work, the first language entropy component comprised a variety of entropy measures, including for reading (generally), for speaking (generally), with family, and with friends. We thus referred to that latent component as "general language entropy." Here, the first latent factor is more precisely related to distributions of language usage for internal purposes and mental computations, for which we had multiple measures. Other measures, that indexed, for example, reading in various contexts, are distributed

across other factors (i.e., factors for externally driven or work purposes and media consumption). Of note, we also assessed reading in a more fine-grained manner than before (i.e., reading for fun, reading for work, or reading online), allowing for greater specificity of this domain. This shift in the factor loadings, particular when measures are assessed in a granular, context-dependent, manner, is important to consider for future studies. It also suggests that when language usage is probed "generally" or "globally" (e.g., overall percentage of time spent reading), participants may respond with a particular domain in mind, such as usage for internal purposes or with family members, as opposed to how they tend to use language in the everyday environment. Thus, for maximally sensitive measurement of usage and exposure, it is important to assess context-specific language usage.

We also found that the distributions of language usage with family and friends loaded onto this first factor, though the loadings were muted compared to other, more important, measures. This further suggests aspects of language use that are driven primarily by the participant's intentions and that these intentions are crucially shaped interactions with family and friends, which likely occur early in on life.

When comparing the factor structures for sheer L2 exposure and language entropy, there is considerable overlap. This is, perhaps, not entirely surprising because measures of L2 exposure (together with exposure for each other language reported by the participant) figured into the computation of language entropy. The overlapping results here suggest that the domains of language usage (i.e., the communicative contexts) that drive language use are common for language entropy and L2 exposure, at least for this population of primarily university students living in Montréal. However, for reasons we return to later, it is important not to conflate these two factors, as they tap into different theoretical constructs and may thus relate to different factors of bilingual language experience such as language control processes.

The factor analyses, and the questionnaire on which they are based, have several limitations that could be addressed in future studies. Specifically, future questionnaires should more systematically probe crucial contexts of language usage with a similar number of items. For example, here, the factor indexing media usage was dominated by a single item (listening to TV/radio), likely contributing to the poor sampling accuracy of this measure related to language entropy and to inconsistent loadings across language entropy and sheer exposure constructs. Moreover, it may be beneficial to include a similar number of items related to language production and language comprehension crossed by the written and spoken modalities to provide a fuller picture of bilingual language usage. This would give the most accurate assessment of language exposure and usage across communicative contexts. Finally, future questionnaires should require that participants respond to all questions related to the communicative contexts and explicitly specify if they do not use a language in a specific context to reduce missing data.

Measuring language proficiency: Latent structure of objective and subjective proficiency measures

The factor analysis on language proficiency measures revealed evidence for three latent factors indexing (a) overall L2 proficiency (objective and subjective), (b) subjective L1 proficiency, and (c) objective L1 proficiency. This suggests that

participants in our sample were generally able to judge their objective L2 performance through subjective measures, at least in a relative manner between participants. In other words, people who score high on the L2 verbal fluency task, tended to rate themselves more highly on the self-reported L2 measures (i.e., higher L2 abilities and lower L2 accent). The observation of convergent patterns between objective and subjective L2 proficiency measures stands in contrast to previous work indicating that bilinguals are not able to accurately judge their performance in the L2 (e.g., Tomoschuk et al., 2019). At the same time, subjective measures did not always correlate well with objective measures, reflected in the two latent factors indexing L1 objective proficiency and L1 subjective proficiency. This finding is somewhat surprising, because where individuals' thoughts about language ability diverged from objective performance was in the native language, which is frequently more dominant than the L2. Ultimately, these findings illustrate a key point made by Tomoschuk et al. that self-ratings (and by extension, the extent to which they line up with objective measures) may depend on aspects of the population being sampled, as some populations may tend to rate themselves higher or lower in line with cultural expectations and norms. Regarding the present participants, all participants were sampled from the Montréal area. This location is somewhat unique because bilingualism and multilingualism are the norm and there is contact between many languages, including mainly French and English. Thus, we echo calls to incorporate subjective and objective measures of proficiency (e.g., Surrain & Luk, 2019; Tomoschuk et al., 2019) so that the validity of each can be assessed.

