TURUN YLIOPISTO
UNIVERSITY OF TURKU

VERBAL FLUENCY TASKS AND LANGUAGE ATTRITION

Identifying Lexical Processing Strategies in Finnish-English Bilinguals Immersed in an English Environment

Nana Lehtinen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. B OSA – TOM. 705 | HUMANIORA | TURKU 2024

TURUN YLIOPISTO
UNIVERSITY OF TURKU

VERBAL FLUENCY TASKS AND LANGUAGE ATTRITION

Identifying Lexical Processing Strategies in Finnish-English Bilinguals Immersed in an English Environment

Nana Lehtinen

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. B OSA – TOM. 705 | HUMANIORA | TURKU 2024

University of Turku

Faculty of Social Sciences
Department of Psychology and Speech-Language Pathology
Speech-Language Pathology
Doctoral Programme of Social and Behavioural Sciences

Supervised by

Adjunct Professor Kati Renvall
University of Turku
Turku, Finland

Professor Marja-Liisa Helasvuo
University of Turku
Turku, Finland

Reviewed by

Professor Scott Jarvis
Northern Arizona University
Flagstaff, USA

Professor Mira Goral
The City University of New York
New York, USA

Opponent

Professor Scott Jarvis
Northern Arizona University
Flagstaff, USA

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

ISBN 978-951-29-9983-5 (PRINT)
ISBN 978-951-29-9984-2 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)
Painosalama, Turku, Finland 2024

UNIVERSITY OF TURKU
Faculty of Social Sciences
Department of Psychology and Speech-Language Pathology
Speech-Language Pathology
NANA LEHTINEN: Verbal Fluency Tasks and Language Attrition: Identifying Lexical Processing Strategies in Finnish-English Bilinguals Immersed in an English Environment
Doctoral Dissertation, 186 pp.
Doctoral Programme of Social and Behavioural Sciences
December 2024

ABSTRACT

Language is a living, dynamic entity that constantly evolves to meet the social needs of its speakers. For bilingual individuals, these social requirements can vary significantly depending on their linguistic environment. When social shifts and changes in language contexts lead to reduced use and exposure of the first language in an immersive second-language environment, the resulting changes in the native language and its usage can be referred to as first language attrition. These changes can be observed, for example, among emigrants who have left their native country and lived for an extended period in a second-language environment.

Changes in the native language and its use are frequently observed at the lexical level, both in informal conversations and formal lexical tasks. Verbal Fluency (VF) tasks are a versatile tool for investigating lexical retrieval. Language attrition research shows that immersed bilinguals tend to produce fewer words in VF tasks compared to monolinguals. However, the underlying factors that hinder optimal performance have not been systematically identified, and the value of these tasks as a research tool in language attrition studies has been questioned. While VF total scores offer a general measure of task performance, they do not provide insight into the processing strategies underlying lexical retrieval, such as frequency and type of errors, the temporal distribution of words, and the clustering and switching patterns employed during VF task performance.

This dissertation by publication consists of three studies. The overarching aim of the dissertation is to identify processes underlying VF task performance in a group of immersed Finnish-English bilinguals living in Northern California. The first study (Study I) introduced and validated a systematic method for comprehensive analysis of VF task performance, including total scores, errors, temporal parameters, and clustering and switching strategies. The subsequent articles (Study II and Study III) applied this analysis method to VF task data from immersed bilinguals living in Northern California, US in the native language, Finnish (L1), and their second language, English (L2). Task performance in L1 was also compared to that of a control group of Finnish speakers living in Finland. Two extralinguistic variables were included to consider the impact of the shared language background on VF task performance in the immersed bilingual group: duration of immersion in the L2 environment and the frequency of L1 use during this time. Studies II and III demonstrate that the analysis methods introduced in Study I can effectively identify similarities and differences in language processing strategies employed during VF task performance.

Immersed bilinguals performed similarly in L1 and L2 VF tasks. Study II demonstrated that more frequent L1 use was associated with a higher total score, and this positive influence extended to both languages. These findings suggest that the observed strategies utilized in VF task performance are not limited to L1 but reflect general language processing strategies within the immersed bilingual group, emphasizing the importance of including both languages of immersed bilingual speakers in language attrition studies.

Differences in L1 performance between the immersed bilingual group and the control group living in Finland were detected in Studies II and III, suggesting potential cognitive adaptations in the group of immersed bilinguals. Immersed bilinguals utilized rapid lexical retrieval more efficiently than the control group. They also relied more systematically on clustering to reach a higher total score, while appearing to lack flexibility in switching between clusters once a cluster was exhausted. This could indicate that immersed bilinguals rapidly generated words belonging to a single cluster at the beginning of the task (e.g., cluster “pets” in the category “animals”) but were less successful in switching to a new category or in returning to a previous cluster (e.g., “pets” - “jungle animals” - “pets”), resulting in reduced overall performance. The findings from this dissertation offer cautious optimism that the proposed method could identify group specific characteristics of lexical retrieval in bilingual groups immersed in a second language environment.

In conclusion, this dissertation demonstrates that a detailed analysis of VF task performance can enhance our understanding of the processes underlying lexical retrieval in varied bilingual groups. In addition to language attrition studies, the method proposed in Study I could be applied across diverse populations and settings. Considering that VF tasks are among the most common tasks for evaluating lexical retrieval in neuropsychological language assessments, the application of the proposed analysis methods extends to various research fields, including, but not limited to, speech pathology, neuropsychology, and linguistics.

KEYWORDS: Language attrition, verbal fluency task, bilingualism, semantic verbal fluency, phonemic verbal fluency

TURUN YLIOPISTO
Yhteiskuntatieteellinen tiedekunta
Psykologian ja logopedian laitos
Logopedia
Nana Lehtinen: Sanasujuvuustehtävät ja ensikielen attritio: Leksikaaliset prosessointistrategiat englanninkielisessä kieliympäristössä asuvien suomi-englanti kaksikielisten puhujien tehtäväsuoriutumisessa
Väitöskirja, 186 s.
Yhteiskunta – ja käyttäytymistieteiden tohtoriohjelma
Joulukuu 2024

TIIVISTELMÄ

Kieli on elävä, dynaaminen kokonaisuus, joka muuntuu vastaamaan puhujansa sosiaalisia tarpeita. Sosiaaliset tarpeet muuttuvat elämäntilanteiden vaihtuessa ja elämäntilanteiden muutoksiin voi liittyä merkittäviä kieliympäristön vaihdoksia. Kun kieliympäristön vaihtuminen johtaa äidinkielen käytön vähenemiseen toisen kielen kieliympäristössä, muutoksia kaksikielisten puhujien äidinkielessä ja sen käytössä kutsutaan *ensikielen attritioksi.* Ensikielen attritiota tavataan esimerkiksi siirtolaisilla, jotka ovat muuttaneet pois synnyinmaastaan ja jotka ovat asuneet pitkään toisen kielen kieliympäristössä.

Sanastotason attritiota voidaan havaita sekä vapaassa puheessa että muodollisissa tehtävissä. Sanasujuvuustehtävät ovat monipuolinen työkalu sanahaun tutkimiseen. Kieliattritiotutkimuksissa on todettu, että siirtolaiset tuottavat äidinkielellään sanasujuvuustehtävissä vähemmän sanoja verrattuna yksikielisiin puhujiin. Alentuneen suorituksen taustalla vaikuttavia tekijöitä ei kuitenkaan ole systemaattisesti tunnistettu, ja tehtävien informaatioarvoa kieliattrition tutkimusmenetelmänä on kyseenalaistettu. Sanasujuvuustehtävien kokonaispistemäärät antavat yleisen kuvan tehtävässä suoriutumisesta, mutta eivät kerro strategioista, joita puhujat käyttävät sanahaussa. Sanahaun strategioita voidaan tutkia analysoimalla tehtävässä tuotettuja virheitä, sanojen ajallista jakautumista tehtävässä tai sanojen tuottamisessa hyödynnettyjä sanakategorian sisäisiä klusterointi- ja vaihtostrategioita.

Tämä väitöskirja koostuu kolmesta julkaisusta (tutkimus I, II ja III). Väitöskirjan kokonaistavoitteena on tunnistaa prosesseja, jotka ohjaavat Pohjois-Kaliforniassa asuvien suomalaisten siirtolaisten suoriutumista sanasujuvuustehtävissä. Ensimmäisessä artikkelissa esiteltiin ja validoitiin systemaattinen menetelmä sanasujuvuustehtävien kattavaan analysointiin, mukaan lukien kokonaispistemäärä, virheet, ajalliset parametrit sekä sanakategorian sisäiset klusterointi- ja vaihtostrategiat. Tutkimuksissa II ja III tätä menetelmää sovellettiin Pohjois-Kaliforniassa asuvien suomalaisten siirtolaisten suoriutumiseen sanasujuvuustehtävissä. Tehtäväsuoriutumisen taustalla vaikuttavia prosesseja vertailtiin siirtolaisten äidinkielessä (suomi, L1) ja ympäristön käyttämässä kielessä (englanti, L2). Tutkimuksissa arvioitiin myös kahden taustamuuttujan vaikutusta tehtäväsuoriutumiseen. Nämä muuttujat olivat aika, jonka siirtolaiset ovat asuneet toisen kielen kieliympäristössä, sekä äidinkielen käyttöfrekvenssi tänä aikana. Lisäksi molemmissa tutkimuksessa verrattiin siirtolaisten tehtäväsuoriutumista Suomessa asuvista suomenpuhujista muodostetun verrokkiryhmän tehtäväsuoriutumiseen.

Tutkimuksessa I esitetyillä analyysimenetelmillä voitiin tunnistaa samankaltaisuuksia ja eroja sanasujuvuustehtävässä suoriutumisen taustalla vaikuttavissa kielellisissä prosesseissa. Tutkimukset II ja III osoittivat, että siirtolaiset hyödynsivät samankaltaisia prosesseja äidinkielessä ja toisessa kielessään, ja ne siirtolaiset, jotka käyttivät äidinkieltä arjessaan useammin, tuottivat enemmän sanoja molemmilla kielillä. Nämä huomiot viittaavat siihen, että siirtolaisten sanahaussa ilmenevät tyyppipiirteet ulottuvat yleisiin kielellisiin prosessointistrategioihin sen sijaan, että muutokset rajoittuisivat äidinkieleen. Tulosten perusteella voidaan todeta, että kieliattritiotutkimuksia ei tule rajoittaa ainoastaan äidinkielessä tapahtuviin muutoksiin vaan kaksikielisyyden vaikutuksen arviointi on tärkeä ulottaa molempiin kieliin.

Siirtolaisten ja Suomessa asuvien verrokkien tehtäväsuorituksissa havaittiin eroja tutkimuksissa II ja III. Siirtolaiset käyttivät puoliautomaattista, nopeaa sanahakua tehokkaammin kuin verrokit. He myös turvautuivat systemaattisesti klusterointistrategiaan saavuttaakseen korkeamman kokonaispistemäärän mutta eivät vaikuttaneet hyödyntävän vaihtostrategioita yhtä tehokkaasti kuin verrokit. Tämä viittaa siihen, että siirtolaiset mahdollisesti tuottivat nopeasti yhteen klusteriin kuuluvia sanoja tehtävän alkuvaiheessa (esim. klusteri lemmikit eläinkategoriassa) mutta eivät onnistuneet vaihtamaan joustavasti kategorioiden välillä tai palaamaan aiempiin kategorioihin, kun kategoria oli tyhjennetty (esim. lemmikit - viidakon eläimet - lemmikit). Tehdyt havainnot viittaavat mahdollisiin kognitiivisiin sopeutumisstrategioihin siirtolaisten ryhmässä, ja nämä sopeutumisstrategiat saattavat heijastua tehtäväsuoriutumiseen alentuneena kokonaispistemääränä.

Väitöskirjatutkimus osoittaa, että sanasujuvuustehtävien yksityiskohtainen analyysi voi lisätä tietoa sanahaun prosesseista ja niissä tapahtuvista muutoksista siirtolaisten kielten tutkimuksessa. On myös huomattava, että tässä tutkimuksessa esitellyt analysointimenetelmät eivät rajoitu kaksikielisyystutkimukseen. Sanasujuvuustehtävät ovat yksi yleisimmistä tehtävistä neuropsykologisissa testipatteristoissa. Hyödyntämällä tässä tutkimuksessa esitettyjä analyysimenetelmiä erilaisissa tutkimusasetelmissa voidaan lisätä tietoa erilaisten kohderyhmien kielellisissä toiminnoissa tapahtuvista muutoksista eri tutkimusaloilla, mukaan lukien logopedian, neuropsykologian sekä kielitieteen alojen tutkimus.

ASIASANAT: Kieliattritio, sanasujuvuustehtävä, kaksikielisyys, semanttinen sanasujuvuus, foneeminen sanasujuvuus

Acknowledgements

The story of this research project began in Glasgow, Scotland. It was at the library of the University of Strathclyde where I, an undergraduate exchange student from the University of Oulu, Finland, stumbled upon a book about language attrition. As a student of Speech and Language Pathology with a strong interest in acquired communication deficits in adults, the idea of gradual, non-pathological language loss was fascinating to me.

Upon returning home, I was fortunate to have the support of Prof. Matti Lehtihalmes, who encouraged my interest in this multidisciplinary field. He advised me to seek insights from various areas and to foster a balance of collaboration and independence that has shaped my work since. This guidance allowed me to complete my Master's thesis on language attrition in 2004, supported by my supervisors Prof. Matti Lehtihalmes, Timo Lauttamus, Pekka Hirvonen, and Harri Mantila. During this time, I also had the privilege of meeting Prof. Monika Schmid, who graciously invited me to join the early European Graduate Network on Language Attrition meetings in Amsterdam. Her enthusiasm, dedication, and generosity continue to inspire me to this day. While decades have passed since these discussions, I thank the aforementioned mentors for opening my eyes to the possibility of pursuing a PhD in this field.

However, after graduating, the fascinating world of adults with acquired communication deficits led me to pursue other professional goals. Working as a speech therapist in this field has been an invaluable education, providing insights into various language processing challenges. One of my favorite tools in learning about language processing is the focus of this study—the verbal fluency task. I would like to express my sincere gratitude to the individuals I have worked with, as well as to my colleagues and dear friends at the speech therapy practice Puheklinikka, whose support and insights have been instrumental to both my professional development and the progression of this research.

The next chapter of this research project began in 2014, ten years after my Master's thesis, when I relocated from Turku, Finland, to California, USA. As a new emigrant myself, the idea of building a dataset of Finnish spoken in Northern California slowly began to take shape. I owe my deepest gratitude to Dr. Lotta

Weckström at the University of California, Berkeley, who not only convinced me this project was possible but facilitated connections with many of the participants in this study. Lotta also introduced me to the Migration Institute of Finland, linking my work to the broader field of migration studies. Throughout this project, Lotta has remained a valuable mentor and friend, always encouraging me and ensuring progress, especially during the times when only baby steps were being made.

Fittingly, this current chapter of my research project is being written again ten years later, in 2024. Throughout my PhD journey, I have received invaluable support from diverse perspectives. I am deeply grateful to my supervisors, Adj. Prof. Kati Renvall and Prof. Marja-Liisa Helasvuo, who embraced my PhD proposal with open arms and a can-do attitude from the start. They have always been readily available for guidance while also allowing me the freedom to explore and learn at my own pace when needed, thus enabling this project to move forward. Kati has provided consistent support within the Department of Speech and Language Pathology, even from afar. Her attention to detail and perseverance in getting things just right have perfectly balanced my enthusiastic—and at times impatient—approach. Beyond her personal support, Kati has provided valuable resources, including the opportunity to collaborate with dedicated research assistants who happily tapped into their networks to locate suitable control participants and offered their help in data recording and analysis. Marja-Liisa has provided diligent and constructive feedback to guide the overall process. She has generously shared her deep knowledge of the Finnish language in designing experimental tasks for data collection. Through her, I was also introduced to the possibility of depositing the data in the Language Bank of Finland (Kielipankki), enabling broader explorations of Finnish spoken in California in future studies. I truly admire her ability to quickly pinpoint areas needing attention and move things forward efficiently, all while maintaining a thoughtful approach and an incredibly reassuring presence.

Alongside my supervisors, I owe my deepest gratitude to my colleagues, co-authors, and friends, MPhil Ida Luotonen and Dr. Anna Kautto. They were crucial in encouraging me to rethink the scope of this dissertation and to dive deeper into one task from the collected data. Working with Ida to develop the task analysis method used in this study was an absolute delight. I hope our discussions on the potential of these tasks continue far beyond this dissertation. To Anna, I owe my heartfelt thanks for sparking my interest in the possibilities of statistical analysis. This dissertation would not have been possible without her guidance, expertise, patience, and plain hard work. I am in awe of how she breaks down complex concepts into bite-sized pieces of information, leaving you eager to explore further. I sincerely hope this is just the beginning! I also wish to extend my gratitude to Adj. Prof. Raymond Bertram for the brainstorming sessions and preliminary analysis

related to my research. I look forward to transforming the ideas inspired by these discussions into future studies.

During the data collection for this research project, I had the privilege of meeting and interacting with amazing people who generously shared their stories with me. I am deeply grateful to the participants for their willingness to take part in this project. I would also like to acknowledge the participants in both groups for kindly agreeing to share the collected data for future research. As for the Language Bank of Finland, I wish to thank Mietta Lennes for her patience while guiding me in preparing the data for deposit, and for providing this invaluable service and resource for researchers. This project was financially supported by the Turku University Foundation, Finlandia Foundation National, Migration Institute of Finland, UTUGS, and the Alfred Kordelin Foundation.

Throughout this process, the support of the language attrition research community has been incredibly uplifting. While progress at times has been slow, the energy gained from these interactions has kept it steadily moving forward. I eagerly await the next chapter of this story. Hopefully, it will not take another ten years to take shape!

Finally, to my family and friends, near and far. You inspire and amaze me every day. I am lucky to exist in this time and space with you all.

December 2024

Nana Lehtinen

Table of Contents

Figures

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Lehtinen, N., Luotonen, I., & Kautto, A. (2023). Systematic administration and analysis of verbal fluency tasks: Preliminary evidence for reliable exploration of processes underlying task performance. *Applied Neuropsychology: Adult*, *30*(6), 727–739. https://doi.org/10.1080/23279095.2021.1973471

II Lehtinen, N., Kautto, A., & Renvall, K. (2024). Frequent native language use supports phonemic and semantic verbal fluency in L1 and L2: An extended analysis of verbal fluency task performance in an L1 language attrition population. *International Journal of Bilingualism, 28(5),* 884–906. https://doi.org/10.1177/13670069231193727

III Lehtinen, N., Kautto, A., & Renvall, K. (2024). Efficacy of clustering and switching strategies in verbal fluency tasks in a Finnish-English language attrition population. *The Language Learning Journal*, *52*(2), 218–231. https://doi.org/10.1080/09571736.2023.2294060

1 Introduction

People in modern societies live, learn, interact, and work in versatile language environments. It is estimated that approximately half of the world's population are bilingual (Grosjean, 2010; Grosjean, 2013). Definitions of bilingualism vary widely, ranging from requiring early exposure to two languages and achieving native-like fluency in both, to encompassing any ability to speak, read, or understand a second language at any level of proficiency (Grosjean, 2021; Romaine, 1995). Bilinguals acquire and use their languages in different contexts and for different purposes to enable maximal, functional communication. These purposes can vary a great deal during the lifespan of a bilingual individual, often resulting in fluctuating outcomes in terms of language proficiency and use (Grosjean, 2010).

First language attrition is a distinctive phenomenon within the broad spectrum of bilingualism. It refers to the gradual, non-pathological change in a native language when the use of the native language is reduced and multiple languages are activated over extended periods (Gallo et al. 2021; Schmid, 2011b; Schmid & Köpke, 2017a; Schmitt & Sorokina, 2024). Favorable conditions for first language attrition arise when young adults migrate to a different country, and a proficiently mastered native language (L1) falls into disuse or is used in conjunction with an environmental language (L2) over a prolonged period (Schmid, 2011b; Schmid, 2019; Schmid & Jarvis, 2014; Schmid & Köpke, 2017a; Schmid & Yilmaz, 2018). The International Organization for Migration (2024) reports that the global migrant population is steadily growing with nearly 281 million people currently residing outside their country of birth—a threefold increase since 1970. Migration poses varied changes in social requirements and language environment circumstances, potentially accelerating the changes observed in involved languages and their use (Aitchison, 2013; Sharwood-Smith & van Buren, 1991).

The history of Finnish emigration to the USA is rather well documented (e.g., Kostiainen, 2014). During the Great Migration (1870-1945), approximately 370 000 Finns emigrated to the US, creating some of the largest Finnish emigrant communities (Library of Congress, 2024; Mikkonen, 2024). These early emigrates tended to reside in close-knit communities. In some areas, it was possible to continue using Finnish as the primary language in everyday life for the first-generation emigrants, with their children learning English as they entered the school system (Hirvonen & Lauttamus, 2000; Martin & Jönsson-Korhola, 1993). The majority of studies on the Finnish language in the US focus on the language of early emigrants and their descendants. These language corpora are mainly collected between 1960 and 1990, often with minimal documentation on the background of the participants (e.g. Hirvonen & Lauttamus, 2000; Virtaranta, 1993).

In the modern society, the ease of travel and communication across borders can facilitate the decision to relocate or emigrate. According to the 2019 US Census, 22% of the population in the US reported speaking a language other than English at home, with the ability to speak English among these individuals improving in the recent years. Language diversity within the US is also projected to grow as various groups aim to maintain the use of their native language as part of their cultural identity (Dietrich & Hernandez, 2022). This suggests a trend towards balanced bilingualism among emigrants partly due to recent emigrants arriving with a higher level of English proficiency (Dietrich & Hernandez, 2022).

In Europe, 54% of residents are able to hold a conversation in at least one additional language to their mother language and the EU promotes educational policies that support multilingualism (European Commission, 2012). It can be argued that this number is even higher in Finland. Finland has two official languages—Finnish and Swedish—and the education system supports multilingualism. During primary education, students are required to study at least two additional languages beyond their first language (Oikeusministeriö, 2024; Opetushallitus, 2024; Perusopetuslaki, 1998). Thus, it is reasonable to assume that more recent Finnish emigrants are well equipped to function as modern world citizens. These recent emigrants tend to have less need or inclination to settle in close-knit Finnish communities, enabling them to immerse themselves more fully in the English-speaking environment than earlier generations (Leinonen, 2014). This leads to a more pronounced shift in their language environment and reduction in exposure to their first language (L1).

Finnish is considered a small language, spoken as a first language by the majority of the Finnish population (85%), yielding to approximately 4.8 million speakers (Tilastokeskus, 2024). In the US, Finnish Americans represent a small minority. According to the 2019 U.S. Census 606 028 individuals reported Finnish ancestry, with 44 353 residing in California (U.S. Census Bureau, 2022a). Notably, only 4.7% of these individuals reported using language other than English at home, and just 0.5% indicated that they speak English less than "very well" (U.S. Census Bureau, 2022b). As a member of the Finno-Ugric language family, Finnish also has unique linguistic characteristics (Helasvuo, 2008). Thus, the variant of the modern Finnish language spoken by Finnish emigrants in the US provides an intriguing opportunity for research on first language attrition.

1.1 Key Concepts

Within the broad field of bilingualism, *language attrition* can be studied, described, documented, explained, and understood in the same way as any other phenomenon concerning language (Andersen, 1982). Research on language attrition focuses on changes at the individual level, measured as changes in language ability, performance, and use (Schmid, 2011b; Schmitt & Sorokina, 2024). Language attrition studies explore individual language performance and seek to understand the impact extralinguistic factors have on language and its use. The term language attrition can also refer to loss of a second or foreign language (Mehotcheva & Köpke, 2019) or language loss during childhood, when native language input is reduced or stops due to environmental changes before the first language is fully mastered (e.g., during an international adoption process) (Pierce et al., 2019). In this dissertation, I use the term language attrition to specifically refer to changes in the established native language (L1) of adult individuals who have experienced extended separation from their L1 environment.