Do individual differences in bilingual language experience discriminate acquisition history and predict language proficiency?

We then assessed whether individual differences along these continuous dimensions can be used to predict membership of participants among the three language acquisition groups (simultaneous, L1 English, and L1 French), and that individual differences in language entropy and sheer L2 exposure can be used to predict language proficiency. We generally found that the professional L2 exposure and the objective L1 proficiency factors discriminated between the groups. Individuals with higher professional L2 exposure scores were more likely to have French as the L1 and less likely to have English as the L1 relative to simultaneous bilinguals. Individuals with higher objective L1 proficiency scores were less likely to have French as the L1 relative to simultaneous bilinguals. This finding is likely related to characteristics of the sample. In particular, most of the participants were students at McGill University, an English-speaking campus. Thus, participants who speak French as a native language in this environment will inherently have greater exposure to the L2 in this setting. In contrast to the results for sheer L2 exposure, language entropy was not as important in discriminating the language groups. This finding is consistent with previous work showing only small relationships between language entropy and L2 AoA (Gullifer et al., 2018; Gullifer & Titone, 2020a).

In terms of predicting language proficiency as a function of continuous individual differences in bilingual language experience, we found that high L2 proficiency (a component representing subjective and objective proficiency) was associated with high sheer L2 exposure for internal purposes and high entropy associated with externally directed

or professional purposes. Subjective L1 proficiency decreased as a function of sheer L2 exposure in professional settings, but objective L1 proficiency was generally unrelated to these individual differences. In other words, individual differences in bilingual language experience were associated with L2 outcomes (self-reported and objective measures), but they were only associated with self-reported measures in the L1. These results confirm the idea that language proficiency is not a unitary construct and that people become proficient bilinguals by way of distributing usage of their languages differentially in different communicative contexts to achieve various communicative goals.

Regarding the distribution of language use, due to the fact that sheer L2 exposure and language entropy shared similar latent factor structures, one might argue that the two constructs are tracking the same underlying construct and that it would be difficult to effectively disentangle the two. Moreover, given that the introduction of language entropy results in yet another construct to measure in bilingualism, researchers may be tempted to opt for easier measures like global L2 exposure. However, we reiterate that a reliance on global measures may hold unintended consequences. Moreover, we have shown here sheer L2 exposure and language entropy exhibit different contributions in L2 proficiency depending on the communicative context.

Differential contributions of language entropy and sheer L2 exposure are most evident in our work in the domain of executive control and brain organization (Gullifer et al., 2018; Gullifer & Titone, 2020b). In that work, language entropy is the best predictor for measures of executive control relative to other individual differences measures (including exposure, and L2 AoA). The specific import of language entropy for control processes is predicted by theoretical perspectives like the adaptive control hypothesis (Abutalebi & Green, 2016; Green & Abutalebi, 2013; Green & Wei, 2014; Grosjean, 1997, 2001) and recent empirical evidence (Gullifer et al., 2018; Gullifer & Titone, 2020a; Sulpizio et al., 2019) whereby people adapt their control processes as a function of the language demands of the communicative or interactional context, and in particular the distributions of language usage that occurs in that context (e.g., in single language contexts vs. dual language contexts). Of note, predictions of these theories cannot be easily addressed through global L2 exposure with standard statistical practices that assume linear relationships between predictors and outcome variables because single language contexts and monolingual modes occur at both ends of the L2 exposure spectrum. Thus, language entropy should be considered in future work (perhaps jointly with other measures) if there is substantive reason to assess the relative distributions of language usage (e.g., balance of L1 – L2+ usage).

Bilingualism as a multidimensional spectrum of experiences

In this study, we offer another demonstration of how bilingual language experience can be quantified as a multidimensional spectrum of experiences. This perspective is increasingly becoming the norm in the field of bilingualism (e.g., Anderson, Hawrylewicz, et al., 2018; Baum & Titone, 2014; Dash et al., 2019; DeLuca et al., 2019, 2020; Gullifer et al., 2018; Gullifer & Titone, 2020a; Li, Legault, et al., 2014; Sulpizio et al., 2019; Tiv, Gullifer, et al., 2020). Previously, it was common to collapse bilinguals and monolinguals into groups. However, a dichotomous approach assumes that all participants are homogenous in their language

experience, and it washes away important variance that is now known to influence language processing, language control, cognitive control, and brain structure and function. Here we jointly account for multiple sources of variation that include, on the one hand, static historical aspects of language exposure and, on the other hand, current language experience.

Of importance, there are several dimensions of bilingual experience we did not measure or assess here that are, nonetheless, part of the bilingual experience. We did not assess the code-switching behavior of the participants or their experiences with translanguaging. Yet engagement in these flexible language practices is thought to be another important determinant of how the language and control systems of bilinguals become adapted through interactional contexts (e.g., Adler, Valdes Kroff, & Novick, 2020; Green & Wei, 2014). Moreover, while our measures of sheer L2 exposure and language entropy tap into the extent of current language exposure within an immersion context, they do not account for the duration of exposure. Duration of exposure has been shown to contribute to adaptive modulations in neuroanatomical brain structure and language processing (e.g., DeLuca et al., 2019; Dussias & Sagarra, 2007). A final key issue in this domain is validity, reliability, and generalizability of the measures and latent constructs that index the bilingual language experience, including L2 AoA, sheer language exposure, language entropy, and language proficiency. Future work should assess the joint contribution of all of these factors and their applicability to other populations. Many of these constructs are based on self-report, and there are concerns about whether participants are able to reliably assess their experiences. To ensure reliable and valid measures, detailed questionnaires (with multiple questions estimating each factor of interest) should be used to discern the constructs of interest. Sample sizes should be sufficiently large to employ statistical approaches that allow for the assessment of construct validity (such as factor analysis or structural equation modeling). Even objective measures should be validated to ensure that they are assessing the construct of interest (e.g., language proficiency) and not potentially erroneous constructs.

Conclusions

Bilingualism is a multidimensional spectrum of experiences composed of several continuous constructs associated with language usage and exposure across multiple languages. To build a complete picture of bilingualism, it is crucial to account for as many of these factors to the greatest extent possible. A wealth of literature suggests that the timing of language exposure and the amount of language exposure are crucial constructs to consider when studying bilingual and multilingual populations. However, these constructs do not fully capture the full range of bilingual experience; in particular, they do not capture the diversity associated with language use in the social world or for internalized purposes.

A key challenge in obtaining a full picture of the bilingual experience is the requirement to engage with a high dimensional space. We proffer language entropy as a means to estimate a core construct related to social language usage and as a means to help reduce complexity of the space. Language entropy is a highly flexible measure of language balance: it can be applied to situations where more than two languages are in use, and it can be applied to many types of data that provide information about

language usage. While we estimate language entropy from self-reported data, similar (and perhaps more precise) estimates could also be derived from naturalistic production. Moreover, we show how data reduction techniques such as factor analyses or principal component analysis can also be used to further reduce the dimensionality of a problem and to identify the latent structure for constructs of interest. A precise quantification of bilingualism in all of its complexity will better enable researchers to tackle broader problems and questions in bilingualism, language processing, language acquisition, and language control.

Note

1. Of note, various forms of entropy have been used previously to quantify language use in other domains: for example, entropy has been used to assess diversity in multilingual language use on Twitter (e.g., Eleta & Golbeck, 2014) and the distribution of programming language choice among software developers (Krein, MacLean, Delorey, Knutson, & Eggett, 2009). We have provided an *R* package to compute language entropy from language history data (available on github at https://github.com/jasongullifer/languageEntropy; Gullifer & Titone, 2018), and a related measure can be computed automatically by Li, Zhang, Yu, and Zhao (LHQ 3.0; https://blclab.org/lhq3/; 2019).