Uniform methods for data collection and analyses are needed to reliably study the international and interdisciplinary field of language attrition (Köpke & Schmid, 2004; Schmid & Köpke, 2019). The need and opportunity to unify language attrition research methods was recognized in the early 2000's by researchers Monika Schmid and Barbara Köpke, whose efforts led to forming the European Graduate Network on Language Attrition (Schmid, 2011b; Schmid & Köpke, 2019). Their work has yielded guidelines and tools for collecting and analyzing data on language proficiency and extralinguistic background factors in the form of Language Attrition Test Battery (LATB) (Schmid, 2011b; see also Köpke & Schmid, 2004; Schmid & Köpke, 2009; Schmid & Köpke, 2019). The LATB is a collection of tests, instruments and analysis methods that can be applied to studies on language attrition in various languages. The protocol includes lexical tasks, grammatical tasks, and free speech samples. To systematically extract data for extralinguistic factors that may impact individual language performance, the LATB incorporates a comprehensive sociolinguistic questionnaire (Schmid, 2011b). The LATB framework has enabled attrition researchers to conduct their investigations consistently, leading to a relatively comparable database of studies on speakers of different languages in varied language environments (Jarvis, 2019; Schmid & Köpke, 2009). This research project is the first to apply LATB guidelines for data collection in a L1 Finnish-speaking population in an L2 English environment, thereby contributing to the evolving database.

The participants in this study are healthy, older adults who are late sequential bilinguals fully immersed in their second language. In language attrition research, the term *attriter* is can be used to refer to participants fitting this description (e.g., Dostert, 2009; Lazaridou-Chatzigoga & Karatsareas, 2022; Schmid, 2011a; Schmid

& Dusseldorp, 2010; Schmid & Jarvis, 2014; Schmid & Keijzer, 2009; Schmid & Köpke, 2009). Consequently, the term *attriter* is used here when discussing prior studies to align with established conventions. Additionally, in Studies II and III, I have adopted the term *attriter* to describe the participants in this dissertation to situate these studies within the language attrition framework. However, referring to participants as attriters without confirmed evidence of language attrition is controversial (e.g., Kasparian & Steinhauer, 2017; Schmid & Köpke, 2017b). Thus, in this work, I refer to the focus group of this dissertation as *immersed bilinguals* to reflect their linguistic background befitting typical participants in first language attrition studies.

This dissertation focuses on lexical retrieval. The LATB incorporates *Verbal Fluency (VF)* tasks as a controlled approach for measuring lexical fluency and accuracy (Jarvis, 2019; Schmid, 2011b, Schmid & Köpke, 2009). VF tasks are a versatile tool for examining rapid lexical retrieval. In a VF task the participant generates as many unique words as possible within a set time frame (e.g., 60 seconds), adhering to a given criterion (Strauss et al., 2006). The LATB protocol suggests VF task analysis at the total score level (Schmid, 2011b), following typical scoring protocol (Strauss et al., 2006). Many language attrition studies have reported significant differences between attriters and control groups' VF total scores with attriters underperforming in the semantic VF tasks (SVF). However, the underlying factor impeding optimal performance in the language attrition groups has not been systematically established (for an overview, see Schmid & Jarvis, 2014; Schmid and Köpke, 2009). In their comprehensive examination on lexical access and diversity in first language attrition, Schmid and Jarvis (2014) note that while total scores demonstrate robust group differences between the groups, they contribute very little to identifying participants as attriters. They suggest that the use of VF task as tools to detect lexical attrition may need to be reassessed, in line with Schmid and Köpke, (2009).

Language attrition is not the only field of research that has voiced concerns over the lack of explanatory power for VF task total scores (Thiele et al., 2016). An increasing number of studies across diverse populations are incorporating additional analyses to improve the explanatory value of VF tasks (Becker & Salles, 2016; Johns et al., 2018; Kim et al., 2019; Oberg & Ramírez, 2006; Thiele et al., 2016). These analyses include investigating errors and the temporal distribution of words within the task as well as tracking the use of clustering and switching strategies participants employ in their lexical retrieval. Applying these additional analyses to VF task performance in studies focusing on language attrition may facilitate the identification of underlying processes that hinder participants optimal performance in VF tasks.

1.2 Dissertation Structure

This dissertation consists of three articles. The first article (Study I) proposes and validates a systematic, in-depth analysis protocol for VF task data in a Finnish-speaking population in Finland. The second and third articles (Studies II and III) apply the proposed analysis method to VF data from a group of Finnish-English L2 immersed bilinguals with an extended residence in Northern California, US. These studies contrast L1 and L2 performance within the immersed bilingual group to identify strategies employed in lexical processing. The impact extralinguistic variables have on these strategies is also evaluated. L1 performance in the immersed bilingual group is contrasted to that of a group of Finnish speakers living in Finland to identify differences and similarities in lexical retrieval strategies between the groups to pinpoint group specific characteristics in lexical processing.

In the following chapters I first provide a concise overview of previous literature on language attrition and VF task analyses and outline the aims of this dissertation. The chapter on methods details the participants, data and data analysis. After providing an overview of the studies included in this dissertation I discuss the findings from these studies in relation to existing literature. In the discussion I aim to draw conclusions on the strategies that drive VF performance in a group of immersed bilinguals and the factors that potentially influence the use of these strategies. Before final conclusions, I discuss the limitations and major contributions of this dissertation and explore future prospects arising from the studies included in this dissertation.

2 Review of Literature

In this chapter I briefly describe existing literature on first language attrition with a focus on the lexical domain and extralinguistic variables relevant to this dissertation. I will then describe existing literature on verbal fluency (VF) tasks with a focus on the processes that support or potentially hinder optimal performance in these tasks.

2.1 Language Attrition

First language attrition is a complex phenomenon that can impact all components of language and communication (Schmid & Köpke, 2019; Schmitt & Sorokina, 2024). While it is also described as language loss, language attrition mainly consists of a series of subtle, yet consistent, changes, which are not always immediately observable in language performance. Language attrition primarily manifests during online processing (Jarvis, 2019; Schmid, 2013). Observations of diminished fluency, slower reaction times, speaker insecurities, reduced accuracy and diversity, overall simplification in L1 use, and incorporation of L2 elements into L1 can be regarded as language attrition (Schmid, 2011b; Schmid & Köpke, 2009).

Observations of language attrition markers are closely linked to cross-linguistic influence in online speech production (Gallo et al., 2021; Sharwood-Smith & van Buren, 1991; Schmid, 2013). L2 influence on L1 is experienced by all bilinguals as a part of the dynamic waxing and waning of languages within the bilingual individual (Grosjean, 2010, 2013; Schmid, 2011b; Schmid & Köpke, 2009). It is widely recognized that all languages coexisting in the bilingual brain interact continuously in a bidirectional process, altering the cognitive and neural dynamics of both languages and resulting in shifts in individual language dominance patterns over time (Grosjean, 1989; 2013; Gurunandan et al., 2022; Laine & Lehtonen, 2018; Linck & Kroll, 2019; Treffers-Daller, 2019).

Within the continuum of bilingualism, language attrition is characterized by its onset due to lack of exposure and reduced use of L1 alongside immersion in a second language environment (Gallo et al., 2021; Schmid, 2011b; Schmid & Köpke, 2009). This shared language history generates unique opportunities to explore L1 and L2 functional plasticity affecting lexical retrieval in both languages in populations under these circumstances. Many early first language attrition studies limit their explorations to comparisons between language attrition populations and L1 monolingual speakers residing in the L1 speaking country (Schmid, 2019; Schmid & Dusseldorp, 2010; Schmid & Yilmaz, 2018). However, evaluating the dynamics of L1 and L2 and the impact extralinguistic variables have on both languages is imperative for a holistic understanding of the processes that impact language task performance (Schmid & Köpke, 2019). Thus, examining performance in L2 alongside L1 can offer insight into the individual factors contributing to L1 language attrition or retainment as well as L2 acquisition (Bylund & Ramirez, 2016; Gurunandan et al., 2022; Runnqvist & Costa, 2011; Schmid & Köpke, 2017a; Schmid & Yilmaz, 2018).

In the field of language attrition, the most frequently researched area is the lexical domain. Markers of language attrition at the lexical level can be observed as a less readily accessible lexicon or reduced lexical retrieval (Köpke & Schmid, 2004; Schmid, 2011b; Schmid & Jarvis, 2014). A more thorough description of lexical attrition is provided in Section 2.1.1. Changes in structural systems, such as phonetics, phonology, morphology, and syntax have been documented, but these systems are generally considered rather stable and resistant to attrition once they are fully mastered in the native language (Schmid, 2011b). Structural changes can manifest as: borrowing (e.g. using grammatical sentence frames from L2), restructuring (e.g. changes in use of allophones), convergence (e.g. complex case systems merging several cases to one larger category), shift (e.g. preferring and overusing structures accepted in both L1 and L2) and possibly even losing underlying knowledge of the native language (Schmid, 2011b; Sharwood-Smith & van Buren, 1991).

2.1.1 Lexical Attrition

It is widely believed that the early signs of language attrition can be detected as lexical attrition (e.g. Opitz, 2011; Schmid 2011b; Schmid, 2013; Schmid & Köpke, 2019). Participants in language attrition studies also typically report problems with lexical access (Lazaridou-Chatzigoga & Karatsareas, 2022; Schmid, 2013). Measures to investigate lexical attrition include analysis of lexical accuracy, diversity and fluency (Andersen, 1982; Jarvis, 2019; Schmid & Köpke, 2009). To gain a comprehensive understanding of lexical access and abilities the Language Attrition Test Battery (LATB; Schmid, 2011b) recommends eliciting free speech samples and implementing controlled tasks during data collection (Schmid & Jarvis, 2014). For free speech samples, LATB suggests eliciting data via a film retelling task to ensure task comparability across studies (specifically a 9 min 58s sequence from the Charlie Chaplin movie 'Modern Times' following Perdue, 1993). For controlled tasks the LATB suggests including picture naming tasks, picture-word matching tasks and verbal fluency tasks.

Lexical accuracy analysis from free speech samples has demonstrated effortful word-finding, observed as hesitations, false starts, filled or empty pauses, self-corrections, and circumlocutions (Schmid & Köpke, 2009), as well as errors with lexico-semantic errors being the most prevalent (Jarvis, 2019). Controlled lexical tasks (picture naming tasks and picture-word matching tasks) have proven effective in detecting a loss of accuracy among attriters (for an overview, see Schmid & Köpke, 2009). Hesitations, self-corrections and errors in VF tasks can also be analyzed as diminished accuracy markers.

Lexical diversity is typically assessed by analyzing free speech samples. A common method of analysis is calculating the Type-Token Ratio (TTR, the total number of different lemmas in relation to the total number of words produced in a speech sample) (Köpke & Schmid, 2004). Although this approach has limitations and can yield inconsistent findings (Schmid & Jarvis, 2014) it remains a popular method for language attrition studies (Schmid & Köpke, 2009). Lexical diversity can also be evaluated by assessing the ability to generate high- and low-frequency words in VF tasks (e.g., "bug", "earwig") (Sandoval et al., 2010)

Lexical fluency, characterized by slowed down retrieval or failure, can be measured in free speech samples by speech rate or by the total number of words generated. Factors such as hesitations, filled or empty pauses, self-corrections, and circumlocutions can reduce lexical fluency (Jarvis, 2019). Lexical tasks, namely, picture naming tasks and picture-word matching tasks have demonstrated increased reaction times in language attriters compared to monolinguals (Schmid, 2011b; Schmid & Köpke, 2009).

For a lexical fluency measure under controlled conditions, the LATB proposes VF task total scores (Jarvis, 2019; Schmid, 2011b; Schmid & Köpke, 2009). There is ample data on semantic verbal fluency (SVF) task total scores with language attrition populations systematically generating lower scores than monolinguals (e.g., Badstübner, 2011; Dostert, 2009; Schmid, 2011a; Schmid & Dusseldorp, 2010; Schmid & Keijzer, 2009; Schmid & Köpke, 2009; Schmid & Jarvis, 2014; Opitz, 2011), but the application of phonemic verbal fluency (PVF) tasks remains limited (Jarvis, 2019). VF tasks, their analysis, and findings in general bilingual populations and specifically in language attrition studies are described in detail in Section 2.2.

2.1.2 Extralinguistic Variables

Sociolinguistic parameters can impact language performance and shifts in language dominance patterns (Grosjean, 2013; Gurunandan et al., 2022; Laine & Lehtonen, 2018; Linck & Kroll, 2019; Treffers-Daller, 2019; Yilmaz, 2019). Language attrition populations typically exhibit considerable variation in their language performance with some attriters exhibiting more attrition than others (Schmid & Cherciov, 2019; Schmitt & Sorokina, 2024). To investigate the factors that impact language performance within language attrition populations the LATB includes a comprehensive sociolinguistic questionnaire for documenting extralinguistic parameters in a uniform manner. The questionnaire is introduced and discussed in detail in Schmid (2011b), Schmid and Cherciov (2019), and Schmid and Dusseldorp (2010), and it is available online[1] in English with translations to several languages, along with tools for analyzing the data. To examine the impact of the shared language history on task performance, the present study includes two variables from the questionnaire: Length of Residence in the L2 language environment (LoR) and the frequency of current L1 use.

LoR represents the duration of reduced native L1 exposure and extensive exposure to L2. It has been demonstrated that LoR can predict task performance when the time spent in L2 environment is less than ten years, with the effect stabilizing after this period (Opitz, 2011; Schmid, 2011a; Schmid, 2019). In her overview, Schmid (2019) summarized that of the 49 language attrition studies that included LoR as a predictor, none of the studies with a LoR of ten or more years showed a significant effect. This stabilization effect is partially attributed to the initial years of immersion aligning with a period of rapid L2 learning in an immersive environment (Linck & Kroll, 2019). However, while increased exposure to L2 can trigger L1 attrition, attrition measures do not directly correlate with L2 acquisition or proficiency levels (Sorokina & Mugno, 2024; Yilmaz, 2019).

Exploring the impact of L1 use on language performance is more complex than measuring LoR. Cumulative data gathered via the sociolinguistic questionnaire suggests that the frequency of L1 use as a single measurement does not systematically affect L1 performance (Schmid, 2013; for an overview see Schmid, 2019). However, it has been shown that the use of L1 in different life domains can have varying implications for changes in L1 performance. Guidelines for data analysis for the sociolinguistic questionnaire include classifying L1 use into three categories (Schmid & Dusseldorp, 2010):

- Use of the L1 in informal situations where code-switching between languages is not inhibited (with partner, children, and bilingual friends)
- Use of the L1 in situations where L2 is highly active, but suppressed and switching between languages is inappropriate (e.g., work related contexts or groups that socially do not accept switching)
- Passive exposure to non-attrited L1 (media, reading, visits to L1-speaking country)

Studies following these guidelines have shown a positive effect for frequent use of L1 in a professional setting on low-frequency word retrieval (Yilmaz & Schmid, 2012), general proficiency (as measured by a c-test), grammaticality judgment task performance, lexical diversity (Schmid & Dusseldorp, 2010) and verbal fluency (Schmid, 2011a; Schmid & Dusseldorp, 2010; Schmid & Jarvis, 2014). Conversely, frequent L1 use in informal interactions with peers has been linked with increased variability in the phonemic domain, resembling developing a foreign accent (De Leeuw et al., 2010). As a result, it has been suggested that attriters who have spent an extended time in an L2 environment and frequently use L1 with peers reflect a contact-induced language change in the community at a larger scale (Schmid, 2011a; Schmid & Köpke, 2017a). Research also indicates that although LoR and frequent L1 use do not independently predict attrition (Schmid, 2011a; Schmid, 2019; Schmid & Dusseldorp, 2010; Schmid & Yilmaz, 2018), the influence of LoR can become significant when coupled with minimal L1 use (de Bot & Clyne, 1994; Schmid, 2011a).

The impact of the extralinguistic variables that define the shared background of language attriters continues to be a relevant question, despite inconsistent findings (Schmid, 2019). Participants in language attrition report experiencing attrition as natural extension of limited use of L1 over several years in L2 dominant environment (e.g., Lazaridou-Chatzigoga & Karatsareas, 2022). Cumulative data from larger studies and multiple smaller studies utilizing established methods could provide insights into how these shared extralinguistic characteristics influence language attrition, thereby contributing to a more comprehensive understanding of the phenomenon.

2.2 Verbal Fluency Tasks

Verbal fluency (VF) tasks are among the most commonly utilized tools for studying lexical retrieval in neuropsychological language assessments, serving various purposes and subject groups. They are employed in multiple research fields, including speech pathology, neuropsychology, linguistics, and medicine, across different languages and cultures (Strauss et al., 2006).

VF tasks are seemingly simple to administer and analyze. Essentially, the participant is instructed to generate as many unique words as possible, adhering to a given criterion within a set time frame (e.g., 60 seconds), and total score (the number of words) is calculated to represent task performance. Common VF task types are phonemic (PVF) and semantic (SVF). The PVF task requires participants to generate words beginning with a specific phoneme or letter. The SVF task involves participants producing words that belong to a specific semantic category, such as “animals” or “clothes” (Strauss et al., 2006).

All VF tasks necessitate sustained focus on task and require inhibition as participants search and select words that meet the given criteria while inhibiting unsuitable candidates (Shao et al., 2014; Whiteside et al., 2016). Yet, PVF and SVF tasks probe lexical retrieval with differing cognitive and executive function demands. PVF is typically thought to probe strategic cognitive organization, initiation, inhibition, and maintenance of effort, without the support of the hierarchical organization of semantic memory (Barry et al., 2008; Santos Nogueira et al., 2016; Strauss et al., 2006). SVF is believed to rely on a more automatic systematic semantic lexical search based on hierarchical mental lexicon, and memory organization, reflecting everyday language use (e.g., generating a shopping list) (Patra et al., 2020; Strauss et al., 2006).

The different emphases of VF tasks tend to be more pronounced in bilinguals compared to monolinguals (Friesen et al., 2015; Patra et al., 2020; Rosselli et al., 2000; Schmid & Köpke, 2009). Examining task performance in both task types in older bilingual and monolingual populations can provide insights into the mechanisms behind lexical retrieval. Similarities in PVF and SVF task performance between these groups may indicate a general susceptibility to aging in the retrieval process. Differences between PVF and SVF task performance can reveal group-specific subcomponents in lexical retrieval strategies (Goral, 2004). Thus, including both task types, PVF and SVF, in studies on bilingual performance could guide identification of differing language processing strategies between bilingual populations.

As stated above, performance in VF tasks is mostly assessed by the total score (Strauss et al., 2006; Thiele et al., 2016). Total scores are easily measurable, providing general information on language task performance, but they lack explanatory power of the processes that impact the outcome. To probe the processes that support or hinder optimal VF performance, additional analyses of errors, temporal parameters, and clustering and switching strategies have been implemented, utilizing varied approaches and analysis methods (e.g., Becker & Salles, 2016; Johns et al., 2018; Kim et al., 2019; Oberg & Ramírez, 2006; Thiele et al., 2016). In the following, I briefly discuss findings from previous studies implementing these analysis methods as relevant to this dissertation.

2.2.1 Total Scores

Norms for total scores, including the effects of age, education, and gender for monolingual populations have been published in various languages (Ardila, 2020; Cavaco et al., 2013; Goral, 2004; Oberg & Ramírez, 2006; Olabarrieta-Landa et al., 2017; Pereira et al., 2018; Quaranta et al., 2016; Santos Nogueira et al., 2016; Vicente et al., 2021). Typically, normative data describe higher total scores for SVF tasks than PVF tasks (Cavaco et al., 2013; Santos Nogueira et al., 2016; Strauss et al., 2006). Higher total scores are generally associated with higher education, particularly in PVF tasks (Oberg & Ramírez, 2006; Santos Nogueira et al., 2016; Tallberg et al., 2008; Tombaugh et al., 1999; Troyer, 2000) and total scores tend to decrease with age, especially in SVF tasks (Ardila, 2020; Goral, 2004; Lanting et al., 2009; Santos Nogueira et al., 2016; Strauss et al., 2006; Tallberg et al., 2008; Tombaugh et al., 1999; Troyer, 2000). Regarding gender, studies have shown either no effect or a slight female advantage with an emphasis on PVF (Scheuringer et al., 2017). However, comparing studies and their outcomes is challenging as many studies fail to provide accurate descriptions of the task administration, scoring, and analysis methods applied (Olabarrieta-Landa et al., 2017; Thiele et al., 2016).

In the field of bilingualism studies, both PVF and SVF tasks are utilized to investigate lexical access (Rosselli et al., 2002), vocabulary knowledge (Luo et al., 2010; Rosselli et al., 2002), dominance patterns of languages (Gollan et al., 2002; Roberts & Le Dorze, 1997), the role of executive functions (Luo et al., 2010; Marsh et al., 2019; Patra et al., 2020), and cross-linguistic fluency strategies (Roberts & Le Dorze, 1997; Rosselli et al., 2002). Research on general bilingual populations shows that bilinguals produce fewer words than monolinguals in SVF tasks (Gollan et al., 2002; Ljungberg et al., 2013; Patra et al., 2020; Rosselli et al., 2000; Sandoval et al., 2010). In PVF tasks, performance varies, with some studies showing similar performance between groups (Rosselli et al., 2000; Rosselli et al., 2002), and others showing bilinguals outperforming monolinguals (Luo et al., 2010; Marsh et al., 2019; Patra et al., 2020; Sandoval et al., 2010). It has also been shown, that in general bilingual populations, VF task total scores tend to decline with age, especially in SVF tasks (Ardila, 2020; Goral, 2004; Lanting et al., 2009; Santos Nogueira et al., 2016; Strauss et al., 2006; Tallberg et al., 2008; Tombaugh et al., 1999; Troyer, 2000).

Consistent with general bilingual studies, numerous language attrition studies have identified significant differences in SVF total scores between attriters and controls, with attriters underperforming compared to monolinguals (e.g., Badstübner, 2011; Dostert, 2009; Lazaridou-Chatzigoga & Karatsareas, 2022; Schmid, 2011a; Schmid & Dusseldorp, 2010; Schmid & Jarvis, 2014; Schmid & Keijzer, 2009; Schmid & Köpke, 2009; Opitz, 2011). While PVF tasks are less frequently used in the field of language attrition (Jarvis, 2019; Schmid & Köpke, 2009), Badstübner (2011) found no difference between groups in PVF performance, suggesting that PVF performance remains unaffected by L1 attrition, while Lazaridou-Chatzigoga and Karatsareas (2022) showed that attriters were outperformed by monolinguals also in PVF.

Efforts to identify the underlying factors of consistent group differences in SVF have considered the frequency of L1 exposure, attitudinal component (i.e., attitude towards a specific language) and education level, but no systematic findings have been reported (for an overview, see Schmid, 2019). In addition, SVF total scores do not systematically correlate with measures of other language domains (e.g., picture naming, or lexical diversity) (Schmid, 2011b; Schmid & Köpke, 2009), and they have been shown to demonstrate limited predictive power in profiling individual speakers as attriters or monolinguals (Schmid & Jarvis, 2014). Thus, it has been proposed that the effectiveness of VF total scores as a tool for detecting lexical attrition may require re-evaluation (Schmid & Jarvis, 2014; Schmid & Köpke, 2009).

2.2.2 Errors

While scarce in normative data, errors are a part of VF task performance across all populations (Crowe, 1998; Gollan et al., 2011; Sandoval et al., 2010). Typical error types in VF tasks include repetitions, items out of category, and non-items (Thiele et al., 2016).

In bilingual environments, cross-language intrusion errors are of particular interest, as these can be linked to language interference (Gollan et al., 2002; Gollan et al., 2011). The occurrence of language intrusion errors can also signal language dominance, with more frequent errors generated in the non-dominant language (Sandoval et al., 2010).

Regarding language attrition populations, Badstübner (2011) found that attriters generated more errors than monolinguals in SVF tasks. These errors were described as direct L2 intrusions and incorrect lexical items (such as partial recall of an L1 word), potentially indicating L2 dominance within the subject group.