REFERENCES

Abutalebi, J., & Green, D. W. (2016). Neuroimaging of language control in bilinguals: Neural adaptation and reserve. *Bilingualism: Language and Cognition*, **19**, 689–698. doi: 10.1017/s1366728916000225

Adler, R. M., Valdes Kroff, J. R., & Novick, J. M. (2020). Does integrating a code-switch during comprehension engage cognitive control? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **46**, 741–759. doi: 10.1037/xlm0000755

Anderson, J. A. E., Hawrylewicz, K., & Bialystok, E. (2018). Who is bilingual? Snapshots across the lifespan. *Bilingualism: Language and Cognition*. Advance online publication. doi: 10.1017/s1366728918000950

Anderson, J. A. E., Mak, L., Keyvani Chahi, A., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, **50**, 250–263. doi: 10.3758/s13428-017-0867-9

Baum, S., & Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics*, **35**, 857–894. doi: 10.1017/S0142716414000174

Beatty-Martínez, A. L., & Dussias, P. E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language*, **95**, 173–189. doi: 10.1016/j.jml.2017.04.002

Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2020). Interactional context mediates the consequences of bilingualism for language and cognition. *Journal of Expimental Psychology: Learning, Memeroy, and Cognition*, **46**, 1022–1027 doi: 10.1037/xlm0000770

Berken, J. A., Chai, X., Chen, J.-K., Gracco, V. L., & Klein, D. (2016). Effects of early and late bilingualism on resting-state functional connectivity. *Journal of Neuroscience*, **36**, 1165–1172. doi: 10.1523/JNEUROSCI.1960-15.2016

Berken, J. A., Gracco, V. L., Chen, J.-K., & Klein, D. (2015). The timing of language learning shapes brain structure associated with articulation. *Brain Structure and Function*, **221**, 3591–3600. doi: 10.1007/s00429-015-1121-9

Bice, K., & Kroll, J. F. (2019). English only? Monolinguals in linguistically diverse contexts have an edge in language learning. *Brain and Language*, **196**, 104644. doi: 10.1016/j.bandl.2019.104644

Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign—Language skills: Implications for teachers and researchers. *Language Learning*, **39**, 313–338.

Burnham, K. P., & Anderson, D. R. (2004). Sociological methods & research. *Sociological Methods Research*, **33**, 261–304.

Commins, N. L. (1989). Language and affect: Bilingual students at home and at school. *Language Arts*, **66**, 29–43.

Dash, T., Berroir, P., Joanette, Y., & Ansaldo, A. I. (2019). Alerting, orienting, and executive control: The effect of bilingualism and age on the subcomponents of attention. *Frontiers in Neurology*, **10**. doi: 10.3389/fneur.2019.01122

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 7565–7574. doi: 10.1073/pnas.1811513116

DeLuca, V., Rothman, J., Bialystok, E., & Pliatsikas, C. (2020). Duration and extent of bilingual experience modulate neurocognitive outcomes. *NeuroImage*, **204**. doi: 10.1016/j.neuroimage.2019.116222

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish-English bilinguals. *Bilingualism: Language and Cognition*, **10**, 101–116.

Eleta, I., & Golbeck, J. (2014). Multilingual use of Twitter: Social networks at the language frontier. *Computers in Human Behavior*, **41**, 424–432. doi: 10.1016/j.chb.2014.05.005

Flege, J. E., Munro, M. J., & Mackay, I. R. A. (1995). Effects of age of 2nd-language learning on the production of english consonants. *Speech Communication*, **16**, 1–26. doi: 10.1016/0167-6393(94)00044-B

García, O., & Wei, L. (2012). Translanguaging. *The encyclopedia of applied linguistics* (pp. 1–7). New York: Wiley.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal Cognitive Psychology*, **25**, 515–530. doi: 10.1080/20445911.2013.796377

Green, D. W., & Wei, L. (2014). A control process model of code-switching. *Language, Cognition, and Neuroscience*, **29**, 499–511. doi: 10.1080/23273798.2014.882515

Grosjean, F. (1985). The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual & Multicultural Development*, **6**, 467–477.

Grosjean, F. (1997). The bilingual individual. *Interpreting*, **2**, 163–187.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, **1**, 131–149.

Grosjean, F. (2001). The bilingual's language modes. *One mind, two languages: Bilingual language processing*. Oxford: Blackwell.

Grosjean, F. (2016). The complementarity principle and its impact on processing, acquisition, and dominance. *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 66–84). Cambridge: Cambridge University Press.

Gullifer, J. W., Chai, X. J., Whitford, V., Pivneva, I., Baum, S., Klein, D., & Titone, D. (2018). Bilingual experience and resting-state brain connectivity: Impacts of L2 age of acquisition and social diversity of language use on control networks. *Neuropsychologia*, **117**, 123–134. doi: 10.1016/j.neuropsychologia.2018.04.037

Gullifer, J. W., & Titone, D. (2018). Compute language entropy with {languageEntropy}. Retrieved from https://github.com/jasongullifer/languageEntropy

Gullifer, J. W., & Titone, D. (2020a). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, **23**, 283–294. doi: 10.1017/s1366728919000026

Gullifer, J. W., & Titone, D. (2020b). Engaging proactive control: Influences of diverse language experiences using insights from machine learning. *Journal of Experimental Psychology: General*. doi: 10.1037/xge0000933

Jia, G., Aaronson, D., & Wu, Y. (2002). Long-term language attainment of bilingual immigrants: Predictive variables and language group differences. *Applied Psycholinguistics*, **23**, 599–621.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, **21**, 60–99.

Kałamała, P., Szewczyk, J., Chuderski, A., Senderecka, M., & Wodniecka, Z. (2020). Patterns of bilingual language use and response inhibition: A test of the adaptive control hypothesis. *Cognition*, **204**, 104373.

Klein, D., Mok, K., Chen, J.-K., & Watkins, K. E. (2014). Age of language learning shapes brain structure: A cortical thickness study of bilingual and monolingual individuals. *Brain and Language*, **131**, 20–24.

Kousaie, S., Chai, X. J., Sander, K. M., & Klein, D. (2017). Simultaneous learning of two languages from birth positively impacts intrinsic functional connectivity and cognitive control. *Brain and Cognition*, **117**, 49–56. doi: 10.1016/j.bandc.2017.06.003

Kousaie, S., & Phillips, N. A. (2011). Age-related differences in interlingual priming: A behavioural and electrophysiological investigation. *Aging Neuropsychology and Cognition*, **18**, 22–55. doi: 10.1080/13825585.2010.510555

Krein, J. L., MacLean, A. C., Delorey, D. P., Knutson, C. D., & Eggett, D. L. (2009, June). Language entropy: A metric for characterization of author programming language distribution. In *4th Workshop on Public Data about Software Development*.

Li, P., Legault, J., & Litcofsky, K. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex*, **58**, 301–324.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, **17**, 673–680.

Li, P., Zhang, F., Yu, A., & Zhao, X. (2019). Language history questionnaire (lhq3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*. Andvance online publication.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, **25**, 605–621. doi: 10.1080/20445911.2013.795574

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language Hearing Research*, **50**, 940–967. doi: 10.1044/1092-4388(2007/067)

Revelle, W. R. (2017). Psych: Procedures for personality and psychological research. Retrieved from https://CRAN.R-project.org/package=psych

Ripley, B. D. (2002). *Modern applied statistics with s.* London: Springer.

Schrauf, R. W. (2002). Comparing cultures within-subjects: A cognitive account of acculturation as a framework for cross-cultural study. *Anthropological Theory*, **2**, 98–115.

Schrauf, R. W. (2009). English use among older bilingual immigrants in linguistically concentrated neighborhoods: Social proficiency and internal speech as intracultural variation. *Journal of Cross-Cultural Gerontology*, **24**, 157–179.

Shannon, C. E. (1948). The mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423.

Sulpizio, S., Del Maschio, N., Del Mauro, G., Fedeli, D., & Abutalebi, J. (2019). Bilingualism as a gradient measure modulates functional connectivity of language and control networks. *NeuroImage*, **205**, 116306. doi: 10.1016/j.neuroimage.2019.116306

Surrain, S., & Luk, G. (2019). Describing bilinguals: A systematic review of labels and descriptions used in the literature between 2005–2015. *Bilingualism—Language and Cognition*, **22**, 401–415. doi: 10.1017/S1366728917000682

Tiv, M., Gullifer, J., Feng, R., & Titone, D. (2020). Using network science to map what Montréal bilinguals talk about across languages and communicative contexts. *Journal of Neurolinguistics*, **56**, 100913. doi: 10.31234/osf.io/bmry2

Tiv, M., O'Regan, E., & Titone, D. (2020). *Bilingual social cognition: Investigating the relationship between bilingual language experience and mentalizing*. Retrieved fron doi: 10.31234/osf.io/nbsxh

Tomoschuk, B., Ferreira, V. S., & Gollan, T. H. (2019). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition*, **22**, 516–536.