2.2.3 Temporal Parameters

Temporal analysis of VF tasks is conducted to detect differences in the speed and distribution in which the words are generated within the task. Temporal parameters of VF task performance can be analyzed by counting the number of words generated during shorter time segments within the total time (e.g., 10, 15, or 20-second segments within the 60-second total time). Participants typically produce most words in the early stages of the task, employing a semi-automatic rapid retrieval process (Crowe, 1998; Fernaeus & Almkvist, 1998; Fernaeus et al., 2008; Venegas & Mansur, 2011). In the later segments of the tasks lexical retrieval becomes more effortful, with fewer and more infrequent words being generated.

Studies have demonstrated that bilinguals generate fewer words than monolinguals during the initial stages of a VF task, but the difference between the groups tends to diminish as the task progresses (Luo et al., 2010; Sandoval et al., 2010). This pattern has also been demonstrated for attriters in SVF tasks. Schmid and Jarvis (2014) calculated the number of words in six ten-second segments within the 60-second timeframe in two groups of attriters in comparison to a monolingual group. Their analyses showed that attriters generated fewer words than monolinguals in all segments, with the most pronounced difference in the second and third segment.

The pattern of slowing down in the early segments is often interpreted to reflect more language interference in the early stages of the task, where bilinguals may produce high-frequency words they know in both languages (e.g., “cat”) rather than low-frequency words (e.g., “bobcat”) they may only know in one language (Gollan et al., 2011; Luo et al., 2010; Sandoval et al., 2010; Schmid & Jarvis, 2014; Yilmaz & Schmid, 2018). However, highly automatic language skills, which are relied upon during the early stages of task performance, are considered to be more resistant to language attrition (Goral, 2004; Segalowitz, 1991). It is also worth noting that similar neural activation patterns have been observed during temporal analysis for L1 and L2 performance, suggesting a common bilingual effect in both languages (Gurunandan et al., 2022).

2.2.4 Clustering and Switching

Clustering and switching strategies in VF tasks can be analyzed for various research objectives, such as differential diagnostics in neuropsychological populations (Johns et al., 2018; Thiele et al., 2016; Troyer et al., 1997; Troyer, 2000) and cross-linguistic fluency strategies in bilingualism studies (Roberts & Dorze, 1997; Rosselli et al., 2002). Clustering refers to the ability to form sub-clusters within a category (e.g., within the category of animals: farm animals, pets, wild animals), while switching refers to the ability to transition between these clusters (e.g., shifting from farm animals to pets, then to wild animals and then back to farm animals). Successful performance in a VF task involves both strategies, typically measured as the mean cluster size and the number of switches (Strauss et al., 2006; Thiele et al., 2016; Troyer et al., 1997; Troyer, 2000).

In PVF, task-congruent clusters, namely phonemic clusters, are based on phonemic characteristics such as identical onset-nucleus sequences, as in 'simple, simile, sieve'. The generation of these clusters necessitates a detailed non-routine search reliant on phonemic attributes, without the assistance of semantic categorization (Luo et al., 2010; Strauss et al., 2006; Troyer et al., 1997; Troyer, 2000). Generating semantic clusters in SVF engages verbal semantic memory and semantic categorization within a category. The transition between clusters, or switching, involves higher executive functions, including cognitive flexibility (Thiele et al., 2016; Troyer, 2000; Troyer et al., 1997). Troyer, Moscovitch, and Winocur (1997) demonstrated that, within a monolingual population, switching had a more significant impact on achieving optimal fluency than clustering in PVF and in SVF, both clustering and switching strategies held equal importance in achieving a high total score.

Alongside task-congruent clusters, task-discrepant clustering (semantic clusters in PVF and vice versa) represents an additional strategic approach that requires more effort and intentionality. Compared to task-congruent clustering this process involves higher cognitive functions and task-discrepant clusters in PVF may indicate the activation of semantic mechanisms during a demanding phonemic task (Abwender et al., 2001; Sung et al., 2013).

Within the field of bilingualism, clustering and switching strategies have been analyzed to examine executive processes (Mardani et al., 2020; Marsh et al., 2019; Patra et al., 2020) and cross-linguistic productivity strategies (Roberts & Le Dorze, 1997; Rosselli et al., 2002). It has been hypothesized that bilinguals who routinely alternate between languages in daily life may exhibit superior cognitive flexibility through more efficient switching strategies compared to monolinguals (Gollan et al., 2002). Mardani et al. (2020) reported that bilinguals, with no abrupt change in their language environment, relied more on switching to achieve a higher total score than monolinguals in L1 PVF. In SVF, bilinguals generated smaller clusters than monolinguals, interpreted as a disadvantage in L1 verbal-semantic memory performance. Conversely, Patra et al. (2020) demonstrated that bilingual emigrants (with an average of 7.48 years spent in an L2 environment) maintained PVF clustering more efficiently than monolinguals in L2. This was interpreted as superior executive performance in the bilingual group.

Regarding studies on language attrition populations, Ammerlaan (1996) noted that attriters heavily relied on effortful strategies partly resembling clustering (looking for clues in their environment or translating from L2 in SVF and going through the alphabet in PVF), although his study did not include an analysis of these strategies. Based on the literature reviewed for this dissertation, no studies that focus on language attrition per se have employed a robust VF task clustering and switching analysis.

It is crucial to acknowledge that direct comparisons across studies using different data analysis methods for clustering and switching should be approached with caution. Diverse methodologies for determining clusters and calculating switches can yield varied results (Abwender et al., 2001; Ross, 2003; Strauss et al., 2006; Thiele et al., 2016; Troyer, 2000; Troyer et al., 1997). A comprehensive description of analyzing clusters and switches in VF data for this study can be found in Study I, Appendix A.

In addition, cross-linguistic studies have underscored the importance of considering diverse language backgrounds and language exposure when selecting VF categories and determining clusters in bilingual populations (Abwender et al., 2001; Gollan et al., 2011; Olabarrieta-Landa et al., 2017; Roberts & Le Dorze, 1997; Rosselli et al., 2002). Factors such as differing semantic category structures between bilingual and monolingual groups (Rosselli et al., 2002), and the impact of life experiences in a specific language on the semantic structure of category vocabulary (Roberts & Le Dorze, 1997), should be accounted for when evaluating clusters in bilingual performance. These factors can be considered in data analysis by calculating naturally occurring clusters rather than relying on predetermined categories, particularly in SVF tasks (Abwender et al., 2001; Gollan et al., 2011; Olabarrieta-Landa et al., 2017; Roberts & Le Dorze, 1997; Rosselli et al., 2002). Calculating naturally occurring clusters allows for all unique, individual clustering and natural switching strategies to be accounted for without cultural or language related premises. For more comprehensive information on naturally occurring clusters, the reader is directed to Study I, Appendix A.

3 Aims of the Study

The primary objective of this dissertation is to investigate lexical task performance in a group of Finnish-English bilinguals experiencing reduced exposure and use of their first language in an immersive second language environment. The specific aim of this study is to identify group-specific processing strategies that drive optimal task performance in semantic and phonemic verbal fluency (VF) tasks.

First, this dissertation aims to establish systematic guidelines and provide tools for comprehensive analyses of VF tasks to explore strategies that underlie optimal VF performance (Study I). By developing a systematic method for analysis that extends beyond total scores, this dissertation aims to demonstrate that commonly used but often under-analyzed VF tasks can serve as a valuable tool for detailed investigations of lexical retrieval.

The second aim is to utilize the proposed method to identify similarities and differences in L1 and L2 lexical processing strategies within the group of Finnish-English immersed bilinguals and evaluate the impact extralinguistic variables have on these strategies (Studies II and III).

The third aim is to investigate the characteristics of language processing strategies among immersed bilinguals by comparing the strategies they use in their L1 to those used by a group of Finnish speakers residing in Finland (Studies II and III)

In addition to the abovementioned aims of this dissertation, this PhD study aims to enable future studies on language spoken by Finnish emigrants in Northern California, US. During data collection sessions free speech samples and formal tasks were recorded in addition to the VF tasks analyzed for this dissertation. This data is described in Appendix 1, and it is stored and made available for future research through the Language Bank of Finland[1].

[1] www.kielipankki.fi

4 Methods

This chapter introduces the participants, data, and data analysis of this study and discusses ethical considerations. First, the two groups of participants and the background factors related to their language history as relevant to this dissertation are described. Methods for data collection, data content and analyses are then outlined.

4.1 Participants

Two groups of healthy, neurotypical adults participated in this study: L1 Finnish speakers immersed in an L2 English language environment referred to here as *the immersed bilingual group* (N = 38) and Finnish speakers living in Finland referred to as *the control group* (N = 50).

Participants for the immersed bilingual group were recruited first, followed by Finnish speakers residing in Finland. Groups were matched for age, educational level (lower level, without an academic degree; higher level, with an academic degree), and gender to minimize the potential effect of demographic variables. Exclusion criteria for both groups included a history or presence of cardiovascular, neurological, psychiatric, and developmental language or speech disorders, severe hearing loss, substance abuse, and age above 80 years.

The Ethics Committee of the University of Turku has approved all experimental procedures. Participants were informed about their rights, including the option to withdraw from the study at any time and the option to decline having their collected data deposited in the Finnish Language Bank. No monetary compensation was provided for their participation.

4.1.1 Immersed Bilingual Group

The immersed bilingual group (N = 38) comprised first-generation Finnish emigrants residing in Northern California. Participants were recruited through local Finnish groups and associations in Northern California and via introductions from existing participants based on the following criteria: 1) first language is Finnish, second language is English, 2) first-generation emigrant, moved from Finland to the

US after 12 years of age, 3) age range 30-80, 4) lived in an English-language environment for a minimum of 20 years and used English as their main academic and/or work language during this time. Based on these recruitment criteria, 41 individuals were interviewed.

All potential participants spoke Finnish as their first language (L1) and English as their second language (L2). They had spoken Finnish as their primary language before emigration. Two of the recruits had emigrated to the US before the age of 12 (one participant at the age of 9 and one at the age of 11). Both participants spoke fluent Finnish and were included despite their younger age at the time of emigration. Three of the recruits were over the age of 80 and were excluded. Included participants had lived in an L2 environment for at least 20 years (Length of Residence in years, LoR, M = 34.24, SD = 10.83, range 20 – 50) and had used English as their academic or work language. The group's average age was 60.90 (SD = 8.42, range 45–79), with education levels divided into two categories: no academic degree (n = 16) and academic degree (n = 22). The group consisted of 29 female participants and 9 male participants.

All participants had emigrated either as adults or as older children with their families between the years 1948 and 1998 by their own choice (age at emigration M = 26.68, SD = 7.38, range 9 – 48). Those who arrived with their parents opted to remain in the United States as adults. While not employed as inclusion criteria or for participant profiling in this dissertation, the following questions pertaining to language background and perceived language change from the sociolinguistic questionnaire are included here for informational purposes:

- Do you see yourself as bilingual? “No” n = 0, “Yes” n = 38
- In general, how would you rate your Finnish language proficiency before you moved to the USA? “None” n = 0, “Bad” n = 0, “Good enough to get by” n = 0, “Good” n = 15, “Very good” n = 23
- In general, how would you rate your Finnish language proficiency at present? “None” n = 0, “Bad” n = 0, “Good enough to get by” n = *6*, “Good” n = 26, “Very good” n = 6
- Has the balance between your Finnish and English changed during the time you have lived in the USA? “No” n = 5, “Yes” n = 32
- Do you think your Finnish language proficiency has changed since you moved to the USA? “Yes, I think it has become worse” n = 32,” No” n = 5, “Yes, I think it has become better” n = 1
- In general, how would you rate your English language proficiency at present? “None” n = 0, “Bad” n = 0, “Good enough to get by” n = *1*, “Good” n = 16, “Very good” n = 21

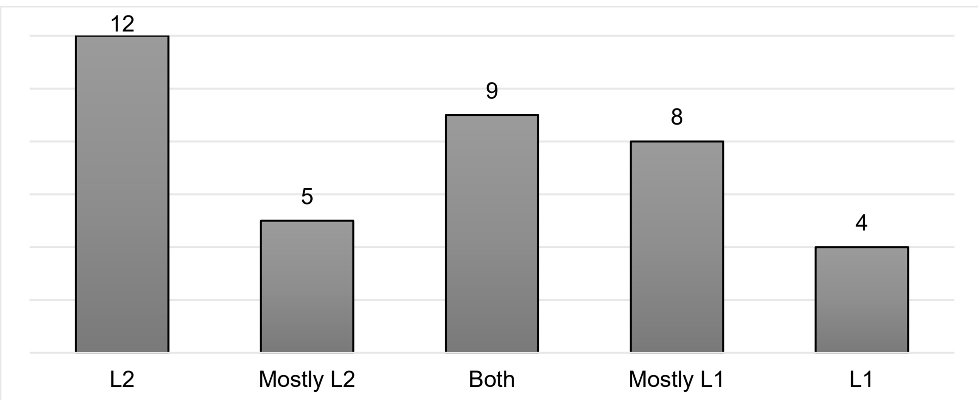
- Do you feel that you are equally proficient in Finnish and English? “No, more proficient in English” *n* = 10, “Yes” *n* = 18, “No, more proficient in Finnish” *n* = 9
- Do you feel more comfortable speaking Finnish or English? “English” *n* = 10, “Finnish” *n* = 2, “No preference” *n* = 26

As shown above, at the time of the interview all participants identified themselves as bilinguals. The majority of participants reported their L1 proficiency as very good before emigration and good at the time of the interview. The majority also reported that their language balance had shifted while living in the US. In terms of proficiency, most participants reported that they had experienced a decline in their L1 proficiency. As for L2, the majority of participants reported their L2 proficiency at the time of the interview as very good. The majority of participants perceived their proficiency in L1 and L2 as equal, and most reported feeling equally comfortable speaking both languages, with a larger number indicating that they felt more comfortable speaking L2 than L1.

For language dominance, the participants were asked to evaluate the language they are “best” at on a 5-point Likert scale with answers visualized in Figure 1. This distribution suggests that, on average, the perceived dominance in the immersed bilingual group was balanced bilingualism with a slight preference for English (L2) [*M* = 2.66 (*SD* = 1.40, range 1 – 5)].

- Language dominance (i.e., the language you are “best” at)? “English” *n* = 12, “Mostly English” *n* = 5, “Both Finnish and English” *n* = 9, “Mostly Finnish” *n* = 8, “Finnish” *n* = 4

Figure 1. Self-Reported Language Dominance in the Immersed Bilingual Group.



Note. *N* = 38. Question: Language dominance (i.e., the language you are “best” at)?

4.1.2 Control Group

The control group (*N* = 50) consisted of first language Finnish speakers who had lived exclusively in Finland and used Finnish as their primary language in daily life. The recruitment of participants for the control group aimed to closely match the immersed bilingual group in terms of age, gender, and education. Recruiting participants for the control group proved to be more straightforward than recruiting immersed bilinguals and all interested applicants that fit the requirements were included. Having a larger control group than immersed bilinguals was considered to strengthen the analyses as statistical methods applied are not dependent on equal group size and larger group size improved the reliability of the analyses. The average age in the control group was 62.58 (*SD* = 7.59, range 49 – 79). Education background was determined in two tiers: without an academic degree (*n* = 27), with an academic degree (*n* = 23). The group consisted of 35 female participants and 15 male participants.

The vast majority of participants in the control group identified as monolingual at the time of the interview. However, all participants had been exposed to at least two other languages in a school setting due to Finland's language policy. For informational purposes, the following list shows answers to questions pertaining to perceived monolingualism and language background:

- Are you bilingual? “No” *n* = 49, “Yes” *n* = 1 (Finnish-Swedish)
- What languages did you learn as a child (before school)? “Finnish” *n* = 49, “Finnish and other” *n* = 1 (Swedish), “Other” *n* = 0
- What language do you use in your everyday life? “Finnish” *n* = 47, “Finnish and other language, but mostly Finnish” *n = 3*, “Two languages, equally” *n = 0,* “Finnish and other language, but mostly other” *n = 0,* “Other language than Finnish” *n =0*
- What has been the primary working language in your work career? “Finnish” *n* = 37, “Finnish and other language, but mostly Finnish” *n = 13*, “Equally with two languages” *n = 0,* “Finnish and other language, but mostly other” *n = 0,* “Other language than Finnish” *n =0*

4.2 Data

In this section, I detail the data analyzed for this dissertation, namely two extralinguistic variables derived from the sociolinguistic questionnaire and the verbal fluency (VF) tasks. Beyond the data analyzed for this dissertation, additional tasks were administered during this research project as described Appendix 1.

Sociolinguistic Questionnaire (Immersed Bilinguals: L1 or L2; Controls: L1)

Background information for the immersed bilingual group was collected using the sociolinguistic questionnaire included in the Language Attrition Test Battery. The questionnaire comprised 71 questions, covering numerous self-reports on different aspects of language history, use, attitudes, and affiliations. The author translated the questionnaire into Finnish, and immersed bilinguals had the choice to complete it in either Finnish or English. For the control group, an abridged version of the questionnaire was used to confirm the use of Finnish in everyday contexts. All questionnaires are included as supplemental material in Study II.

Verbal Fluency Task (Immersed Bilinguals: L1 and L2; Controls: L1)

The VF task was administered in two subtypes: phonemic (PVF) and semantic (SVF). For the L1 PVF task, participants were given 60 seconds to name words beginning with the most frequent word-initial letters in Finnish (K, A, P) following the high-frequency dictionary approach (Mardani et al., 2020; Oberg & Ramírez, 2006; Schmid, 2011a). For L2 VF tasks participants were asked to name words beginning with three letters commonly used for PVF tasks in English (F, A, S) (based on Borkowski et al., 1967, see Strauss et al., 2006). For the SVF task, participants were given 60 seconds to name items within a semantic category, specifically concrete category "animals" and abstract category "emotions" in L1 and L2. To limit the complexity of the semantic analysis, only the concrete, linguistically neutral, category "animals" (Pekkala et al, 2009) was included in this dissertation. Detailed description for task administration is given in "The Instruction Manual for Administration and Scoring Verbal Fluency Tasks" (Study I, Appendix A).

Data Collection Sessions

Data for each participant in the immersed bilingual group were collected during a single session in Northern California, US. Participants were instructed to use only one language at a time, despite being aware that the interlocutor was bilingual. Data for the control group were collected in Finland in a single session. At the start of each session, participants were given information about the study in their preferred language and asked to complete a consent form, including permission to deposit the data into the Language Bank of Finland. Participants were given a choice to complete the sociolinguistic questionnaire before or during the session. The questionnaire was used to guide a semi-structured interview to build rapport with the participants. Conversations that occurred during the interview were not recorded.

A description of additional experimental tasks administered during the sessions is provided in Appendix 1. For the immersed bilingual group, the tasks were conducted first in one language, followed by a brief pause before being administered in the second language. The order of languages (L1 vs. L2) and the order of different experimental language tasks within the language were pseudo-randomized. For the control group the order of experimental tasks in L1 was pseudo-randomized. All sessions were conducted in a quiet environment, such as the participants' homes or a clinical setting.

4.3 Data Analysis

This section offers a concise summary of the data analysis methods used in the three studies constituting this dissertation. Statistical analyses and their outcomes are described in Chapter 5.

4.3.1 Extralinguistic Variables

The participant groups were matched based on age, education, and gender. These variables were extracted from the sociolinguistic questionnaire and analyzed to determine whether there were any significant differences between the groups. No significant differences were detected between the groups for age (immersed bilinguals $M = 60.90$, controls $M = 62.60$, $z = -0.97$, $p = .332$), education ($\chi(1) = 0.49$, $p = .27$), or gender ($\chi(1) = 0.43$, $p = .51$).

The LATB sociolinguistic questionnaire for immersed bilinguals incorporates guidelines for data analysis, including the formulation of compound variables for L1 use across various domains. These guidelines for data analysis are available online[4]. Preliminary investigations showed that it was not possible to reliably consolidate participant data to formulate suggested compound variables for this dataset. Previous literature suggests examining L1 use in different domains, specifically for professional purposes (Schmid & Dusseldorp, 2010; Schmid, 2011a; Schmid & Köpke, 2017; Yilmaz & Schmid, 2012). However, as only three participants reported regular L1 use for work, this could not be used as a separate variable reliably for this dissertation.

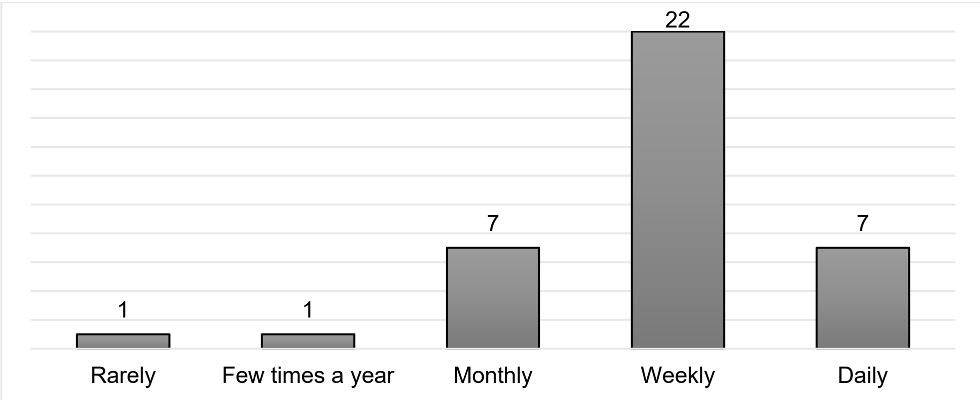
- What has been the primary working language in your work career? “Only English” $n = 25$, “Both, but mostly English” $n = 10$, “Both, equally” $n = 2$, “both, but mostly Finnish” $n = 1$, “only Finnish” $n = 0$

Following these preliminary analysis, two variables were extracted from the sociolinguistic questionnaire to evaluate the impact of shared language history on VF task performance in Studies II and III. These variables were: length of residence in the L2 environment (LoR) and frequency of L1 use. LoR was measured in years from emigration to the interview date (M = 34.24 years, SD = 10.83, range 20 – 70).

Frequency of L1 use was estimated on a 5-point Likert scale, with 57.9% of participants reporting weekly L1 use. The distribution of responses regarding frequency of L1 use is illustrated in Figure 2.

- How often do you speak Finnish? “Rarely” n = 1, “Few times a year” n = 1, “Monthly” n = 7, “Weekly” n = 22, “Daily” n = 7

Figure 2. Frequency of L1 Use on a 5-point Likert Scale within the Immersed Bilingual Group.



Note. *N* = 38. Question: How often do you speak Finnish?

4.3.2 Verbal Fluency Tasks

During data elicitation, all responses in the PVF and SVF tasks were recorded for subsequent verification and scoring. A research assistant transcribed the audio tracks, which were then verified by the first author of Study I. In case of disagreement, the second author of Study I was asked to check the transcription. If no agreement was reached, the entry would have been scored as unintelligible. There were no such instances in the data.

All VF tasks were scored based on: (1) total score, (2) number of acceptable words generated in 15-second segments within 60 seconds, (3) quantity and types of errors, (4) mean cluster size for task-congruent clusters (semantic for semantic and phonemic for phonemic), and the number of switches calculated from task-congruent clusters, and (5) the number of task-discrepant clusters (semantic clusters in phonemic task and vice versa). The process for administering, scoring and analyzing VF task data is outlined in detail in Appendix A of Study I.

Total Scores

The total score represents the sum of acceptable words generated during a 60-second trial. All unique words exhibiting semantically distinctive features were scored as separate items. Synonyms were considered individual items, reflecting the versatility of vocabulary. Inflections of previously produced items were scored as errors (repetitions). Homonyms were scored as separate items when the participant indicated a semantic distinction (e.g., the Finnish word “kuusi” meaning “six” or “spruce”). For the semantic category “animals”, insects were accepted, while fantasy and imaginary animals were not accepted and were scored as categorical errors (items out of category).

Errors

Errors were excluded from the total score and classified into five categories: 1. repetition, 2. categorical error (item out of category, such as “plant” under the semantic trial “animals” or a word with the wrong initial phoneme under phonemic trials), 3. rule-break error (such as proper names), 4. paraphasia and nonword, and 5. language intrusion (word in a different language than the target language, excluding established loan words).

Temporal Parameters

To investigate temporal parameters of VF task performance the total time of 60 seconds was divided into four 15-second segments: 1–15 s, 16–30 s, 31–45 s, and 46–60 s. The score for each 15-second time segment was calculated following the protocol for total scores and errors. Any word was attributed to the time segment in which the participant began producing it.

Clustering and Switching

Clusters were defined as two or more consecutively generated words that belong to the same semantic or phonemic category or subcategory. Categories were based on naturally occurring clusters for each participant. Thus, semantic clusters sometimes represented overarching categories (e.g., birds; items found in the kitchen) and at other times more detailed categories (e.g., birds of prey, water birds; cooking utensils, silverware). A transition between two words that did not belong to the same cluster, including single-word transitions, was considered a switch. General and specific rules for phonemic and semantic clustering in both task types are described in Study I, Appendix A, including comprehensive examples and sample trials.

To verify the reliability of the suggested clustering and switching analyses, an inter-rater reliability analysis was calculated using Intraclass Correlation Coefficient (ICC) for all task-congruent and task-discrepant, phonemic, and semantic cluster sizes in both VF tasks. A two-way random-effects model with a single measurement and absolute agreement was selected to show the level of agreement achieved between the two raters (Koo & Li, 2016).

For the phonemic cluster size, the number of phonemic switches, and the number of semantic switches, the ICC analysis showed an excellent degree of reliability between the two raters as ICCs are above 0.90 (Koo & Li, 2016). For the semantic cluster size, the reliability between the raters was good; ICC value between 0.75 and 0.90 (Koo & Li, 2016). A blind review by a third rater showed that lower ICC in the semantic clustering analysis resulted primarily from differences between the raters in instances where there were multiple semantically acceptable ways to form clusters. As subjective semantic categorization will always be present in semantic clustering analysis (Tröger et al., 2019), this result was determined to demonstrate acceptable semantic variation between raters.

5 Overview of the Studies and Results

In this chapter I provide an overview of the three articles that form the foundation of this dissertation by publication. The first article introduces a systematic method for VF task analyses. The subsequent two articles apply the proposed method to a group of immersed bilinguals, contrasting L1 and L2 performance, evaluating the impact of extralinguistic variables on task performance and comparing L1 performance with that of a control group residing in Finland. The analyses were divided between the studies as follows: Study II investigated total scores, errors and temporal aspects and Study III focused on clustering and switching analyses. As discussed in Section 1.1., the published articles II and III adopt the terms *language attriter* to refer to the group of immersed bilinguals and *monolingual* to refer to the group of Finnish speakers living in Finland, to align with established conventions in language attrition studies. Data and analysis scripts are available online via the Center for Open Science.[2]

In this chapter, Section 5.1 outlines the process and outcomes for developing and verifying the method for VF task analysis (Study I). Section 5.2. details analysis between L1 and L2 within the group of immersed bilinguals for total scores, temporal aspects, errors, and clustering and switching strategies. It also discusses the impact of extralinguistic variables on task performance within the immersed bilingual group (Studies II and III). Section 5.3 focuses on group comparisons between immersed bilinguals and the control group residing in Finland. It outlines hypotheses and results for total scores, temporal aspects, errors, and clustering and switching analysis from the comparison between the groups (Studies II and III).

[2] Study I: [https://osf.io/kh8f3/?view_only=9eb45e1eabd641e0b421be8e2808ccc8], Study II: [https://osf.io/fue3k/?view_only=6b6762f07e2243d6b8548c0992dce9f1], Study III: [https://osf.io/95q3j/?view_only=9a0e3e3d8dee4c6da143f4fcca290bab]

5.1 Administration and Analysis of Verbal Fluency Tasks (Study I)

The aim for Study I was to generate guidelines for systematic implementation and scoring of VF tasks and to verify the proposed method. In this article, instructions for administration, scoring, and analyses for VF total scores, errors, temporal parameters, clustering, and switching were developed and compiled to form "Instruction Manual for Administration and Scoring Verbal Fluency Tasks" (Study I, Appendix A). To validate the proposed method's reliability, the analysis method was demonstrated in a sample of middle-aged and older Finnish-speaking adults residing in Finland (N = 50, introduced in Section 4.1).

The assessments for method validation explored whether task type (PVF vs. SVF), age, education, or gender could predict the total score, and the number of words generated across four 15-second segments. Additionally, the frequency and types of errors in PVF and SVF tasks were examined. Regarding the use of clustering and switching strategies, it was first assessed whether the size of task-congruent clusters, the number of switches, or their interactions could predict the total score. Subsequently, an investigation was conducted to determine whether the frequency of task-discrepant clusters could predict the total score. The analyses were conducted utilizing R software by the third author of Study I (R Core Team, 2019).

Total Scores

To examine the impact of task type, participant age, education, and gender on VF total scores, a linear mixed-effects model was implemented. In addition, descriptive data on total scores for age brackets of 49–59, 60–69, and 70–79 were provided to supplement the analyses.

Consistent with multiple normative datasets (for an overview, see Strauss et al., 2006), healthy, neurotypical middle-aged and older adults generated a higher total score in SVF than in PVF tasks. In the PVF, the total scores for each phoneme mirrored the category size of word-initial phonemes in the Finnish language, as anticipated (Gollan et al., 2002). Education was assessed on a 2-tier scale, revealing a positive correlation with performance on both fluency types, which is in line with findings from Pereira et al. (2018). As hypothesized, no correlation was found between age and task performance in either task type, potentially due to the relatively narrow age range in the sample (Ardila, 2020).

Errors

The frequency and type of errors were reported as raw scores, given the small number of errors in the data. Mean values were presented to facilitate comparison between tasks. The number of participants who generated errors was detailed for each task separately.

Error frequency and variety followed error profiles described by Crowe (1998) and Gollan et al. (2011), with repetitions emerging as the most common error type. Consequently, the generation of errors, particularly repetitions, appeared to be a part of the verbal retrieval process across tasks for a healthy, older, monolingual population.

Temporal parameters

For the temporal parameters of VF tasks, the number of words generated in four 15-second segments within the total 60 seconds were investigated. To determine if the task type (PVF vs. SVF) predicted the number of words produced during each 15-second segment of the task, the number of words produced was modeled as a function of each 15-second segment. Model selection procedure based on analysis of variance and Bayesian Information Criterion (BIC) values proved the model without participant background information (age, education and gender) to be the most parsimonious fit for the data in Study I.

Task performance in 15-second segments revealed a decrease in performance as the task progressed during both task types, potentially indicating an increase in effort required for word retrieval as time progressed (Crowe, 1998). The slope of decline was more substantial in SVF (as illustrated in Section 5.3, Figure 7). The results also corroborated findings from earlier literature, demonstrating a strong association between performance in the first quartile and the overall score (Venegas & Mansur, 2011).

Clustering and Switching

To evaluate the effectiveness of clustering and switching, the total score was modeled as a function of task-congruent mean cluster size, the number of switches, and their interactions. Separate models were used for PVF and SVF tasks. In the PVF model, participant intercept and slope for the task (K, A, or P) were incorporated as random factors. Since the SVF dataset only had one task per participant, using a mixed-effects model was deemed unnecessary, and a simple linear regression was employed instead. Model selection suggested the model with no background variables as additional predictors to be the most parsimonious fit for the PVF, and the model with education as an additional predictor to be the most parsimonious fit for the SVF.

The analysis detected that task-congruent clustering and switching were effective strategies in both PVF and SVF, aligning with the findings of Troyer et al. (1997).

The main effect of education in the SVF was consistent with the observation in the total score model reported earlier in this chapter.

To examine the use of task-discrepant clusters, the total score in the PVF was modeled as a function of the number of semantic clusters, and the total score in the SVF was modeled as a function of the number of phonemic clusters. As the order of phonemic trials was not randomized, trial order was included as a predictor to investigate whether the position of the phonemic trial affected the number of semantic clusters. Trials with no task-discrepant clusters were excluded from the analysis. Based on a model selection procedure, the models with no background variables were selected as the most parsimonious ones.

The analysis revealed that task-discrepant clustering was common in PVF trials, aligning with Abwender et al. (2001). Notably, the number of semantic clusters predicted the total score in PVF, but the number of phonemic clusters in SVF did not.

Conclusion

Study I presented comprehensive guidelines for PVF and SVF task administration, scoring, and analyses including total scores, temporal parameters, errors, and clustering and switching with strong inter-rater reliability in a sample group of 50 older, healthy participants. The guidelines for these proposed analyses were detailed in “The Instruction Manual for Administration and Scoring Verbal Fluency Tasks” (Study I, Appendix A).

Analyses for method validation demonstrated consistent outcomes with previous studies, thereby affirming the reliability of the proposed guidelines for administering, scoring, and analyzing VF performance. While applying the proposed method to diverse target groups and larger data pools is required for further validation, this study is expected to guide future research in various fields by offering a straightforward scoring framework for PVF and SVF tasks. Guidelines for data analysis are intended to serve as reliable, language-neutral tool for thorough analyses of VF task performance in a variety of clinical and research settings.

5.2 Comparison Between L1 and L2 within Immersed Bilinguals (Studies II and III)

This section focuses on comparisons between VF task performance in L1 and L2 within the group of immersed bilinguals. Evaluating the dynamics of L1 and L2 and the impact extralinguistic variables have on both languages is imperative as all languages in a bilingual brain interact continuously in a bidirectional process and can alter the cognitive and neural dynamics of both languages (Grosjean, 2013;

Gurunandan et al., 2022; Laine & Lehtonen, 2018; Linck & Kroll, 2019; Treffers-Daller, 2019). In this dissertation, performance between L1 and L2 was evaluated for total scores, errors, temporal parameters (Study II) and the use of clustering and switching strategies (Study III) in the immersed bilingual group. The analyses were conducted utilizing R software by the second author of Studies II and III (R Core Team, 2019).

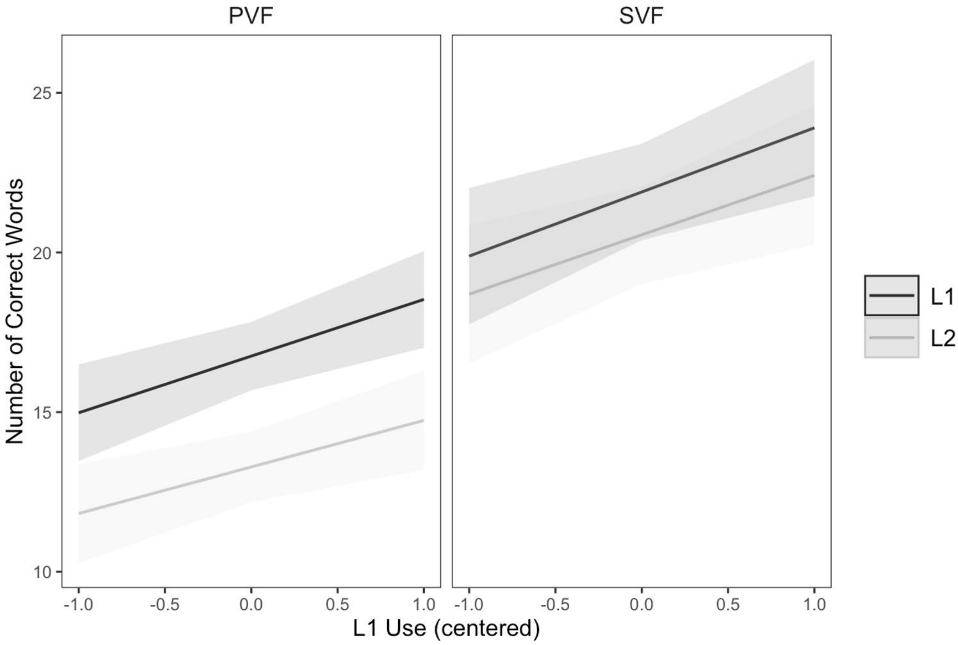
Total Scores

The group of immersed bilinguals self-reported a balanced bilingualism with a slight preference for L2. This was expected to be reflected in the total scores as comparable L1 and L2 performance or stronger performance in their self-reported dominant language (L2) (Roberts & Le Dorze, 1997; Rosselli et al., 2002). Differences between languages were expected to be more pronounced in SVF than in PVF (Goral, 2004; Friesen et al., 2015; Patra et al., 2020; Rosselli et al., 2000; Schmid & Köpke, 2009). For extralinguistic variables, shorter LoR and more frequent L1 use were expected to positively affect L1 total scores in PVF and SVF but these variables were expected not be strong predictors of performance independently (de Bot & Clyne, 1994; Schmid, 2013; Schmid, 2019).

A linear mixed-effects model was employed to examine total scores as a function of task language (L1/L2), task type (PVF/SVF), LoR, and frequency of L1 use and their interactions in Study II. The model selection procedure based on BIC values (L1 use, LoR, or both, combined with task language and fluency type) suggested the model with L1 use but no LoR as the best fit for the data.

Contrary to expectations, a higher number of acceptable words was generated in L1 than L2 across fluency types. Frequent use of L1 supported performance in both languages and this is illustrated in Figure 3.

Figure 3. Predicted Values of Correct Words in the Immersed Bilingual Group in One Semantic and Three Phonemic Tasks (Combined) in L1 and L2 with Frequency of L1 Use as the Predictor.



Note. Reprinted from "Frequent native language use supports phonemic and semantic verbal fluency in L1 and L2: An extended analysis of verbal fluency task performance in an L1 language attrition population," by N. Lehtinen, A. Kautto, and K. Renvall, 2024, *International Journal of Bilingualism 28(5),* 884-906. (https://doi.org/10.1177/13670069231193727). Reprinted under Creative Commons Attribution 4.0 License.

Errors

Participants were healthy, neurotypical adults. Thus, minimal errors were expected across languages. Immersed bilinguals were projected to generate more cross-language intrusion errors in their self-reported less dominant language (L1) (Sandoval et al., 2010; Gollan et al., 2011) and longer LoR and less frequent L1 use was expected to result in an increased number of errors in L1 but not L2.

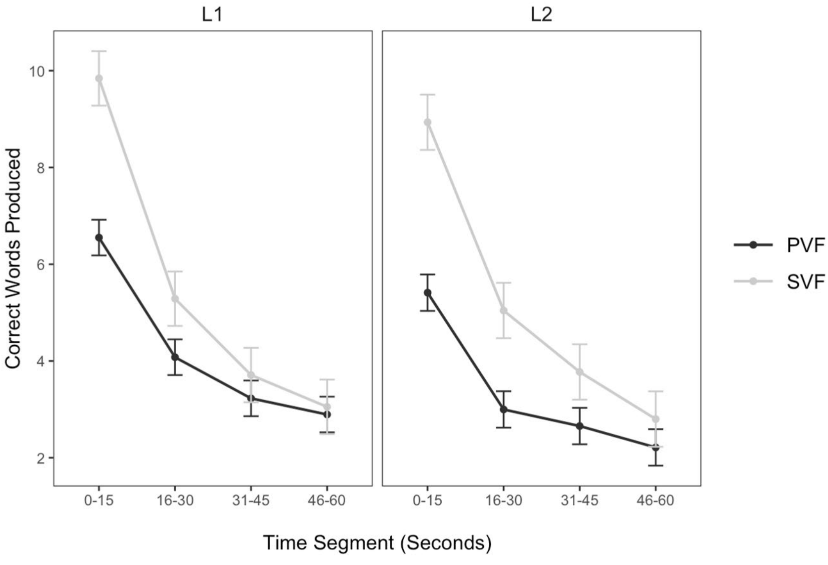
The analysis in Study II showed that immersed bilinguals generated minimal errors in both languages, with repetitions being the most common error type. A Wilcoxon Rank-Sum Test was used to determine that there were no significant differences in the number of errors between L1 and L2. As the number of errors was minimal, the impact of extralinguistic variables on errors was excluded from the analysis.

Temporal Parameters

For temporal parameters, immersed bilinguals were projected to employ rapid retrieval strategies more efficiently in their self-reported stronger language (L2), particularly in SVF (Crowe, 1998; Fernaeus & Almkvist, 1998; Fernaeus et al., 2008; Venegas & Mansur, 2011). Shorter LoR and more frequent L1 use were expected to facilitate rapid retrieval in L1 but these variables were not projected to be strong independent predictors of performance (de Bot & Clyne, 1994; Schmid, 2013; Schmid, 2019).

In Study II, the temporal profile of task performance and its association with task attributes (SVF / PVF and L1 / L2) was investigated by modeling the total number of acceptable words produced during four 15-second time windows, with L1 use and LoR as predictors. The temporal performance profile within the task mirrored expectations by presenting similarly for both languages, as visualized in Figure 4.

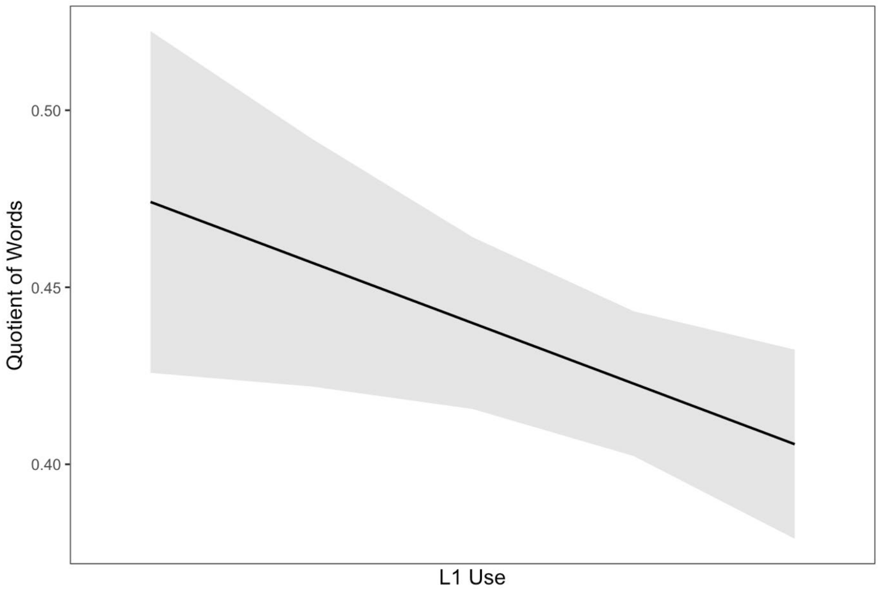
Figure 4. Predicted Values of Correct Words in the Four 15-Second Time Segments in One Semantic and Three Phonemic Verbal Fluency Tasks in the Immersed Bilingual Group in L1 and L2.



Note. Error Bars Represent 95% Confidence Intervals. Reprinted from "Frequent native language use supports phonemic and semantic verbal fluency in L1 and L2: An extended analysis of verbal fluency task performance in an L1 language attrition population," by N. Lehtinen, A. Kautto, and K. Renvall, 2024, *International Journal of Bilingualism 28(5),* 884-906 (https://doi.org/10.1177/13670069231193727). Reprinted under Creative Commons Attribution 4.0 License.

To further examine temporal performance variation between the languages, the quotient of acceptable words produced during the first time segment in all tasks and the impact of the frequency of L1 use were analyzed. Immersed bilinguals generated a smaller quotient of total words in L1 (42%) than in L2 (44%) during the first 15-second time segments, but this difference was not statistically significant. Contrary to expectations, immersed bilinguals who used L1 more frequently generated a smaller quotient of words in the initial stage of the tasks in both languages than those who used L1 less frequently, illustrated in Figure 5.

Figure 5. Predicted Values of the Quotient of Correct Words in the First 15-Second Time Segment in the Immersed Bilingual Group in L1 and L2 (Combined) in All Task Types (Combined) with Frequency of L1 Use as Predictor.



Note. Reprinted from "Frequent native language use supports phonemic and semantic verbal fluency in L1 and L2: An extended analysis of verbal fluency task performance in an L1 language attrition population," by N. Lehtinen, A. Kautto, and K. Renvall, 2024, *International Journal of Bilingualism 28(5),* 884-906 (https://doi.org/10.1177/13670069231193727). Reprinted under Creative Commons Attribution 4.0 License.

Clustering and Switching

As L2 was acquired later in life, immersed bilinguals were expected to demonstrate a robust semantic network in L1 by relying more efficiently on clustering than switching in L1 compared to L2 (Luo et al., 2010; Strauss et al., 2006; Troyer et al., 1997; Troyer 2000). Frequent L1 use was anticipated to support switching in both languages and task types, potentially reflecting cognitive flexibility stemming from more frequent L1 use in an L2-dominant environment (e.g, Gollan et al., 2002).

Clustering and switching analyses in Study III were conducted using linear mixed-effects models to predict total scores as a function of task-congruent cluster size, the number of switches and language (L1/L2) separately for PVF and SVF with LoR and L1 use as predictors. The analysis demonstrated that immersed bilinguals employed clustering and switching strategies with equal efficiency to achieve high scores across languages, as well as in both task types. The equal efficacy of both strategies mirrors typical monolingual SVF performance as described by Troyer, Moscovitch, and Winocur (1997). LoR or frequency of L1 did not appear to influence the efficacy of either strategy.

Summary of Findings

In sum, comparison between L1 and L2 demonstrated strong proficiency in L1 and similar lexical retrieval strategies in both languages. Frequent use of L1 supported overall task performance, and proportionally slowed down the performance in the initial stage of the task in both languages. Analysis on clustering and switching revealed similar efficacy of clustering and switching strategies in both languages with performance mirroring typical SVF performance in the control group of Finnish speakers residing in Finland.

5.3 Group Comparison Between Immersed Bilinguals and the Control Group (Studies II and III)

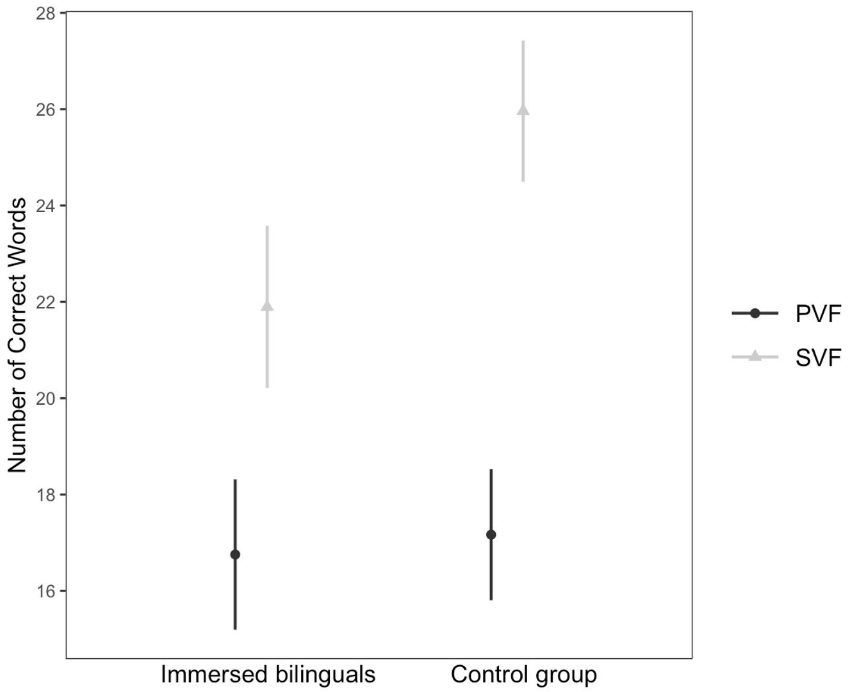
This section provides an overview of L1 performance between the group of immersed bilinguals and the control group of Finnish speakers living in Finland for VF total scores, errors, temporal parameters (Study II) and the use of clustering and switching strategies (Study III). By comparing the L1 performance between these groups, this dissertation aims to contribute to the developing infrastructure of language attrition studies that adhere to the LATB protocol developed by the European Graduate Network on Language Attrition (Schmid, 2011b).