Vaid, J., & Menon, R. (2000). Correlates of bilinguals' preferred language for mental computations. *Spanish Applied Linguistics*, **4**, 325–342.

van Buuren, S., & Groothuis-Oudshoorn, K. (2010). Mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, **45**, 1–68.

Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, **19**, 73–80. doi: 10.3758/s13423-011-0179-5

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, **9**, 111–125.

APPENDIX A

TABLE A.1

	Dependent variable			
	Fluency F1			
	(1)	(2)	(3)	(4)
Constant	0.092	0.104	**0.129**	0.086
	(−0.190, 0.373)	(−0.145, 0.353)	**(−0.115, 0.372)**	(−0.182, 0.353)
AoA.L2.year	−0.012	−0.017	**−0.022**	−0.010
	(−0.069, 0.046)	(−0.068, 0.033)	**(−0.072, 0.028)**	(−0.064, 0.044)
Entropy F1	0.447**		**−0.351**	−0.273
	(0.245, 0.650)		**(−0.707, 0.005)**	(−0.934, 0.388)
Entropy F2	0.221*		**0.193***	0.346*
	(0.011, 0.430)		**(0.011, 0.374)**	(0.034, 0.659)
Entropy F3	−0.001		**−0.005**	0.100
	(−0.198, 0.197)		**(−0.176, 0.166)**	(−0.186, 0.386)
AoA.L2.year: exposureF1				0.027
				(−0.114, 0.168)
AoA.L2.year: exposureF2				0.046
				(−0.034, 0.125)
AoA.L2.year: entropyF1				−0.023
				(−0.148, 0.102)
AoA.L2.year: entropyF2				−0.028
				(−0.086, 0.030)
AoA.L2.year: entropyF2				−0.023
				(−0.078, 0.032)
Exposure F1		0.505**	**0.819****	0.719*
		(0.297, 0.713)	**(0.448, 1.190)**	(0.086, 1.352)
Exposure F2		0.199	**0.169**	−0.049
		(−0.007, 0.404)	**(−0.033, 0.371)**	(−0.424, 0.327)
AIC	213.6	193.2	**190.7**	196.5
BIC	228.1	205.3	**210.1**	228
Observations	83	83	**83**	83

(Continued)

TABLE A.1 (*Continued*)

	Dependent variable			
	Fluency F1			
	(1)	(2)	(3)	(4)
R^2	0.261	0.408	**0.465**	0.492
Adjusted R^2	0.223	0.385	**0.423**	0.413
Residual std. error	0.841 ($df = 78$)	0.748 ($df = 79$)	**0.725 ($df = 76$)**	0.731 ($df = 71$)
F statistic	6.883** ($df = 4$; 78)	18.146** ($df = 3$; 79)	**11.016** ($df = 6$; 76)**	6.243** ($df = 11$; 71)

*$p < .05$. **$p < .001$.

TABLE A.2

	Dependent variable			
	Fluency F2			
	(1)	(2)	(3)	(4)
Constant	−0.053	**−0.108**	−0.111	−0.087
	(−0.373, 0.267)	**(−0.389, 0.174)**	(−0.398, 0.176)	(−0.404, 0.230)
AoA.L2.year	0.010	**0.024**	0.025	0.021
	(−0.055, 0.076)	**(−0.034, 0.081)**	(−0.034, 0.084)	(−0.043, 0.086)
Entropy F1	−0.134		0.147	0.366
	(−0.365, 0.096)		(−0.273, 0.567)	(−0.417, 1.150)
Entropy F2	−0.008		0.046	0.036
	(−0.247, 0.230)		(−0.168, 0.260)	(−0.334, 0.407)
Entropy F2	0.011		0.058	−0.062
	(−0.213, 0.236)		(−0.144, 0.260)	(−0.402, 0.277)
AoA.L2.year:exposureF1				0.103
				(−0.064, 0.270)
AoA.L2.year:exposureF2				−0.071
				(−0.165, 0.023)
AoA.L2.year:entropyF1				−0.058
				(−0.207, 0.090)
AoA.L2.year:entropyF2				0.005
				(−0.064, 0.074)
AoA.L2.year:entropyF2				0.020
				(−0.045, 0.085)