Total Scores

Immersed bilinguals were expected to achieve lower total scores in SVF than the control group but demonstrate comparable performance in PVF, following previous literature (Gollan et al., 2002; Ljungberg et al., 2013; Luo et al., 2010; Marsh et al., 2019; Patra et al., 2020; Roberts & Le Dorze, 1997; Rosselli et al., 2000; Rosselli et al., 2002; Sandoval, 2010; Schmid & Jarvis, 2014; Schmid & Köpke, 2009).

In line with the hypothesis, a linear mixed-effects model with total scores as the outcome variable and task type (PVF/SVF) and participant group as predictors in Study II indicated that immersed bilinguals generated fewer acceptable words in the SVF than the control group, while performance in PVF task was similar across groups, as shown in Figure 6.

Figure 6. Comparison of Predicted Values of Number of Correct Words in One Semantic and Three Phonemic Verbal Fluency Tasks (Combined) in L1 Between the Immersed Bilingual and Control Group Using Participant Group as a Predictor.



Note. Adapted from "Frequent native language use supports phonemic and semantic verbal fluency in L1 and L2: An extended analysis of verbal fluency task performance in an L1 language attrition population," by N. Lehtinen, A. Kautto, and K. Renvall, 2024, *International Journal of Bilingualism 28(5),* 884-906 (https://doi.org/10.1177/13670069231193727). Reprinted under Creative Commons Attribution 4.0 License.

Errors

The occurrence of errors was expected to be minimal overall, with a lower number of error-free trials in the group of immersed bilinguals (Badstübner, 2011; Gollan et al., 2002; Gollan et al., 2011; Sandoval et al., 2010).

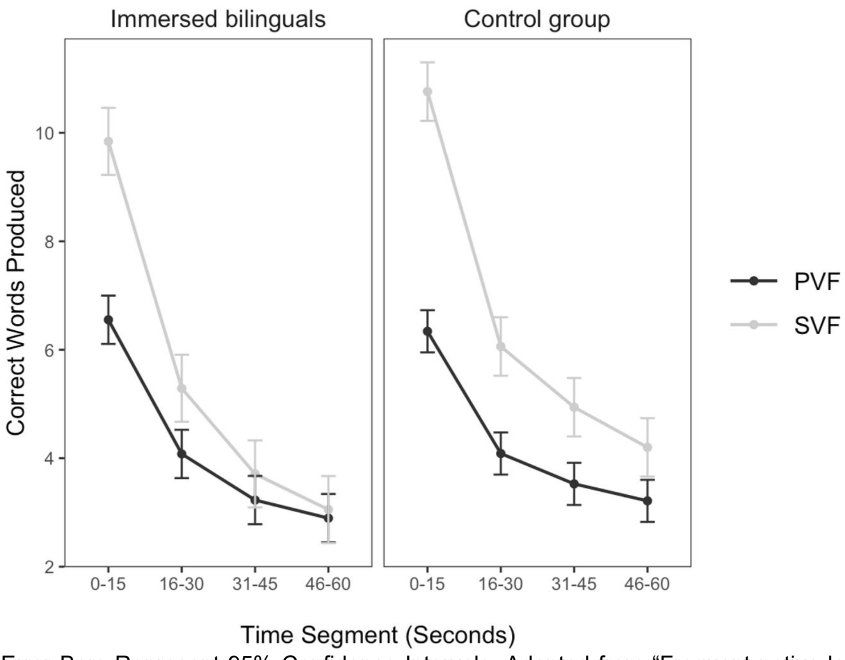
The analysis in Study II revealed that immersed bilinguals made more errors than controls in the PVF, but error counts were comparable in SVF, as determined by the Wilcoxon Rank-Sum Test. Qualitatively, repetitions were the most common error in both groups and fewer immersed bilinguals than controls produced error-free trials in both tasks.

Temporal Parameters

For temporal parameters it was anticipated for immersed bilinguals to retrieve words more slowly than the control group in the early stages of the task, indicating bilingual disadvantage due to language interference, especially in SVF (Schmid & Jarvis, 2014).

To investigate performance changes during the task, a model with the number of correct words generated during 15-second intervals as the outcome variable, and participant group, time sequence, and task type as predictors was fitted in Study II. The temporal performance profile in L1 did not differ between immersed bilinguals and the control group when measured as performance across 15-second segments, as visualized in Figure 7.

Figure 7. Predicted Values of Words Produced in the Four 15-second Time Segments in One Semantic and Three Phonemic Verbal Fluency Tasks in the Immersed Bilingual and Control Group.



Note. Error Bars Represent 95% Confidence Intervals. Adapted from “Frequent native language use supports phonemic and semantic verbal fluency in L1 and L2: An extended analysis of verbal fluency task performance in an L1 language attrition population,” by N. Lehtinen, A. Kautto, and K. Renvall, 2024, *International Journal of Bilingualism 28(5),* 884-906 (https://doi.org/10.1177/13670069231193727). Reprinted under Creative Commons Attribution 4.0 License.

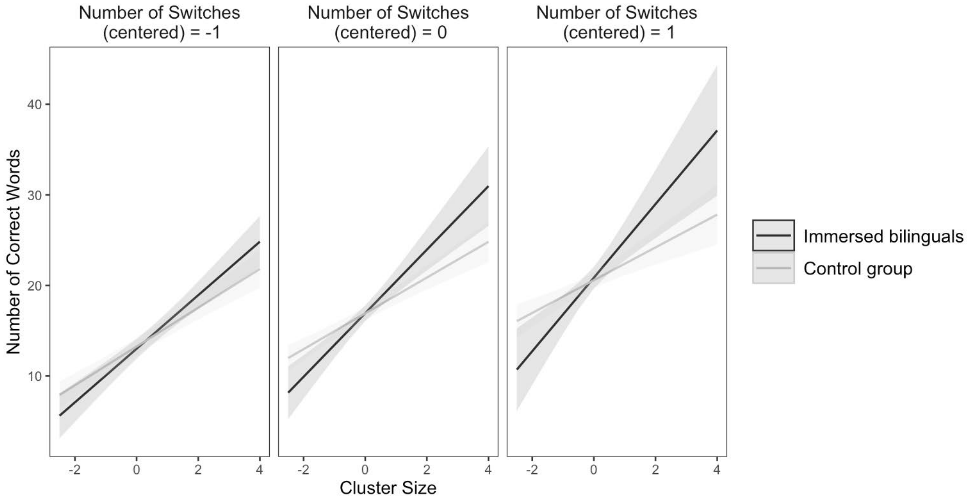
To further compare the ability to generate words rapidly in the early stages of the task between immersed bilinguals and the control group the ratio of correct words generated during the first 15-seconds to the total score was modeled in Study II, with task type and participant group as predictors. Immersed bilinguals produced a higher quotient of words than controls in the first 15-second segment, suggesting that immersed bilinguals were more efficient in generating words in the initial stage of the task than the control group.

Clustering and Switching

Immersed bilinguals were expected to demonstrate superior executive flexibility by relying more on switching than the control group in PVF (Gollan et al., 2002; Patra et al. 2020). In SVF, immersed bilinguals were anticipated to display a weaker semantic foundation and greater cognitive flexibility than controls by relying more on switching than clustering strategies to achieve a high total score (Roberts & Le Dorze, 1997; Gollan et al., 2002).

The analysis on clustering and switching strategies between the groups in Study III was conducted by modeling total scores as a function of task-congruent cluster size, the number of switches, and the participant group separately for PVF and SVF. Participant ID was used as a random factor in the PVF model as the PVF task data included three observations for each participant (K, A, and P). Contrary to expectations, no significant differences were found in the effectiveness of switching strategies between the groups in PVF in Study III. Moreover, immersed bilinguals appeared to generate task-congruent clusters more efficiently than the control group in PVF, as visualized in Figure 8.

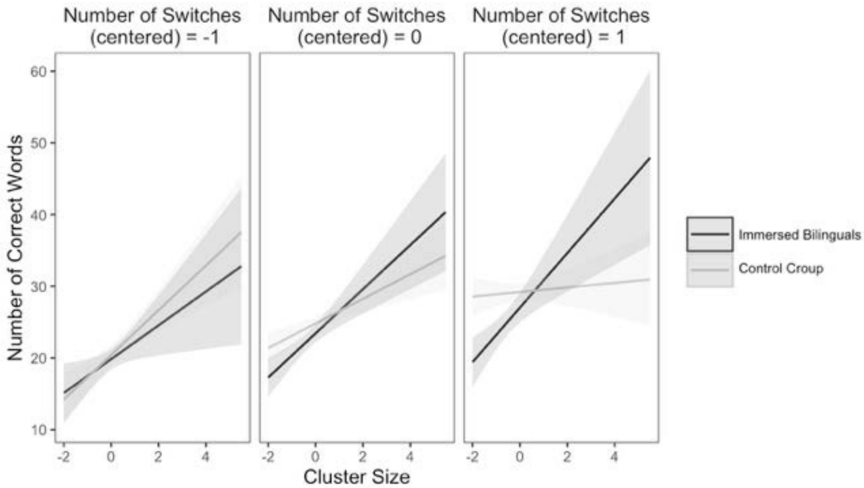
Figure 8. Number of Correct Words as a Function of Task-Congruent Cluster Size (Centered), the Number of Switches (Centered), and Participant Group in the Phonemic Verbal Fluency Task.



Note. Adapted from “Efficacy of clustering and switching strategies in verbal fluency tasks in Finnish-English language attrition population,” by N. Lehtinen, A. Kautto, and K. Renvall, 2023, *The Language Learning Journal*, *52(2), 218–231* (https://doi.org/10.1080/09571736.2023.2294060). Reprinted under Creative Commons CC BY License.

This reliance on clustering for optimal performance was also evident in SVF, where the correlation between cluster size and total score remained consistent in the group of immersed bilinguals, regardless of the number of switches, as visualized in Figure 9. In line with expectations, the control group employed clustering and switching strategies more evenly to achieve higher scores, aligning with Troyer, Moscovitch, and Winocur (1997).

Figure 9. Number of Correct Words as a Function of Task-Congruent Cluster Size (Centered), the Number of Switches (Centered), and Participant Group in the Semantic Verbal Fluency Task.



Note. Adapted from "Efficacy of clustering and switching strategies in verbal fluency tasks in Finnish-English language attrition population," by N. Lehtinen, A. Kautto, and K. Renvall, 2023, *The Language Learning Journal*, *52(2), 218–231* (https://doi.org/10.1080/09571736.2023.2294060). Reprinted under Creative Commons CC BY License.

Summary of Findings

Taken together, the findings from comparisons between immersed bilinguals and the control group of Finnish speakers residing in Finland show promise in locating the underlying strategies responsible for the detected group differences in VF task total scores. Immersed bilinguals generated a larger portion of the total number of words in the first 15-second segment compared to the control group. Thus, immersed bilinguals seemed to rely more on rapid retrieval in L1 than the control group and the disadvantage in SVF did not appear to be due to a slower initiation in the immersed bilingual group, contrary to the hypothesis. Clustering and switching analysis showed that immersed bilinguals relied more on clustering strategies than the control group and did not seem to be able to boost their performance by efficient switching as well as the control group. Participants in the control group utilized both strategies, clustering and switching, more equally to achieve a higher total score. Utilizing these strategies, immersed bilinguals reached a lower total score in SVF and comparable performance in PVF, regardless of immersed bilinguals generating more errors in PVF than the control group.

6 Discussion

The primary aim of this dissertation to identify processes that define and drive VF task performance among a group of Finnish-English bilinguals experiencing reduced exposure and use of their first language in an immersive second language environment. To accomplish this overarching objective, three distinct studies were conducted, each with its own set of sub-objectives. In this discussion, I aim to draw conclusions about the analytical methods developed in this dissertation and the findings related to strategies that either support or impede optimal performance in VF tasks within a bilingual Finnish-English population immersed in an English-speaking environment. I will also discuss the significance of the findings in relation to lexical processing in bilinguals with varied backgrounds.

Following the structure of Chapter 5, the VF task analysis method developed and utilized in this dissertation is discussed in Section 6.1. Section 6.2 focuses on the comparison between L1 and L2 VF task performance within the immersed bilingual group and the impact of extralinguistic variables they can have on task performance. Section 6.3 compares the performance in L1 between the group of immersed bilinguals and a control group of Finnish speakers residing in Finland. Prior to presenting the major contributions and final conclusions, the contributions and limitations of this dissertation will be discussed, along with potential future research directions that emerge from these studies.

6.1 Developing Guidelines for a Systematic Administration and Analysis (Study I)

The objective for the first publication of this dissertation (Study I) was to establish systematic guidelines and provide tools for data collection and comprehensive analyses of VF tasks. Study I demonstrated and verified a comprehensive method for investigating lexical processing strategies in VF task performance in a group of healthy, neurotypical middle-aged and older Finnish speaking adults. This approach was deemed reliable for analyzing the linguistic and cognitive processes in various clinical groups and the guidelines on conducting the proposed analyses were compiled into “The Instruction Manual for Administration and Scoring Verbal Fluency Tasks” (Study I, Appendix A). However, considering the rather limited sample sizes and small age range of participants, further exploration of the suggested analyses in larger data samples of healthy participants with a broader age range, as well as across languages and in varied clinical groups, is needed to solidify findings from Study I.

The proposed method was designed for straightforward adaptation to multiple languages, incorporating language and culture-specific considerations into the analyses. This was achieved by selecting one semantic category, “animals,” which is culturally and linguistically relatively universal, and by utilizing language-specific PVF categories. For the clustering and switching analysis, clusters were determined based on naturally occurring patterns for each participant.

Task categories included three phonemic prompts and one semantic prompt. Although a balanced number of categories across task types would be ideal, drawing parallels for task analysis between PVF and SVF tasks can be complex. Analysis of letter categories in PVF is more straightforward, and multiple letter categories can be combined for analysis in a relatively simple manner. In SVF, the semantic component of tasks limits their comparability both across and within populations with diverse demographic and cultural backgrounds, as well as between different semantic categories themselves (e.g., Abwender et al., 2001; Olabarrieta-Landa et al., 2017; Opitz, 2011; Roberts & Dorze, 1997; Rosselli et al., 2002). To facilitate data analysis guidelines that are not language specific and to reduce the complexity of semantic analysis, this dissertation includes only one semantic category: the concrete, culturally and linguistically neutral, and widely researched category of 'animals' (e.g., Pekkala et al., 2009).

The prompts in PVF were selected to include the two most frequent word-initial consonants (P, K) and the most frequent word-initial vowel (A) of the target language, Finnish, following the high-frequency dictionary approach to account for individual characteristics of the language (Mardani et al., 2020; Oberg & Ramirez, 2006). However, certain languages have strong traditions for including specific prompts in PVF tasks. For instance, in English, it is typical to use the letter combinations F, A, S, based on early research (Borkowski et al., 1967). Consequently, there is a substantial body of research based on certain letter combinations that do not follow the high-dictionary approach (Strauss et al., 2006). Therefore, it is left to the researcher to evaluate whether it is more valuable for their research objective to use letters conforming to earlier research in a specific language or select letter prompts to match their word-initial frequency in the target language for comparability between studies conducted in different languages.

For clustering and switching analysis it is important to bear in mind that different methods for determining clusters and calculating switches can result in varied outcomes (Abwender et al., 2001; Thiele et al., 2016), making comparisons between studies unreliable. Many previous studies use predetermined subcategories for clustering. Predetermined categories can be selected following earlier studies, such as Troyer et al. (1997) or determined specifically for a single study. The method proposed in this study includes calculating naturally occurring clusters and switches for each task. Calculating naturally occurring clusters and switches for each participant following uniform rules encompasses all unique clustering strategies and natural switches generated by participants, including language and culture-specific clusters (Roberts & Le Dorze, 1997; Rosselli et al., 2002). Naturally occurring strategies allowed for subcategories in an area of expertise (e.g., birds of prey, aquatic birds), clustering based on geographical semantics (giraffe, monkey), and visual semantics (snake, eel). It also facilitated the use of context to determine the intended category (e.g., forest animals, animals typical to Lapland; Becker & Salles, 2016).

The inter-rater analysis for calculating clusters showed that the method for determining naturally occurring cluster was sufficiently reliable. However, as manual analysis of semantic clustering strategies involves a subjective semantic component (Tröger et al., 2019), a more significant margin of error between raters in the semantic than in phonemic cluster analysis was detected. Precise instructions for determining naturally occurring clusters were included in the “The Instruction Manual for Administration and Scoring Verbal Fluency Tasks” (Study I, Appendix A) with the aim of minimizing this variability. To further support semantic clustering analysis for naturally occurring clusters, a sample protocol and rater training instructions were also included in the manual, as suggested by Ross (2003).

6.2 Performance Within Immersed Bilinguals in L1 and L2 (Studies II and III)

The immersed bilingual group consisted of individuals whose L1 was Finnish and who had emigrated to an L2 English language environment at an age when L1 was considered well-established. This dissertation aimed to identify similarities and differences in L1 and L2 language processing within this group.

At the time of the interview, the perceived language dominance in the immersed bilingual group was balanced bilingualism, with a slight preference for L2. Participants demonstrated higher VF total scores in L1, indicating its dominance. The qualitative analysis on errors demonstrated minimal errors in L1 and L2. However, the number of errors in L1 was higher in the immersed bilingual group than in the control group, and fewer immersed bilinguals produced error-free trials (see Sections 5.3 and 6.3). These observations can indicate speaker insecurities and reduced accuracy in L1, which are often labelled as general markers of language attrition (Jarvis, 2019; Schmid & Köpke, 2009). As such, they might have influenced the self-reported preference for L2 in the group of immersed bilinguals (Andersen, 1982).

A comparison between L1 and L2 VF task processing strategies revealed similar temporal word distributions across VF tasks in both languages. An interesting direction for future studies would be to explore whether these behavioral findings reflect comparable neural activation patterns during L1 and L2 performance, as observed in adult language learners (Gurunandan et al., 2022). The consistent use of clustering and switching strategies, along with low error rates in both L1 and L2, may provide modest support for this preliminary observation.

Analysis of the impact of extralinguistic variables showed that the Length of Residence (LoR) did not appear as a significant predictor on any of the analysis separately. This was expected as the inclusion criteria included a minimum of 20 years as LoR, with the average length of residence resulting in 34.24 years in the immersed bilingual group. This surpasses the threshold of the first ten years of LoR after which the impact of LoR has been shown to stabilize (Opitz, 2011; Schmid, 2011a; Schmid, 2019). The analysis frequent use of L1 was shown to enhance overall performance but slow down performance in both languages. The influence of frequent L1 use on both languages, rather than on a specific one, further supports a common bilingual effect across languages.

6.3 Performance in L1 Between Immersed Bilinguals and the Control Group (Studies II and III)

In addition to examining L1 and L2 VF task performance within immersed bilinguals, this dissertation also contrasted L1 performance between the immersed bilinguals and a control group of L1 speakers who reside in Finland. The groups were matched for age, gender and education. This approach is consistent with numerous language attrition studies across various language pairs.

Comparisons between the immersed bilingual group and the control group revealed differences in SVF and PVF task performance. In line with previous research, the immersed bilingual group scored lower in total SVF (Badstübner, 2011; Dostert, 2009; Schmid, 2011a; Schmid, 2011b; Schmid & Dusseldorp, 2010; Schmid & Jarvis, 2014; Schmid & Keijzer, 2009; Schmid & Köpke, 2009) and no significant group differences were observed in PVF in this study. The difference between PVF and SVF task performance can indicate the existence of group-specific subcomponents in lexical retrieval strategies (Goral, 2004). In SVF tasks, where lexical retrieval strategies are assumed to rely on hierarchical mental lexicon and memory organization (Patra et al., 2020; Strauss et al., 2006), immersed bilinguals produced fewer acceptable words than the control group, despite similar error rates.

In contrast, in PVF tasks which engage more strategic cognitive organization and require maintenance of effort (Barry et al., 2008; Santos Nogueira et al., 2016; Strauss et al., 2006) both groups generated similar total scores despite the higher error rate in the immersed bilingual group.

Contrary to expectations (Rosselli et al., 2000; Rosselli et al., 2002; Sandoval et al., 2010; Schmid & Jarvis, 2014), analysis of temporal parameters of VF task performance showed that immersed bilinguals did not retrieve fewer words in the early stages of the tasks than the control group. Thus, no temporal indicators of language interference slowing down lexical retrieval were detected to account for the lower total scores in SVF. In fact, immersed bilinguals produced a larger percentage of total words in the initial 15-second segment than the control group. Considering that early task segments expedite word-finding through semi-automatic rapid retrieval strategies (Fernaeus et al., 2008; Sandoval et al., 2010; Venegas & Mansur, 2011), this finding suggests that the immersed bilinguals may use rapid retrieval more efficiently than the control group in the initial 15-second interval, potentially implying that these strategies can be resistant to language attrition (Gollan et al., 2002; Segalowitz, 1991). It can also be noted that despite frequent L1 use slowing down immersed bilinguals in the first 15 seconds of the task segment, they still outperformed the control group, even though the overall performance resulted in lower total scores in SVF for the immersed bilingual group.

To identify processes that impede optimal performance, particularly in SVF in the immersed bilingual group, an analysis of clustering and switching was performed. The comparison between the groups revealed that immersed bilinguals were able to sustain the cognitively demanding strategy of generating task-congruent clusters more efficiently than the control group in both task types. In PVF, immersed bilinguals generated task-congruent phonemic clusters more effectively than the control group, achieving a total score comparable to that of control group. In SVF, the relationship between semantic cluster size and total score remained consistent in the group of immersed bilinguals, regardless of the number of switches. Conversely, the control group utilized clustering and switching strategies more evenly to achieve higher scores, consistent with Troyer, Moscovitch, and Winocur (1997).

Based on these findings, it appears that immersed bilinguals were able to generate clusters more efficiently than the control group especially in the early segments of the task utilizing rapid retrieval strategies, but they were less capable of shifting back and forth between categories once a category had been depleted, unlike their controls. These strategies resulted in lower total scores in SVF in the immersed bilingual group and similar performance between the groups in PVF.

Studies II and III were the first studies on bilingual populations to utilize analysis methods introduced in Study I. Thus, findings from this study cannot be directly compared to previous studies applying varied analysis methods for clustering and switching. However, it is worth noting that findings from this study align with Patra et al., (2020), who investigated the use of clustering and switching strategies in a group of bilingual speakers with an emigration background. Furthermore, the findings of this study contrast with those from Mardani et al., (2020), who investigated clustering and switching strategies in a group of bilingual speakers with no emigration background. While these studies cannot be directly compared, the findings from this dissertation offer cautious optimism. They suggest that the proposed method could potentially identify typical group preferences in optimal VF task processing strategies among immersed bilingual groups within the complex and dynamic interplay of bilingual language processes.

6.4 Limitations and Future Considerations

In this section, I discuss the limitations of the present studies in relation to how these factors could be addressed in future research. First, I consider the proposed method in Study I, followed by its application to the Finnish-English immersed bilingual group in Studies II and III. Next, I discuss factors related to the extralinguistic variables considered in this dissertation. Lastly, the participants and the data analyzed for this study are discussed before concluding with a brief description of the data collected within the research project, but not analyzed in this dissertation.

6.4.1 Analysis Method (Study I)

Considerations for refining the proposed guidelines for VF task administration include clarifying the task instruction for the PVF task and randomizing the order in which the VF task categories are presented. Originally, the word "letter" was selected for the task instruction based on the strong letter-phoneme correspondence of Finnish. In "The Instruction Manual for Administration and Scoring Verbal Fluency Tasks" (Study I, Appendix A) the prompt for the PVF task reads: "In this next category, I ask you to name as many words beginning with the same letter as possible within one minute". As the goal is not to evaluate spelling fluency, the writers suggested that the word "letter" should be replaced with the word "sound" in the task instruction in the future and this suggestion was noted in the instruction manual. This change is particularly important for languages without a strong letter-phoneme correlation.

For this dissertation, the order of VF tasks was not randomized and the SVF task was always presented before the PVF tasks. In Study I more errors were detected in the PVF than SVF, in contrast to expectations (Crowe, 1998). Task-discrepant (semantic) clusters were more prevalent in the PVF than in SVF. Thus, the potential priming effect from the SVF to PVF was considered. No semantic errors were detected in the PVF, and the proportion of participants who generated semantic clusters in the PVF aligned more with phoneme category size than with the order of the presentation of the trials. This suggests that more than the order of the tasks, the overall number of words available in a category can influence how semantic associations are activated in a PVF task. However, randomizing the order of VF tasks in future studies to minimize potential semantic priming from the SVF trial in the subsequent PVF trials can solidify the method further. This suggestion is also noted in "The Instruction Manual for Administration and Scoring Verbal Fluency Tasks" (Study I, Appendix A).

Both task-congruent and task-discrepant clusters were evident in the data. However, the proposed analysis method only accounted for switches in the task-congruent clusters. Generating clear and unambiguous instructions for analysis that encompass both cluster types proved to be too error-prone to be reliably included in the suggested analysis method. In the future, a combination of the two clustering analyses would be ideal in determining the most reliable method for switching analysis.

6.4.2 Applying the Proposed Method to the Immersed Bilingual Group and the Control Group (Study II and Study III)

The analysis methods proposed in Study I were not fully replicated in Studies II and III. Due to the limited amount of data (immersed bilinguals $N = 38$, control group $N = 50$), the analysis of task-discrepant clusters was excluded from Studies II and III. Therefore, potential differences between languages and groups in automatic activation of semantic strategies throughout PVF (Sung et al., 2013) or in reaching out to hierarchical semantic memory in later stages of an effortful PVF task (Abwender et al., 2001) remain to be investigated in future studies.

Differences in the number of errors between languages and groups were investigated using simple statistical analysis due to the scarcity of the data. In future studies, more thorough statistical analyses on the impact of extralinguistic variables on the number and type of errors between L1 and L2 within the immersed bilingual group as well as between immersed bilinguals and a control group might provide insight into the impact extralinguistic variables have on task performance in a first language attrition study setting. Additionally, investigating the temporal distribution of errors could provide valuable perspective into the performance differences in the initial stage of the tasks. Applying these investigations to larger datasets derived from various immersed bilingual populations as well as general bilingual populations can expand findings of this dissertation.

In this dissertation PVF task prompts were selected based on high-frequency dictionary approach for Finnish and a well-established F, A, S letter selection for English. Adhering to a similar selection criterion for PVF tasks in both languages could have improved the reliability of the comparison between PVF task performance in L1 and L2. However, these choices were made to ensure optimal comparability with previous studies and future investigations. Choosing letter prompts based on a high-frequency approach in a smaller language like Finnish allows for comparisons not only within that language, but also with studies conducted in other languages that use the same approach for letter category prompt selection. For English, the F, A, S, letter combination is widely accepted, enabling comparisons with a vast amount of existing literature.

When interpreting the findings from this study, it is important to consider the potential impact of scarce shared cognates on language interference. This study was conducted with a specific pair of languages that share very few cognates (Finnish and English). There may be less competition between high-frequency words in Finnish and English compared with language pairs with a larger shared vocabulary, potentially limiting the applicability of the findings to language pairs with larger shared vocabularies (Schmid & Jarvis, 2014). Therefore, future studies investigating the similarities in lexical processing in immersed L1 Finnish bilingual groups in other L2 environments, and bilingual groups with varied backgrounds can further substantiate the findings from this dissertation.

6.4.3 Impact of Extralinguistic Variables

Limitations of this study include relatively narrow investigations on extralinguistic variables as potential factors impacting VF performance compared to many studies on first language attrition. Although participants provided numerous self-reports on varied aspects of language background, use, attitudes, and affiliations both before and after emigration in the sociolinguistic questionnaire, preliminary explorations following the LATB protocol revealed that it was not possible to reliably consolidate data from those variables for the purposes of this dissertation. Guided by earlier research (e.g., Schmid & Dusseldorp, 2010), this study aimed to investigate the impact of L1 use for professional purposes on VF performance. Unfortunately, the dataset did not allow for such analysis as only three participants in the immersed bilingual group reported using L1 for professional purposes. Therefore, two variables were selected to represent the shared language history of the attrition population: time spent in the second language environment (LoR) and self-reported frequency of overall L1 use measured on a 5-point Likert scale.

The decision to limit the analyses to LoR and frequency of L1 use was also based on the focus of the study. The primary aim was to investigate VF task performance, and the strategies participants employ for optimal performance during these tasks by applying multiple analyses on fine-grained data. To achieve reliable and parsimonious statistical models, the number and complexity of the extralinguistic variables was restricted. The findings from these analyses suggested that the included extralinguistic variables did have an impact on task performance, and future studies including a more extensive analysis of the impact of extralinguistic variables on VF performance are therefore warranted.

However, selecting and determining extralinguistic variables is complex. Reports on language history, attitudes, and affiliations typically involve subjective self-reports spanning several years of varying external circumstances that dictate the use of language in everyday life. Thus, comparability between individuals is not straightforward and self-reports can be considered the best available approximation of past behavior (Bylund & Ramírez-Galan, 2016; Köpke & Schmid, 2004). Variables measuring language use are also typically interconnected in many ways. For example, while the time elapsed since leaving the L1 country is measurable (LoR), assessing the level of deprivation of L1 during this time is more challenging, especially in datasets dating to modern times (Schmid, 2019).

Despite these limitations, exploring the impact of shared language history in language attrition populations is valuable also beyond research interests. For example, participants in language attrition studies often consider it natural that the time spent away from the native country has an impact on the changes they might experience in their L1 (e.g., Lazaridou-Chatzigoga & Karatsareas, 2022). Accordingly, in this study, many participants in the immersed bilingual group expressed concerns about their L1 abilities during interviews, specifically citing the long time since their emigration. Thus, the projected and confirmed results of LoR not appearing as a significant predictor on any of the analysis separately, can motivate and support participants of this study in preserving their native language and cultural identity.

Besides the extralinguistic variables related to language history, a key distinguishing factor between studies focusing on language attrition and general bilingualism studies lies in the age of the participants. Language attrition studies usually involve older adults who emigrated from their native country after puberty, with several years or decades having passed since emigration. In contrast, most studies on bilingualism involve young adults, often students. It has been shown that VF task total scores tend to decline with age, especially in SVF tasks in bilingual populations as discussed in Section 2.2.1. Thus, it is important to consider the potential impact age can have on performance when comparing findings from studies that focus on general bilingual populations to findings from language attrition studies.

For this dissertation, participants in the control group were closely matched to immersed bilinguals and no significant age difference was detected between the groups. In Study I, no effect for age was detected in the group residing in Finland, likely due to the relatively narrow age range (Ardila, 2020). To achieve as parsimonious and informative models as possible, matched groups were considered to effectively control for age in group comparisons.

6.4.4 Participants and Data

The data for this dissertation consists of one lexical task collected from two groups of participants. As for the participants, limitations include absence of screening for cognitive health, and limited information on the language background of the control group. Including a formal assessment of Finnish language proficiency could have further strengthened the study by providing more precise guidelines for participant inclusion and profiling within the immersed bilingual group. For both participant groups, information on cognitive health was collected through a self-reporting questionnaire. In future studies, standardized methods for cognitive screening should be incorporated for reliable sampling.

Since the focus of this dissertation is on the immersed bilingual group, the language history of the control group was screened based on their use of Finnish as a primary language and their permanent residence in Finland. All participants in the control group identified as monolinguals, except for one who identified as a Finnish-Swedish bilingual. No participants in the control group reported being Finnish-English bilinguals, and they were classified as "monolinguals" based on their monolingual identity and active language use to contrast the immersed bilingual group in Studies II and III. However, as the Finnish education system exposes all students to at least two languages during their primary education, and many participants were highly educated, it is reasonable to assume their background was not purely monolingual. Therefore, more detailed information about their language backgrounds is included in Section 4.1, and more precise terminology in Studies II and III would have been warranted.

In a similar vein, participants included in the immersed bilingual group and referred to as "language attriters" in Studies II and III were asked to self-evaluate if they have experienced changes in their L1 proficiency. The sociolinguistic questionnaire included a self-assessment on Finnish language proficiency "before moving to the USA" and "at present" on a 5-point Likert scale as described in Section 4.1. Out of the 38 participants, 32 indicated that their L1 proficiency had changed for the worse during the time elapsed from emigration. This self-report could potentially narrow participants down to "group experiencing attrition". However, as discussed earlier in the Section 6.4.3 self-assessments dating back to long periods of time, in this case even decades, are best subjective approximations of past experiences. Thus, these self-evaluations were not considered reliable criteria for participant exclusion.

Following existing literature, the inclusion criteria for the immersed bilingual group was set to individuals who had emigrated to an L2 environment "at an age when L1 was considered well-established". Literature generally suggests that a reasonable cut-off age for a language to be considered well-established is around puberty, specifically about 12 years of age (e.g., Schmid, 2011). During data collection, two participants reported that they had emigrated at an earlier age, namely at the ages of 9 and 11. These participants were not excluded from the study. The decision to include them in the immersed bilingual group, despite emigrating before age 12, was based on an assessment of their language development and experiences through interviews. This decision was also based on the challenge of recruiting a sufficient sample of participants in the immersed bilingual group.

The scope of this dissertation was limited to the lexical level with a focus on one specific task type, verbal fluency. For a more comprehensive view of lexical access and abilities, it is recommended to use diverse tasks for data collection, including free speech samples (Schmid, 2011b; Schmid & Jarvis, 2014). Furthermore, to thoroughly investigate language attrition within a specific group, it is beneficial to examine language performance and use across multiple language domains. While these wider investigations were beyond the scope of this dissertation, data collected during this research project includes free and controlled speech samples, and morphological tasks focused on the Finnish case system. These tasks are detailed in Appendix 1. The data are deposited and made available for future research via the Language Bank of Finland, with the prospect of enabling further investigations into L1 Finnish and L2 English spoken in California, US, in contrast to Finnish spoken in Finland to determine broader group specific characteristics in the participant groups investigated in this dissertation.

The free and controlled speech samples included in data collection provide ample opportunities for further research into lexical accuracy, diversity, and fluency within the participant groups studied in this dissertation. Additionally, the morphologically rich nature of Finnish (Helasvuo, 2008) offers a unique context for studying language attrition at a structural level. Previous studies on the morphology of Finnish spoken by Finnish emigrants (Hirvonen, 1995, 1998; Jönsson-Korhola, 1993; Larmouth, 1974; Leppänen, 2004) suggest a general simplification and variation in the spontaneous use of the case system among Finnish emigrants. Free and controlled speech samples and morphological tasks targeting the case assignment system provide various opportunities for future studies.

6.5 Major Contributions

The primary contributions of this dissertation are twofold. Firstly, methodological guidelines for a comprehensive and systematic VF task analysis introduced in Study I and applied in Studies II and III demonstrate that these commonly used, but often under-analyzed, tasks can serve as a valuable tool for a detailed analysis of lexical retrieval. Guidelines for VF task data elicitation and analyses are detailed in Study I, Appendix A, and are projected to support future studies across diverse clinical and academic settings.

Secondly, this dissertation deepens understanding of the processes underlying lexical retrieval in a group of immersed bilinguals in a typical first language attrition setting. Studies II and III provide a nuanced perspective on the VF task performance among immersed bilinguals, probing lexical accuracy, and fluency. Immersed bilinguals demonstrated strong performance in L1 contrasting the perceived language dominance of balanced bilingualism with a slight preference for L2. A higher number of errors and fewer error-free trials in in L1 the group of immersed bilinguals compared to a control group of L1 Finnish speakers residing in Finland were interpreted to indicate speaker insecurities and reduced accuracy in L1, which are often labelled as general characteristics of language attrition (Jarvis, 2019; Schmid & Köpke, 2009). As such they potentially influenced the self-reported preference for L2 in the group of immersed bilinguals.

Immersed bilinguals employed similar language processing strategies in L1 and L2 and more frequent L1 use resulted in higher total scores but slowed down performance in both languages. These findings underscore the importance of including both languages in studies that focus on language attrition to enable evaluation of the overall effect extralinguistic variables have on language task performance.

Findings from investigations comparing immersed bilinguals who have experienced reduced exposure and use of their L1 while being immersed in a second language environment, and a control group of L1 speakers living in Finland, suggest potential cognitive adaptations in the group of immersed bilinguals. Immersed bilinguals appeared to use rapid lexical retrieval more efficiently than the control group, even though frequent use of L1 seemed to slow down their performance. The clustering and switching analyses implied that immersed bilinguals might be able to utilize their clustering strategy more efficiently than the control group. However, once immersed bilinguals have depleted a sub-category (cluster), they might be less able to shift back and forth between sub-categories during the VF task. These strategies resulted in lower total scores in SVF and similar performance in the PVF in the immersed bilingual group compared to the control group.

Findings from the clustering and switching analysis partly align with previous literature on bilinguals with an emigrant background but contrast those from bilinguals without a sudden change in language environment. As such, they hint at the potential of clustering and switching analysis as a tool for differentiating between diverse bilingual populations.

6.6 Final Conclusions

This dissertation demonstrated that a comprehensive analysis of VF tasks can yield detailed and nuanced insights into lexical retrieval. First, guidelines for in-depth VF task analyses were generated, verified and compiled into a practical manual for future use in varied clinical populations. Second, these analysis methods were applied to data collected from a group of Finnish-English bilinguals with prolonged immersion in an L2 environment with the aim of identifying differences and similarities in L1 and L2 lexical processing strategies. Third, performance in L1 VF tasks was compared between the immersed bilingual group and their counterparts of L1 Finnish speakers living in Finland, replicating the approach of numerous language attrition studies across various language pairs.

The limitations of this study have been discussed previously. Despite these limitations, this dissertation provides a valuable starting point for future studies investigating VF task performance in varied clinical groups and offers insights into the lexical processing strategies of an immersed bilingual group laying the groundwork for future research in this domain.

Findings from this dissertation underscore the importance of including both L1 and L2 in studies fitting the language attrition framework. This approach showed that while immersed bilinguals demonstrated speaker insecurities in their L1 performance, the strategies they employed for optimal performance in VF tasks were very similar in both languages, resulting in higher total scores in L1 than L2. In addition, the frequency of using their L1 in everyday life had a similar impact across languages. These findings demonstrated that the observed performance was not language-specific but reflected a holistic adaptation of language processing strategies and impact of frequent L1 use in the group of immersed bilinguals. This dissertation also revealed nuanced cognitive adaptations in the immersed bilingual group compared to a control group of L1 speakers living in Finland, particularly in the efficacy of rapid lexical retrieval and the use of clustering and switching strategies. Thus, the proposed method could potentially identify typical group preferences in optimal VF task processing strategies among immersed bilingual populations within the dynamic continuum of bilingualism. In addition to studies focusing on bilingual populations, guidelines for VF task data elicitation and analyses detailed in this dissertation are projected to support future research across diverse clinical populations and academic settings.

Abbreviations

VF	Verbal Fluency
SVF	Semantic Verbal Fluency
PVF	Phonemic Verbal Fluency
L1	Language 1 (native language)
L2	Language 2 (second language, the environmental language)
LoR	Length of Residence (in the second language environment)
LATB	The Language Attrition Test Battery

List of References

Abwender, D. A., Swan, J. G., Bowerman, J. T., & Connolly, S. W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment*, *8*(3), 323–338. https://doi.org/10.1177/107319110100800308

Aitchison, J. (2013). *Language change: Progress or decay?* (4th ed). Cambridge University Press.

Ammerlaan, T. (1996). *"You get a bit wobbly..." – Exploring bilingual lexical retrieval processes in the context of first language attrition.* [Unpublished doctoral dissertation, Katholieke Universiteit]. https://l1attrition.files.wordpress.com/2019/06/ammerlaan-1996.pdf

Andersen, R. W. (1982). Determining the linguistic attributes of language attrition. In R. D. Lambert & B. F. Freed (Eds.), *The Loss of Language Skills.* (pp. 83–118). Newbury House Publishers.

Ardila, A. (2020). A cross-linguistic comparison of category verbal fluency test (ANIMALS): A systematic review. *Archives of Clinical Neuropsychology*, *35*(2), 213–225. https://doi.org/10.1093/arclin/acz060

Badstübner, T. (2011). *L1 attrition: German immigrants in the U.S.* [Doctoral Dissertation, The University of Arizona]. https://repository.arizona.edu/handle/10150/145569

Barry, D., Bates, M. E., & Labouvie, E. (2008). FAS and CFL forms of verbal fluency differ in difficulty: A meta-analytic study. *Applied Neuropsychology*, *15*(2), 97–106. https://doi.org/10.1080/09084280802083863

Becker, N., & Salles, J. F. de. (2016). Methodological criteria for scoring clustering and switching in verbal fluency tasks. *Psico-USF*, *21*(3), 445–457. https://doi.org/10.1590/1413-82712016210301

Borkowski, J. G., Benton, A. L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia*, *5*(2), 135–140. https://doi.org/10.1016/0028-3932(67)90015-2

Bylund, E., & Ramírez-Galan, P. (2016). Language aptitude in first language attrition: A study on late Spanish-Swedish bilinguals. *Applied Linguistics*, *37*(5), 621–638. https://doi.org/10.1093/applin/amu055

Cavaco, S., Gonçalves, A., Pinto, C., Almeida, E., Gomes, F., Moreira, I., Fernandes, J., & Teixeira-Pinto, A. (2013). Semantic fluency and phonemic fluency: Regression-based norms for the portuguese population. *Archives of Clinical Neuropsychology*, *28*(3), 262–271. https://doi.org/10.1093/arclin/act001

Crowe, S. F. (1998). Decrease in performance on the verbal fluency test as a function of time: Evaluation in a young healthy sample. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 391–401. https://doi.org/10.1076/jcen.20.3.391.810

de Bot, K., & Clyne, M. (1994). A 16-year longitudinal study of language attrition in Dutch immigrants in Australia. *Journal of Multilingual and Multicultural Development*, *15*(1), 17–28. https://doi.org/10.1080/01434632.1994.9994554

De Leeuw, E., Schmid, M. S., & Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition*, *13*(1), 33–40. https://doi.org/10.1017/S1366728909990289

Dietrich, S., & Hernandez, E. (2022). *Language use in the United States: 2019* (American Community Survey Reports ACS-50). https://www.census.gov/content/dam/Census/library/publications/2022/acs/acs-50.pdf

Dostert, S. C. (2009). *Multilingualism, L1 attrition and the concept of 'native speaker'* [Doctoral Dissertation]. Heinrich Heine Universitat Düsseldorf.

European Commission. (2012). *Europeans and their languages* (Special Eurobarometer 386). Directorate-General for Communication (European Parliament), Directorate-General for Education, Youth, Sport and Culture. https://op.europa.eu/s/zJXc

Fernaeus, S.-E., & Almkvist, O. (1998). Word production: Dissociation of two retrieval modes of semantic memory across time. *Journal of Clinical and Experimental Neuropsychology*, *20*(2), 137–143. https://doi.org/10.1076/jcen.20.2.137.1170

Fernaeus, S.-E., Östberg, P., Hellström, Å., & Wahlund, L.-O. (2008). Cut the coda: Early fluency intervals predict diagnoses. *Cortex*, *44*(2), 161–169. https://doi.org/10.1016/j.cortex.2006.04.002

Friesen, D. C., Luo, L., Luk, G., & Bialystok, E. (2015). Proficiency and control in verbal fluency performance across the lifespan for monolinguals and bilinguals. *Language, Cognition and Neuroscience*, *30*(3), 238–250. https://doi.org/10.1080/23273798.2014.918630

Gallo, F., Ramanujan, K., Shtyrov, Y., & Myachykov, A. (2021). Attriters and bilinguals: What's in a name? *Frontiers in Psychology*, *12*, 558228. https://doi.org/10.3389/fpsyg.2021.558228

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, *16*(4), 562–576. https://doi.org/10.1037/0894-4105.16.4.562

Gollan, T. H., Sandoval, T., & Salmon, D. P. (2011). Cross-language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological Science*, *22*(9), 1155–1164. https://doi.org/10.1177/0956797611417002

Goral, M. (2004). First-language decline in healthy aging: Implications for attrition in bilingualism. *Journal of Neurolinguistics*, *17*(1), 31–52. https://doi.org/10.1016/S0911-6044(03)00052-6

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*(1), 3–15. https://doi.org/10.1016/0093-934X(89)90048-5

Grosjean, F. (2010). *Bilingual: Life and reality*. Harvard University Press. https://doi.org/10.4159/9780674056459

Grosjean, F. (2013). Bilingualism: A short introduction. In F. Grosjean & P. Li (Eds.), *The Psycholinguistics of Bilingualism* (Vol. 5, pp. 5–25). John Wiley & Sons. https://www.francoisgrosjean.ch/bilin_bicult/25%20Grosjean.pdf

Grosjean, F. (2021). *Life as a bilingual: Knowing and using two or more languages* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108975490

Gurunandan, K., Carreiras, M., & Paz-Alonso, P. M. (2022). Verbal production dynamics and plasticity: Functional contributions of language and executive control systems. *Cerebral Cortex*, 740–753. https://doi.org/10.1093/cercor/bhac097

Helasvuo, M.-L. (2008). Aspects of the structure of Finnish. In *Research in Logopedics, Speech and Language Therapy in Finland. Communication Disorders Accross Languages.* (pp. 9–18). Multilingual Matters.

Hirvonen, P. (1995). Phonological and morphological aspects of Finnish language attrition in the United States. *Proceedings of the International Congress of Dialectogists, Vol. 4.*, *4*, 181–193.

Hirvonen, P. (1998). The Finnish-American language shift. In J. Niemi, T. Odlin, & J. Heikkinen (Eds.), *Language Contact, Variation and Change.* (pp. 136–150). Joensuun yliopiston monistuskeskus. https://doi.org/10.1017/S0272263100304065

Hirvonen, P., & Lauttamus, T. (2000). Code-switching and language attrition: Evidence from American Finnish Interview Speech. *SKY Journal of Linguistics*, *13*, 47–74.

International Organization for Migration. (2024). *World Migration Report 2024*. United Nations.

Jarvis, S. (2019). Lexical attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 240–250). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.20

Johns, B. T., Taler, V., Pisoni, D. B., Farlow, M. R., Hake, A. M., Kareken, D. A., Unverzagt, F. W., & Jones, M. N. (2018). Cognitive modeling as an interface between brain and behavior: Measuring the semantic decline in mild cognitive impairment. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *72*(2), 117–126. https://doi.org/10.1037/cep0000132

Jönssön-Korhola, H. (1993). Lauserakenteesta toisen ja kolmannen polven kielenkäytössä. In P. Virtaranta, H. Jönssön-Korhola, M. Martin, & M. Kainulainen (Eds.), *Amerikansuomi*. Suomalaisen Kirjallisuuden Seura.

Kasparian, K., & Steinhauer, K. (2017). On missed opportunities and convenient "truths." *Linguistic Approaches to Bilingualism*, *7*(6), 709–714. https://doi.org/10.1075/lab.00010.kas

Kim, N., Kim, J.-H., Wolters, M. K., MacPherson, S. E., & Park, J. C. (2019). Automatic scoring of semantic fluency. *Frontiers in Psychology*, *10*, 1020. https://doi.org/10.3389/fpsyg.2019.01020

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Köpke, B., & Schmid, M. S. (2004). Language attrition: The next phase. In M. S. Schmid, B. Köpke, M. Keijzer, & L. Weilemar (Eds.), *Studies in Bilingualism* (Vol. 28, p. 1). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.28.02kop

Kostiainen, A. (2014). Interest in the history of Finnish Americans. In *Finns in the United States: A History of Settlement, Dissent, and Integration* (p. 353). Michigan State University Press. https://ebookcentral.proquest.com/lib/kutu/reader.action?docID=3338350&ppg=1

Laine, M., & Lehtonen, M. (2018). Cognitive consequences of bilingualism: Where to go from here? *Language, Cognition and Neuroscience*, *33*(9), 1205–1212. https://doi.org/10.1080/23273798.2018.1462498

Lanting, S., Haugrud, N., & Crossley, M. (2009). The effect of age and sex on clustering and switching during speeded verbal fluency tasks. *Journal of the International Neuropsychological Society*, *15*(2), 196–204. https://doi.org/10.1017/S1355617709090237

Larmouth, D. W. (1974). Differential interference in American Finnish cases. *Language*, *50*(2), 356–366.