(*Continued*)

TABLE A.2 (*Continued*)

		Dependent variable		
		Fluency F2		
	(1)	(2)	(3)	(4)
Exposure F1		**0.087**	−0.051	−0.400
		(−0.148, 0.322)	(−0.489, 0.387)	(−1.151, 0.350)
Exposure F2		**−0.514*****	−0.526***	−0.251
		(−0.746, −0.281)	(−0.765, −0.288)	(−0.697, 0.194)
AIC	235.1	**213.5**	218.2	224.7
BIC	249.6	**225.6**	237.6	256.2
Observations	83	**83**	83	83
R^2	0.021	**0.227**	0.239	0.270
Adjusted R^2	−0.029	**0.197**	0.179	0.157
Residual std. error	0.957 (*df* = 78)	**0.845 (*df* = 79)**	0.855 (*df* = 76)	0.866 (*df* = 71)
F Statistic	0.424 (*df* = 4; 78)	**7.724*** (*df* = 3; 79)**	3.971** (*df* = 6; 76)	2.387* (*df* = 11; 71)

*$p < .05$. **$p < .01$. ***$p < .001$.

TABLE A.3

		Dependent variable		
		Fluency F3		
	(1)	(2)	(3)	(4)
Constant	0.157	**0.159**	0.134	0.136
	(−0.132, 0.446)	**(−0.124, 0.442)**	(−0.152, 0.420)	(−0.184, 0.455)
AoA.L2.year	−0.037	**−0.039**	−0.031	−0.032
	(−0.096, 0.023)	**(−0.097, 0.019)**	(−0.089, 0.028)	(−0.097, 0.032)
Entropy F1	−0.070		0.212	0.233
	(−0.278, 0.138)		(−0.206, 0.630)	(−0.556, 1.022)
Entropy F2	0.095		0.114	0.175
	(−0.120, 0.310)		(−0.099, 0.327)	(−0.198, 0.548)
Entropy F2	−0.120		−0.109	−0.256
	(−0.323, 0.083)		(−0.311, 0.092)	(−0.598, 0.086)
AoA.L2.year:exposureF1				0.002
				(−0.167, 0.170)
AoA.L2.year:exposureF2				−0.021
				(−0.115, 0.074)

(*Continued*)

TABLE A.3 (*Continued*)

| | Dependent variable | | |
| | Fluency F3 | | |
	(1)	(2)	(3)	(4)
AoA.L2.year:entropyF1				0.005
				(−0.144, 0.155)
AoA.L2.year:entropyF2				−0.018
				(−0.087, 0.052)
AoA.L2.year:entropyF2				0.035
				(−0.030, 0.101)
Exposure F1		**−0.070**	−0.240	−0.297
		(−0.306, 0.167)	(−0.676, 0.196)	(−1.053, 0.459)
Exposure F2		**−0.160**	−0.156	−0.059
		(−0.393, 0.074)	(−0.393, 0.082)	(−0.507, 0.390)
AIC	217.9	**214.5**	217.5	225.9
BIC	232.5	**226.6**	236.9	257.4
Observations	83	**83**	83	83
R^2	0.052	**0.069**	0.101	0.118
Adjusted R^2	0.003	**0.033**	0.030	−0.018
Residual std. error	0.863 ($df = 78$)	**0.850 ($df = 79$)**	0.851 ($df = 76$)	0.872 ($df = 71$)
F Statistic	1.067 ($df = 4; 78$)	**1.943 ($df = 3; 79$)**	1.421 ($df = 6; 76$)	0.865 ($df = 11; 71$)