Lazaridou-Chatzigoga, D., & Karatsareas, P. (2022). Language Attrition and Lived Experiences of Attrition among Greek Speakers in London. *Languages*, *7*(4), 307. https://doi.org/10.3390/languages7040307

Leinonen, J. (2014). One culture, two cultures? Families of Finns in the United States in the twentieth century. In A. Kostiainen (Ed.), *Finns in the United States: A history of settlement, dissent, and integration* (pp. 285–296). Michigan State University Press.

Leppänen, N. (2004). *What happened to Finnish? First language attrition in a second language environment—Case assignment in the Finnish spoken in the USA* [Unpublished Master's Thesis]. University of Oulu, Finland.

Library of Congress. (2024). Nordic and Scandinavian emigration to the United States [Research Guide]. *Finland Emigration*. https://guides.loc.gov/nordic-scandinavian-emigration/finland#s-lib-ctab-28975397-2

Linck, J. A., & Kroll, J. F. (2019). Memory retrieval and language attrition: Language loss or manifestations of a dynamic system? In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 87–97). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.8

Ljungberg, J. K., Hansson, P., Andrés, P., Josefsson, M., & Nilsson, L.-G. (2013). A longitudinal study of memory advantages in bilinguals. *PLoS ONE*, *8*(9), e73029. https://doi.org/10.1371/journal.pone.0073029

Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition*, *114*(1), 29–41. https://doi.org/10.1016/j.cognition.2009.08.014

Mardani, N., Pourjafari, M., Irandegani, M. A., Ahmadi, N., & Baghban, K. (2020). The effect of bilingualism on the processing of clustering and switching in verbal fluency tasks. *Journal of Rehabilitation Sciences & Research*, *7*(3), 114–117. https://doi.org/10.30476/jrsr.2020.86911.1096

Marsh, J. E., Hansson, P., Sörman, D. E., & Ljungberg, J. K. (2019). Executive processes underpin the bilingual advantage on phonemic fluency: Evidence from analyses of switching and clustering. *Frontiers in Psychology*, *10*, 1355. https://doi.org/10.3389/fpsyg.2019.01355

Martin, M., & Jönsson-Korhola, H. (1993). Amerikansuomalaiset ja heidän kieliolonsa. In P. Virtaranta, H. Jönsson-Korhola, M. Martin, & M. Kainulainen, *Amerikansuomi* (pp. 11–20). Suomalaisen Kirjallisuuden Seura.

Mehotcheva, T. H., & Köpke, B. (2019). Introduction to L2 attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 329–348). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.27

Mikkonen, N. (2024). *Tervetuloa, tervemenoa. Suomi muuttoliikkeessä.* Siirtolaisuusinstituutti. Institute of Migration. https://www.doria.fi/handle/10024/188472

Oberg, G., & Ramírez, M. (2006). Cross-linguistic meta-analysis of phonological fluency: Normal performance across cultures. *International Journal of Psychology*, *41*(5), 342–347. https://doi.org/10.1080/00207590500345872

Oikeusministeriö. (n.d.). *Kielelliset oikeudet*. Oikeusministeriö. Retrieved November 9, 2024, from https://oikeusministerio.fi/kielelliset-oikeudet

Olabarrieta-Landa, L., Torre, E. L., López-Mugartza, J. C., Bialystok, E., & Arango-Lasprilla, J. C. (2017). Verbal fluency tests: Developing a new model of administration and scoring for Spanish language. *NeuroRehabilitation*, *41*(2), 539–565. https://doi.org/10.3233/NRE-162102

Opetushallitus. (n.d.). *Toinen kotimainen kieli ja vieraat kielet perusopetuksessa* [Government]. Retrieved November 4, 2024, from https://www.oph.fi/fi/koulutus-ja-tutkinnot/perusopetus/toinen-kotimainen-kieli-ja-vieraat-kielet-perusopetuksessa

Opitz, C. (2011). *First language attrition and second language acquisition in a second language environment* [Doctoral Dissertation, University of Dublin, Trinity College]. http://hdl.handle.net/2262/78154

Patra, A., Bose, A., & Marinis, T. (2020). Performance difference in verbal fluency in bilingual and monolingual speakers. *Bilingualism: Language and Cognition*, *23*(1), 204–218. https://doi.org/10.1017/S1366728918001098

Pekkala, S., Goral, M., Hyun, J., Obler, L. K., Erkinjuntti, T., & Albert, M. L. (2009). Semantic verbal fluency in two contrasting languages. *Clinical Linguistics & Phonetics*, *23*(6), 431–445. https://doi.org/10.1080/02699200902839800

Perdue, C. & European Science Foundation (Eds.). (1993). *Adult language acquisition: Cross-linguistic perspectives*. Cambridge University Press.

Pereira, A. H., Gonçalves, A. B., Holz, M., Gonçalves, H. A., Kochhann, R., Joanette, Y., Zimmermann, N., & Fonseca, R. P. (2018). Influence of age and education on the processing of clustering and switching in verbal fluency tasks. *Dementia & Neuropsychologia*, *12*(4), 360–367. https://doi.org/10.1590/1980-57642018dn12-040004

Perusopetuslaki (Basic Education Act), Pub. L. No. 628/1998 (1998). https://www.finlex.fi/fi/laki/ajantasa/1998/19980628

Pierce, L. J., Genesee, F., & Klein, D. (2019). Language loss and language learning in internationally adopted children: Evidence from behaviour and the brain. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 469–480). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.37

Quaranta, D., Caprara, A., Piccininni, C., Vita, M. G., Gainotti, G., & Marra, C. (2016). Standardization, clinical validation, and typicality norms of a new test assessing semantic verbal fluency. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *31*(5), 434–445. https://doi.org/10.1093/arclin/acw034

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Roberts, P. M., & Dorze, G. L. (1997). Semantic organization, strategy use, and productivity in bilingual semantic verbal fluency. *Brain and Language*, *59*(3), 412–449. https://doi.org/10.1006/brln.1997.1753

Romaine, S. (1995). *Bilingualism* (2nd ed). Blackwell.

Ross, T. P. (2003). The reliability of cluster and switch scores for the controlled oral word association test. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *18*(2), 153–164.

Rosselli, M., Ardila, A., Araujo, K., Weekes, V. A., Caracciolo, V., Padilla, M., & Ostrosky-Solí, F. (2000). Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. *Applied Neuropsychology*, *7*(1), 17–24. https://doi.org/10.1207/S15324826AN0701_3

Rosselli, M., Ardila, A., Salvatierra, J., Marquez, M., Luis, M., & Weekes, V. A. (2002). A cross-linguistic comparison of verbal fluency tests. *International Journal of Neuroscience*, *112*(6), 759–776. https://doi.org/10.1080/00207450290025752

Runnqvist, E., & Costa, A. (2012). Is retrieval-induced forgetting behind the bilingual disadvantage in word production? *Bilingualism: Language and Cognition*, *15*(2), 365–377. https://doi.org/10.1017/S1366728911000034

Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition*, *13*(2), 231–252. https://doi.org/10.1017/S1366728909990514

Santos Nogueira, D., Azevedo Reis, E., & Vieira, A. (2016). Verbal fluency tasks: Effects of age, gender, and education. *Folia Phoniatrica et Logopaedica*, *68*(3), 124–133. https://doi.org/10.1159/000450640

Scheuringer, A., Wittig, R., & Pletzer, B. (2017). Sex differences in verbal fluency: The role of strategies and instructions. *Cognitive Processing*, *18*(4), 407–417. https://doi.org/10.1007/s10339-017-0801-1

Schmid, M. S. (2011a). Contact x time: External factors and variability in L1 attrition. In M. S. Schmid & W. Lowie (Eds.), *Studies in Bilingualism* (Vol. 43, pp. 155–176). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.43.11sch

Schmid, M. S. (2011b). *Language attrition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511852046

Schmid, M. S. (2013). First language attrition. *WIREs Cognitive Science*, *4*(2), 117–123. https://doi.org/10.1002/wcs.1218

Schmid, M. S. (2019). The impact of frequency of use and length of residence on L1 attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 287–303). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.24

Schmid, M. S., & Cherciov, M. (2019). Introduction to extralinguistic factors in language attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 265–276). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.22

Schmid, M. S., & Dusseldorp, E. (2010). Quantitative analyses in a multivariate study of language attrition: The impact of extralinguistic factors. *Second Language Research*, *26*(1), 125–160. https://doi.org/10.1177/0267658309337641

Schmid, M. S., & Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism: Language and Cognition*, *17*(4), 729–748. https://doi.org/10.1017/S1366728913000771

Schmid, M. S., & Keijzer, M. (2009). First language attrition and reversion among older migrants. *International Journal of the Sociology of Language*, *2009*(200), 83–101. https://doi.org/10.1515/IJSL.2009.046

Schmid, M. S., & Köpke, B. (2009). L1 attrition and the mental lexicon. In A. Pavlenko (Ed.), *The bilingual mental lexicon: Interdisciplinary approaches* (p. pp.209-238). Multilingual Matters.

Schmid, M. S., & Köpke, B. (2017a). The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism*, *7*(6), 637–667. https://doi.org/10.1075/lab.17058.sch

Schmid, M. S., & Köpke, B. (2017b). When is a bilingual an attriter? Response to the commentaries. *Linguistic Approaches to Bilingualism*, *7*(6), 763–770. https://doi.org/10.1075/lab.17059.sch

Schmid, M. S., & Köpke, B. (2019). Introduction. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. xxviii–4). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.1

Schmid, M. S., & Yılmaz, G. (2018). Predictors of language dominance: An integrated analysis of first language attrition and second language acquisition in late bilinguals. *Frontiers in Psychology*, *9*, 1306. https://doi.org/10.3389/fpsyg.2018.01306

Schmitt, E., & Sorokina, A. (2024). Guest editorial: Language attrition – A comprehensive introduction. *The Language Learning Journal*, *52*(2), 133–144. https://doi.org/10.1080/09571736.2024.2307641

Segalowitz, N. (1991). Does advanced skill in a second language reduce automaticity in the first language? *Language Learning*, *41*(1), 59–83. https://doi.org/10.1111/j.1467-1770.1991.tb00676.x

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00772

Sharwood Smith, M. S., & Buren, P. V. (1991). First language attrition and the parameter setting model. In H. W. Seliger & R. M. Vago (Eds.), *First Language Attrition* (1st ed., pp. 17–30). Cambridge University Press. https://doi.org/10.1017/CBO9780511620720.002

Siirtolaisuusinstituutti. (2024, August 7). *Emigration*. Siirtolaisuusinstituutti. https://siirtolaisuusinstituutti.fi/en/research/emigration/

Sorokina, A., & Mugno, R. (2024). Embracing bilingualism: L1 retention amidst L2 acquisition. *The Language Learning Journal*, *52*(2), 186–198. https://doi.org/10.1080/09571736.2024.2309184

Strauss, E., Sherman, E. M. S., Spreen, O., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed). Oxford University Press.

Sung, K., Gordon, B., Yang, S., & Schretlen, D. J. (2013). Evidence of semantic clustering in letter-cued word retrieval. *Journal of Clinical and Experimental Neuropsychology*, *35*(10), 1015–1023. https://doi.org/10.1080/13803395.2013.845141

Tallberg, I. M., Ivachova, E., Jones Tinghag, K., & Östberg, P. (2008). Swedish norms for word fluency tests: FAS, animals and verbs. *Scandinavian Journal of Psychology*, *49*(5), 479–485. https://doi.org/10.1111/j.1467-9450.2008.00653.x

Thiele, K., Quinting, J. M., & Stenneken, P. (2016). New ways to analyze word generation performance in brain injury: A systematic review and meta-analysis of additional performance measures. *Journal of Clinical and Experimental Neuropsychology*, *38*(7), 764–781. https://doi.org/10.1080/13803395.2016.1163327

Tilastokeskus. (2024, August 7). *Väestö ja yhteiskunta*. https://stat.fi/tup/suoluk/suoluk_vaesto.html#vaesto-kielen-mukaan

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *14*(2), 167–177.

Treffers-Daller, J. (2019). What defines language dominance in bilinguals? *Annual Review of Linguistics*, *5*(1), 375–393. https://doi.org/10.1146/annurev-linguistics-011817-045554

Tröger, J., Linz, N., König, A., Robert, P., Alexandersson, J., Peter, J., & Kray, J. (2019). Exploitation vs. Exploration—Computational temporal and semantic analysis explains semantic verbal fluency impairment in Alzheimer's disease. *Neuropsychologia*, *131*, 53–61. https://doi.org/10.1016/j.neuropsychologia.2019.05.007

Troyer, A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology*, *22*(3), 370–378. https://doi.org/10.1076/1380-3395(200006)22:3;1-V;FT370

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138

United States Census Bureau. (2022a). *People reporting ancestry. 2023: American community survey 1-Year estimated detailed tables.* (B04006) [Dataset]. https://data.census.gov/table/ACSDT1Y2022.B04006?q=Ancestry&t=Ancestry:Populations%20and%20People&moe=false

United States Census Bureau. (2022b). *Selected population profile in the United States. American community survey, ACS 1-year estimates selected population profiles, table s0201.* [Dataset]. https://data.census.gov/table/ACSSPP1Y2022.S0201?q=finnish.

Venegas, M. J., & Mansur, L. L. (2011). Verbal fluency: Effect of time on item generation. *Dementia & Neuropsychologia*, *5*(2), 104–107. https://doi.org/10.1590/S1980-57642011DN05020008

Vicente, S. G., Benito-Sánchez, I., Barbosa, F., Gaspar, N., Dores, A. R., Rivera, D., & Arango-Lasprilla, J. C. (2021). Normative data for verbal fluency and object naming tests in a sample of European Portuguese adult population. *Applied Neuropsychology: Adult*, 1–12. https://doi.org/10.1080/23279095.2020.1868472

Virtaranta, P. (1993). Amerikansuomen tutkimuksesta. In P. Virtaranta, H. Jönsson-Korhola, M. Martin, & Kainulainen (Eds.), *Amerikansuomi* (pp. 21–34). Suomalaisen Kirjallisuuden Seura.

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal fluency: Language or executive function measure? *Applied Neuropsychology: Adult*, *23*(1), 29–34. https://doi.org/10.1080/23279095.2015.1004574

Yilmaz, G. (2019). L1 attrition, L2 development, and integration. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 303–313). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.25

Yilmaz, G., & Schmid, M. S. (2012). L1 accessibility among Turkish-Dutch bilinguals. *The Mental Lexicon*, *7*(3), 249–274. https://doi.org/10.1075/ml.7.3.01yil

Appendices

Appendix 1. Description of Data Collected in the Research Project: Finnish Language Attrition in the USA

Description of Data Collected in the Research Project: Finnish Language Attrition in the USA

The contemporary Finnish language variant spoken in Northern California, USA, was collected and documented during the research project *Finnish Language Attrition in the USA* between 10/2017-11/2018. Participants were Finnish American emigrants (N = 38) who had lived in an English speaking environment for over 20 years. To contrast the participants and enable group comparisons to Finnish speakers living in Finland, a group of Finnish speakers living in Finland (N = 50) who were matched for age, education and gender were included.

The data described here is deposited and made available for future research through the Language Bank of Finland with the objective of preserving this unique variant of the Finnish language. The data can be accessed for future research following appropriate confidentiality measures determined for this research project. The data can be located via The Language Bank of Finland (www.kielipankki.fi) as "Amerikansuomalaisten puhujien puhuttu suomen kieli" identifier *amersuom*.

Sociolinguistic Questionnaire

Background information for the Finnish American emigrants was collected using the sociolinguistic questionnaire (SQ) included in the Language Attrition Test Battery (Schmid, 2011). The SQ comprises of 71 questions, covering numerous self-reports on different aspects of language history, use, attitudes, and affiliations. The complete SQ is available at www.languageattrition.org, including instructions for use and analysis. The author translated the SQ into Finnish, and participants had the choice to complete it in either Finnish or English. For the control group, an abridged version of the questionnaire was used to confirm the use of Finnish in everyday contexts. Examples of all questionnaires are deposited alongside data derived from the questionnaires. Data is deposited as a spreadsheet file.

Free Speech (Finnish and English)

To record speech samples, a silent movie clip (a 9-minute 57-second segment from Charlie Chaplin's Modern Times, 1936) was used to elicit narration following Perdue, 1993. This task was recorded in Finnish and English for the Finnish American group and in Finnish for the group living in Finland.

As part of their interview, participants in the Finnish American group were invited to share a short story about their immigration journey as a free speech sample in Finnish. Audio files of these speech samples are deposited from those participants who chose to share their story. All data for the free speech samples are deposited as audio files.

WUG Test (Finnish)

A WUG test was administered to assess the participants' ability to inflect words in context. The WUG test is a test designed to gauge the knowledge of morphological rules by asking participants to complete sentences using non-words following productive morphological rules (Berko, 1958). An existing Finnish morphology WUG test (Lyytinen, 2003) was used, and additional tasks were generated to cover all grammatical cases in Finnish (total of 15 cases). The test included 90 sentences in total and the sentences were presented in a pseudo-randomized order. Data for the WUG test is available in Finnish for the Finnish American participant group and the group of Finnish speakers living in Finland. Data is deposited as audio files and text files for the tasks.

References

Berko, J. (1958). The child's learning of English morphology. *WORD, 14*(2–3), 150–177. https://doi.org/10.1080/00437956.1958.11659661

Lehtinen, N., Luotonen, I., & Kautto, A. (2021). Systematic administration and analysis of verbal fluency tasks: Preliminary evidence for reliable exploration of processes underlying task performance. *Applied Neuropsychology: Adult, 30*(6), 727–739. https://doi.org/10.1080/23279095.2021.1973471

Lyytinen, P. (2003). Morfologiatesti. Taivutusmuotojen hallinnan mittausmenetelmä lapsille. [The Finnish morphology test. An evaluation method for the assessment of morphological inflections for children] 2nd. edition. Jyväskylän yliopistona lapsitutkimuskeskus ja Niilo Mäki Instituutti. Jyväskylän yliopisto.

Perdue, C. (1993). *Adult language acquisition: Cross-linguistic perspectives.* Cambridge University Press.

Schmid, M. S. (2011). *Language attrition.* Cambridge University Press. https://doi.org/10.1017/CBO9780511852046

Original Publications

Lehtinen, N., Luotonen, I., & Kautto, A. (2023)
Systematic administration and analysis of verbal fluency tasks: Preliminary evidence for reliable exploration of processes underlying task performance.
Applied Neuropsychology: Adult

I

APPLIED NEUROPSYCHOLOGY: ADULT
2023, VOL. 30, NO. 6, 727–739
https://doi.org/10.1080/23279095.2021.1973471

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Systematic administration and analysis of verbal fluency tasks: Preliminary evidence for reliable exploration of processes underlying task performance

Nana Lehtinen, Ida Luotonen, and Anna Kautto

Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland

ABSTRACT

Verbal fluency (VF) tasks are typically scored by the number of acceptable words generated within an allotted time (i.e., total score). However, total scores do not provide insight into verbal and executive processes underlying VF task performance. Further analyses have been implemented to increase the analytical power of VF tasks, but systematic scoring guidelines are needed. We generated instructions for administration, scoring, and analyses of total scores, errors, temporal parameters, clustering, and switching with strong inter-rater reliability. To investigate the reliability of the proposed analysis, we modeled the performance of Finnish-speaking older adults ($N = 50$) in phonemic (/k/, /a/, and /p/) and semantic (animals) categories. Our results are in line with previous studies: We observed a higher performance on semantic than phonemic fluency ($p \leq 0.001$, $d = 0.91$) and significant effects for education ($p \leq 0.001$, $d = 1.11$) and gender ($p \leq 0.001$, $d = -1.11$), but not for age ($p = 0.10$, $d = 0.48$). Most errors were repetitions. Performance declined over the allotted time frame as measured in 15-s segments (all *ps* < 0.001 with medium to large effect sizes). Task congruent clustering and switching were productive strategies (all *ps* < 0.001 with large effect sizes), and participants generated task discrepant clusters in both phonemic ($p = 0.004$, $d = 0.69$) and semantic tasks ($p = 0.66$, $d = 0.18$). The results substantiate the proposed method, providing evidence that these guidelines are a reliable starting point for VF task performance analyses in various clinical populations investigating VF task performance in depth.

KEYWORDS
Administration; clustering; error analysis; scoring; switching; temporal analysis; verbal fluency

Introduction

Verbal fluency (VF) tasks are widely used for clinical assessment and research purposes in multiple fields, such as speech pathology, neuropsychology, linguistics, and medicine (Strauss et al., 2006). In a VF task, the participant is asked to produce as many words as possible following a specific category in a specified time frame, often 60 s.

The most common VF task types are phonemic verbal fluency (PVF) and semantic verbal fluency (SVF). In the PVF, the participant is asked to produce words beginning with a specific phoneme or letter. This task is also referred to as phonemic fluency, Controlled Oral Word Associations (COWA), or the FAS test. In the semantic verbal fluency task (SVF), also referred to as semantic fluency or category fluency, the participant is asked to generate words belonging to a specific semantic category, such as “animals” or “clothes” (Strauss et al., 2006). Regardless of language context, VF task performance is considered to assess verbal knowledge and executive control. All VF tasks engage language processing, require maintaining focus on the task and selecting words that meet given criteria while inhibiting unsuitable candidates (Shao et al., 2014; Whiteside et al., 2016).

The Cattell–Horn–Carroll Theory of Cognitive Abilities (CHC) classifies word fluency (FW) as a major narrow ability in retrieval fluency (Gr) under the broad ability of long-term storage and retrieval (Glr) (Schneider & McGrew, 2018). The CHC does not differentiate semantic and phonemic fluency per se. However, there is evidence suggesting a distinction between semantic and phonological fluency within the framework (Jewsbury & Bowden, 2017), and multiple neurocognitive studies support this distinction. PVF is typically considered to engage strategic cognitive organization, initiation, inhibition, and maintenance of effort without the support of the hierarchical organization of semantic memory (Barry et al., 2008; Santos Nogueira et al., 2016; Strauss et al., 2006). SVF is considered to rely on a more automatic systematic semantic search based on semantic categorization, hierarchical mental lexicon, and memory organization resembling everyday use of language (e.g., generating a shopping list; Patra et al., 2020; Strauss et al., 2006).

VF task performance is most typically evaluated by the total score, calculated as the number of acceptable words

CONTACT Nana Lehtinen nana.lehtinen@utu.fi Department of Psychology and Speech-Language Pathology, University of Turku, Assistentinkatu 7, Turku 20500, Finland.

Supplemental data for this article can be accessed at publisher’s website.

generated in the given time frame (Strauss et al., 2006; Thiele et al., 2016) with normative data typically describing higher total scores for SVF than PVF tasks (Cavaco et al., 2013; Santos Nogueira et al., 2016; Strauss et al., 2006). While VF total scores can differentiate between healthy subjects and clinical groups, they also reflect verbal and executive processes underlying task performance. Evaluating total scores as the only metric does not provide insight into these processes, limiting the analytical and explanatory power of VF tasks (Becker & Salles, 2016; Johns et al., 2018; Oberg & Ramírez, 2006; Thiele et al., 2016). To investigate processes underlying VF performance, a growing body of literature is implementing additional analyses on temporal parameters, errors, and clustering and switching strategies using both traditional (Thiele et al., 2016) and computational approaches (Johns et al., 2018; Kim et al., 2019).

Norms for total scores and additional measures described above have been published in various languages, including effects of age, education, and gender (Ardila, 2020; Cavaco et al., 2013; Goral, 2004; Oberg & Ramírez, 2006; Olabarrieta-Landa et al., 2017; Quaranta et al., 2016; Santos Nogueira et al., 2016; Vicente et al., 2021). However, many studies fail to give accurate descriptions of administration, scoring, and analysis, making it challenging to compare and contrast studies and their outcomes (Olabarrieta-Landa et al., 2017; Thiele et al., 2016). A systematic approach to VF task analysis would increase the reliability and validity of studies implementing these simple to administer tasks across populations and languages (Becker & Salles, 2016; Thiele et al., 2016). In the following, we highlight aspects of VF task analyses implemented in previous literature and outline a suggestion for a systematic approach to administer, score and analyze phonemic and semantic verbal fluency tasks to increase the analytical and explanatory power of semantic and phonemic VF tasks.

Selecting categories for PVF and SVF tasks

In PVF, it is typical to include three trials. In English, FAS is the most common letter combination; other popular combinations are CFL and PWR. These combinations are selected from "easy letters" (Borkowski et al., 1967) and can, with some reservation, be used interchangeably (Ross, 2003; Strauss et al., 2006). Despite being intended initially for English speakers, a combination of FAS is used in studies conducted in other languages as well (for an overview, see Olabarrieta-Landa et al., 2017). However, it has been shown that letters with high frequency in the target language yield a higher number of words in VF tasks, and selecting language-specific categories for PVF is strongly recommended (Mardani et al., 2019; Oberg & Ramírez, 2006; Tombaugh et al., 1999). A short version of PVF consists of only one trial, often "B" for English (Harrison et al., 2000). As internal reliability between letters is high, it can be justified to reduce the number of trials. However, three phonemic trials are often preferred as it provides a more reliable measure of overall fluency ability (Oberg & Ramírez, 2006; Strauss et al., 2006; Tombaugh et al., 1999).

In SVF, a widely used category across languages is "animals" (Olabarrieta-Landa et al., 2017; Strauss et al., 2006). Other categories, such as "clothing," "vehicles," or "items in a supermarket," are also used. Multiple categories are applied to parallel the number of categories in PVF, but combining data from different semantic categories is more complex than combining multiple phonemic categories. Demographic influences and various cultural settings can influence semantic memory organization, and semantic category size and content can vary between populations (Abwender et al., 2001; Olabarrieta-Landa et al., 2017; Roberts & Dorze, 1997; Rosselli et al., 2002; Strauss et al., 2006; Troyer, 2000). As a semantic category, "animals" are culturally and linguistically relatively neutral, but as all categories, requires specific guidelines for scoring (e.g., how to score variations of the same animal; Olabarrieta-Landa et al., 2017; Pekkala et al., 2009; Roberts & Dorze, 1997). Category "animals" is also included in many neuropsychological test batteries, such as the Western Aphasia Battery (WAB; Kertesz, 1982) and the Consortium to Establish a Registry for Alzheimer's Disease (CERAD; Morris et al., 1989), making it a clinically appropriate choice for a semantic category.

Analyzing VF tasks

Performance in a verbal fluency task is typically analyzed by the total score (i.e., calculating the total number of acceptable words generated during the allotted time) (Strauss et al., 2006; Thiele et al., 2016). Higher total scores are generally associated with higher education, especially in PVF tasks (Oberg & Ramírez, 2006; Santos Nogueira et al., 2016; Tallberg et al., 2008; Tombaugh et al., 1999; Troyer, 2000), and total scores tend to decline with age, especially in SVF tasks (Ardila, 2020; Goral, 2004; Lanting et al., 2009; Santos Nogueira et al., 2016; Strauss et al., 2006; Tallberg et al., 2008; Tombaugh et al., 1999; Troyer, 2000). For gender, studies have shown no effect or a minor female advantage with emphasis on PVF (Scheuringer et al., 2017).

Temporal parameters of VF task performance are typically analyzed by the number of words generated during shorter time segments (10, 15, or 20-s segments) within the total time. Typically, participants produce most words in the early stages of the task using a semi-automatic rapid retrieval process. As time progresses, lexical retrieval becomes more effortful, with fewer and more infrequent words being generated toward the later segments of the task, demonstrating the function of time (Crowe, 1998; Fernaeus & Almkvist, 1998; Venegas & Mansur, 2011). Education has a positive effect in the early time segments in both PVF and SVF with no age effect found (Venegas & Mansur, 2011). In clinical populations, variation in the function of time has been shown to differentiate underlying mechanisms of VF performance in aphasia (Bose et al., 2017) and to have predictive power in Alzheimer's Disease diagnosis (Fernaeus et al., 2008; Venegas & Mansur, 2011). It has also been suggested that as most words are generated in early segments of the task, a shorter total time (30 s) could have enough power to differentiate healthy and patient populations (Fernaeus et al., 2008; Kim et al., 2011).

Error types in VF typically include repetitions, categorical errors, and non-items (Thiele et al., 2016). While errors are relatively scarce in normative data (Crowe, 1998; Gollan et al., 2011), the number and type of errors in varied populations carry value in regards to their research objectives, such as perseverations in Alzheimer's studies (Pekkala et al., 2008) and language intrusions in bilingualism studies (Gollan et al., 2011).

Clustering and switching are strategies needed for optimal fluency performance (Strauss et al., 2006; Thiele et al., 2016; Troyer, 2000; Troyer et al., 1997). Analysis of clustering and switching has been applied for multiple research objectives, such as differential diagnostics in neuropsychological populations (Johns et al., 2018; Thiele et al., 2016; Troyer, 2000; Troyer et al., 1997) and cross-linguistic fluency strategies in bilingualism studies (Roberts & Dorze, 1997; Rosselli et al., 2002). While many studies apply predetermined subcategories for clustering following Troyer et al. (1997), other methods for clustering and calculation switches are also frequently applied. Different methods of determining clusters and calculating switches can result in varied outcomes (Abwender et al., 2001; Thiele et al., 2016).

Clustering refers to the ability to produce words within subcategories (e.g., "words with two same initial phonemes" or "pets"), and it relies on phonemic analysis in PVF and semantic categorization and semantic memory in SVF (Strauss et al., 2006; Thiele et al., 2016; Troyer, 2000; Troyer et al., 1997). Task congruent clustering is a relatively automatic process with participants naturally using phonemic clustering in PVF and semantic clustering in SVF (Troyer et al., 1997). In addition, task discrepant clustering (semantic clustering in PVF, phonemic clustering in SVF) is a prevalent strategy representing automatic semantic activation or the use of intentional and effortful cognitive strategies in PVF (Abwender et al., 2001; Sung et al., 2013). Considering different strategies for clustering, multiple qualities can be attached to one word, and surrounding words often define the intended subcategory (Becker & Salles, 2016). Switching is the ability to move to a new subcategory when the previous subcategory is exhausted. Compared to clustering, switching is considered being a more effortful process. It is considered to involve higher cognitive functions, such as cognitive flexibility and strategic search processes (Patra et al., 2020; Strauss et al., 2006; Thiele et al., 2016; Troyer, 2000; Troyer et al., 1997). Clustering and switching are closely related as the method of calculating clusters determine the number of switches.

In general, higher education predicts a larger cluster size and more switches in both task types, potentially due to a more robust semantic network or larger vocabulary size (Pereira et al., 2018; Troyer et al., 1997). Typically older adults generate larger clusters, possibly reflecting a more extensive vocabulary (Troyer et al., 1997). Older adults also switch less than younger adults signaling an age-related decline in higher executive functions, especially in SVF (Lanting et al., 2009; Troyer, 2000; Troyer et al., 1997). Men tend to generate larger cluster sizes, with women switching more, especially in SVF (Lanting et al., 2009; Weiss et al., 2006).

VF tasks are analyzed across languages and cultures. Thus, language and culture-specific details should be considered when analyzing data and interpreting the results from different populations (Ardila, 2020; Becker & Salles, 2016; Kim et al., 2019; Oberg & Ramírez, 2006). Language-specific scoring guidelines can be essential in languages with productive compounding or extensive use of inflectional and derivational morphemes (Tallberg et al., 2008). In semantic clustering, predefined subcategories can be too narrow or broad to reveal culturally and linguistically unique lexical retrieval strategies (Becker & Salles, 2016; Roberts & Dorze, 1997). Unique retrieval strategies include, for example, dialectical variation and the influence of various cultural settings, as shown in studies investigating bilingual performance in VF tasks (Gollan et al., 2002; Rosselli et al., 2002). Detecting these subtle strategies requires thorough familiarization with the data from the population in question (Olabarrieta-Landa et al., 2017). Computational approaches typically train semantic models to analyze semantic variables using extensive written language corpora of the target language (Johns et al., 2018; Taler et al., 2020; Tröger et al., 2019) and it should be noted that subtle language variations can be underrepresented or omitted in written texts (Kim et al., 2019).

To summarize, VF tasks are a valuable and widely used tool for research and clinical purposes. An extensive amount of studies in multiple research areas have been conducted with varied categories utilizing both manual and computational approaches. Computational modeling has enabled great strides in VF task analysis via automatization, broadening our understanding of large-scale trends of human behavior (e.g., Kim et al., 2019; Taler et al., 2020). However, while detailed manual VF task analysis is time-consuming and can include inconsistencies (Kim et al., 2019), the need for manual analysis remains in smaller-scale studies where computational resources are not available and for specific populations and clinical purposes. Thus, to expedite the process of manual analysis and to increase the reliability and validity of verbal fluency task analysis across studies, comprehensive and precise instructions for administration, scoring, and analyses are warranted.

Aim of this study

In this study, we outlined detailed instructions for administering, scoring, and analyzing semantic and phonemic VF tasks to provide reliable tools for in-depth analysis in various clinical and research settings (Supplementary Appendix A, Instruction Manual for Administration and Scoring Verbal Fluency Tasks). To investigate the reliability of the proposed method, we demonstrated the analysis in a sample of middle-aged and older Finnish-speaking adults.

To investigate the reliability of the proposed method, we described overall task performance and aimed to show whether task type (PVF vs. SVF), age, education, or gender predicts (1) the total score and (2) the number of words generated in four 15-s segments. We also investigated (3) the frequency of errors and error types in PVF and SVF. For clustering and switching strategy use, we first investigated (4) whether task congruent cluster size and the number of switches or their interactions predict the total score. Second, we investigated (5) if the frequency of task discrepant clusters predicts total score.

We expected our results to be in line with previous literature. Based on previous research, we expected to see a higher overall score for SVF than PVF and participants with higher education to generate higher total scores than participants with lower education, especially in PVF. As the age range in our data was rather small, we expected minimal, if any, negative effect for age in total scores or performance in shorter time segments. For temporal parameters, we expected to find a systematic decline in the number of words generated in four 15-s segments. We expected participants with higher education to generate proportionally more words in the first 15-s segment of the task than participants with lower education. For clustering and switching, we expected the use of both strategies to contribute to a higher score. We expected a positive effect of education for task congruent cluster size and the number of task discrepant clusters as well as for the number of switches. We also expected to see more task discrepant clusters generated in the PVF task than in the SVF task with potential predictive power for higher total scores, especially in PVF. As for total scores, we expected age to have a minimal, if any, positive effect on cluster size and a negative effect on switching. We expected very little or no effect for gender overall.

Materials and methods

Participants

A sample of 50 middle-aged and older Finnish speakers with a higher proportion of women than men participated in this study. Sample mirrors the age and gender distribution of neurological deficits in populations (Roy-O'Reilly & McCullough, 2018) and thus serves the purpose of this study as a starting point for future studies consisting of larger clinical and control groups. Based on the calculations described below, the sample size was considered sufficient to demonstrate the proposed analysis methods and to show the effects of the size to have clinical relevance. Background information, health and language history were obtained in an interview setting by a graduate student in speech-language pathology via a comprehensive questionnaire. Participants were community-dwelling monolingual individuals who self-reported no history of language-related deficits or diagnoses (dyslexia, stroke, other neurological disorder) or hearing impairment. Participants who were not able to reliably report meeting the criteria described above were excluded.

Data for this study were collected as a part of an ongoing project on Finnish language attrition consisting of monolingual native Finnish speakers and participants with immigrant backgrounds. Data for monolingual performance in four verbal fluency tasks investigated in this study were extracted from a data pool of language tasks, including five verbal fluency tasks (one concrete semantic task (animals), one abstract semantic task (emotions), three phonemic categories (/k/, /a/, /p/). The research was conducted in accordance with the principles stated in the Declaration of Helsinki and the University of Turku Ethics committee approved all experimental procedures. All participants provided a written voluntary informed consent to participate in the study. They were informed of their right to withdraw at any time and did not receive compensation for participation. Demographic characteristics of participants and group internal comparisons are presented in Table 1.

Table 1. Demographic characteristics of participants and group internal comparisons for participant groups.

Demographic variable	Total $N = 50$	Education <12 years $n = 27$	Education >12 years $n = 23$	p
Age				
Mean (SD)	62.58 (7.59)	62.23 (7.84)	62.96 (7.43)	.75[a]
Range	49–79	49–79	52–79	
Gender				
Female n (%)	35 (70)	20 (74)	15 (65)	.50[b]
Male n (%)	15 (30)	7 (26)	8 (35)	

Note. Education < 12 years = no academic degree; Education > 12 years = academic degree.
[a]Groups were compared using Two sample *t*-test.
[b]Groups were compared using Chi-square goodness of fit test.

Verbal fluency tasks

Data were collected as a part of a larger test battery as described above. Here, we report three phonemic verbal fluency tasks (/k/, /a/, /p/) and the concrete semantic verbal fluency task (animals). The order of VF tasks within the test battery was fixed (semantic, phonemic).

Categories

Phonemic verbal fluency tasks were localized into the Finnish language by using the most frequent word-initial consonants of Finnish /k/ (15,242 words) and /p/ (10,640 words) and the most frequent initial vowel /a/ (4,361 words) (Kielitoimiston Sanakirja, 2021; Leskinen, 1989), allowing comparison to earlier studies and providing a reference point for future studies.

For the semantic verbal fluency tasks, the category selected was "animals." This category is commonly used, culturally and linguistically relatively neutral, and included in many neuropsychological test batteries.

Administration and scoring

For a detailed instructions manual for administration and scoring, see Supplementary Appendix A. All tasks were completed in a quiet environment in a single session. Participants were encouraged to take short breaks in between tasks when needed. Responses were recorded for later verification and scoring. A research assistant transcribed audio tracks, and the authors verified transcripts.

For each trial, participants were asked to produce as many individual words as possible in 60 seconds.

Our goal was to have as simple and straightforward task instructions as possible. We approached this by including the phrase "individual words" in our instructions to discourage participants from using inflections (e.g., *kirja* [book], *kirjani* [my book]) but not to inhibit the use of derivational words that carry independent semantic meaning (e.g., *kirja* [book], *kirjasto* [library]) or compound words (e.g., *kirjakauppa* [book store]). This was a language-specific choice as Finnish has rich derivational and inflectional morphology, and it uses compounding productively to form new words (Helasvuo, 2008; Tyysteri, 2015). Using the wording "individual words" was also considered to guide participants not to produce multiple numerals without adding complexity to the instruction. In addition, we chose to use the word "letter" instead of "phoneme" in PVF, even to assess phonemic, not spelling, fluency. Finnish has strong orthographic transparency, and the words letter and sound are strongly interchangeable (Suomi et al., 2006). The more common word "letter" was selected to simplify instructions as much as possible. The only restriction was the use proper of names.

All verbal fluency tasks were scored for (1) total score, (2) number of acceptable words generated in 15-s segments, (3) number of errors and error types, (4) mean cluster size for task congruent clusters (semantic per semantic and phonemic per phonemic), and the number of switches calculated from task congruent clusters as well as for (5) number of task discrepant clusters (semantic clusters in PVF and vice versa).

We chose to investigate the distribution of words in 15-s segments to demonstrate the function of time comprehensively and to allow a simple combination for 30-s segment analysis if needed. We described the frequency of errors and error types to screen for error frequency and variety in our sample. Lastly, for clustering and switching, we based our analysis on task congruent clusters following Troyer et al. (1997) but chose to use naturally occurring clusters instead of fixed subcategories and to apply the rule for the smallest possible cluster in the analysis. In addition, we tracked the use of task discrepant clusters as suggested by Sung et al. (2013) and Abwender et al. (2001). In regards to detailed rules and instructions on coding semantic and phonemic clusters, see Supplementary Appendix A. Briefly, naturally occurring clusters were determined by calculating the number of words generated in individual subcategories for each participant. Under the semantic condition, naturally occurring clusters are typically taxonomic subcategories of animals (e.g., birds or big cats). However, they can also be formed by environmental semantic connections (e.g., farm animals), geographical semantics (e.g., African animals), or visual semantics (e.g., snake, eel). Under the phonemic condition, clusters are typically formed by the same two initial phonemes or rhyming words, but they can also be formed by structurally similar words that differ only by one sound or vowel sounds.

Table 2. Intraclass correlation coefficients and their 95% confidence intervals for mean cluster sizes and the number of switches.

Variable	ICC $n = 140$	95% CI LL	95% CI UL
Phonemic cluster size	.97	.96	.98
Number of phonemic switches	>.99	.99	>.99
Semantic cluster size	.79	.73	.84
Number of semantic switches	.96	.95	.98

Note. ICC: intraclass correlation coefficient; CI: confidence interval; LL: lower limit; UL: upper limit.

Naturally occurring clusters include all culturally and linguistically unique strategies participants might use (Becker & Salles, 2016; Roberts & Dorze, 1997) and account for individual semantic networks and their possible influence on semantic categorization (Morais et al., 2013). Determining naturally occurring clusters eliminates the need for predefined categories mirroring computational approaches that extract information from natural language (e.g., Kim et al., 2019). Following the rule for the smallest possible cluster size allows for specific tracking of switching. We are aware that analyzing task congruent and task discrepant clusters separately does not account for the effect task discrepant clusters might have on switching. However, to accurately track the use of both cluster types, they are scored and analyzed separately, and only the number of switches based on task congruent clusters is analyzed.

Inter-rater reliability for cluster size and number of switches

To verify the reliability of the analysis for cluster size, two raters coded the data following instructions in Supplementary Appendix A. Based on a larger dataset (3 tasks) and high inter-rater reliability for phonemic clustering in literature (Becker & Salles, 2016; Ross, 2003) 60% of the PVF data ($n = 90$; $n = 30$ for each phoneme) were used for inter-rater reliability analysis. Due to a smaller dataset (1 task) and a semantic component of the analysis, 100% of the SVF data ($N = 50$) were used for inter-rater reliability analysis. We calculated the Intraclass Correlation Coefficient (ICC) for all, task congruent and task discrepant, phonemic, and semantic cluster sizes in both verbal fluency tasks using R (R R Core Team, 2019) and "psych" package (Revelle, 2020). We selected a two-way random-effects model with a single measurement and absolute agreement to show the magnitude of agreement achieved between two raters for these measures as we aim to generalize the reliability of the results to raters who want to use the same analysis in their clinical or research work (Koo & Li, 2016).

For the phonemic cluster size, the number of phonemic switches, and the number of semantic switches, the ICC analysis showed an excellent degree of reliability between the two raters as ICCs are above 0.90 (Table 2) (Koo & Li, 2016). For the semantic cluster size, the reliability between the raters was good; ICC value between 0.75 and 0.90 (Table 2) (Koo & Li, 2016). A blind review by a third rater showed that lower ICC in the semantic clustering analysis resulted primarily from differences between the raters in weak cluster size where there were multiple semantically acceptable ways to form clusters (e.g., see Supplementary Appendix A). As subjective semantic categorization will always be present in semantic clustering analysis, this result was determined to demonstrate acceptable semantic variation between raters.

Data analysis

R software (R Core Team, 2019) with packages dplyr (Wickham et al., 2019), tidyr (Wickham, 2020), lme4 (Bates et al., 2015), and lmerTest (Kuznetsova et al., 2017) were used in data cleanup and analyses. Packages sjPlot (Lüdecke, 2018), jtools (Long, 2020), ggeffects (Lüdecke, 2018), and ggplot2 (Wickham, 2016) were used in tables and figures and packages effect size (Ben-Shachar et al., 2020) and EMAtools (Kleiman, 2017) for calculating effect size estimates. Model assumptions were checked using" check_model"-function from package performance (Lüdecke et al., 2020). Our data and analysis scripts are available at https://osf.io/kh8f3/.

We modeled PVF and SVF performance together. This allowed us to investigate the main effects of task type (the differences in performance between the tasks) but also the effects of demographic variables in PVF and SVF separately as well as in the combined data of both task types. To answer our research question on task type and participant age, education, and gender predicting the total score, we employed a linear mixed-effects model with the total score as a response variable, and task type, gender (female/male), age, education (high/low) and all two-level interactions (interactions of gender, age or education, and task type) as predictors. For modeling purposes, the age variable was scaled and centered to the sample mean so that the estimates would reflect the performance in mean age rather than at 0 years. This was considered to be more informative in this context. To supplement the analysis, we provide descriptive data on total scores for in-age bonds of 49–59, 60–69, and 70–79. The two-level education variable was centered between high and low values so that the model estimates would better reflect performance in the whole population, regardless of the education level. Participant IDs within each level of task type were used as random factors to account for individual variation in performance.

To address our research question on task type predicting the words produced during each 15-s segment of the task, we modeled the number of words produced as a function of each 15-s segment. Participant intercept and individual slope for task type were applied as random factors. To select the most parsimonious model fit to our data, models with participant background information (gender, age, education) as additional predictors were compared to the model with 15-s segments only as a predictor using analysis of variance and BIC values. Based on this model selection procedure, the model without participant background information turned out to be the most parsimonious fit for the data.

The frequency and type of errors are described as raw scores. Mean values are reported to enable comparison between the tasks due to the small number of errors in the data. The number of participants who generated errors is described for all tasks separately.

To address our research question on clustering and switching, we modeled the total score as a function of task congruent cluster size, the number of switches, and their interactions. Separate models were used for PVF and SVF tasks. In the phonemic model, participant intercept and slope for the task (/k/, /a/, or /p/) were used as random factors. Since the semantic dataset only had one task per participant, using a mixed-effects model was unnecessary, and we employed a simple linear regression instead. To control the effects of participant background variables, we also considered participant education, age, and gender and all their interactions as potential predictors in the models. We performed a similar model selection than in the 15-s segment model, which suggested the model with no background variables as additional predictors to be the most parsimonious fit for the phonemic task performance and the model with education as an additional predictor to be the most parsimonious fit for the semantic task performance.

To investigate the use of task discrepant clusters, we modeled the total score in the phonemic tasks as a function of semantic clusters and the total score in the semantic task as a function of the number of phonemic clusters. In the phonemic model, we also included trial order as a predictor in the model to investigate whether the position of the phonemic trial had an effect on the number of semantic clusters as the order of trials was not randomized. All participants used task discrepant clustering in both task types, there were 39 phonemic trials from 30 participants with no task discrepant clusters. Trials with no task discrepant clusters were excluded from the phonemic model. Based on a similar model selection procedure as described above, we chose the models with no background variables as the most parsimonious ones.

We performed power calculations to ensure that the sample size provided adequate power for our outcome measures. However, power calculation for statistical methods we used is not straightforward: because of the model selection procedure, we did not know the exact number of predictors before model fitting. Also, there is no exact method to calculate power for linear mixed models (containing random effects part). We did not have prior estimates about the random effects (i.e., within-subjects variation) available from previous studies to perform appropriate power simulations. Thus, we based our sample size estimations on power calculations for linear regression with similar sample sizes, effect sizes, and the number of predictors. These power calculations suggested a statistical power >0.8 for all models except those examining clusters and switches or task discrepant clusters in the semantic task. In these models, the statistical power was 0.59 and 0.77, respectively. The statistical power was lower in these models because we only had data from one semantic task for each participant as compared to 111–800 data points in other models. In addition, as determining exact degrees of freedom for the test statistics estimated by linear mixed models is difficult, it is also problematic to determine unambiguous p-values (see Baayen et al., 2008). Hence, the statistical significance at the 0.05

Table 3. Descriptive statistics of total scores and temporal parameters in semantic and phonemic verbal fluency tasks.

Variable	Phonemic /k/	Phonemic /a/	Phonemic /p/	Semantic animals
Total score 0–60 s				
Mean	19.38	14.68	17.44	25.96
SD	6.12	5.65	5.50	5.90
Range	5–35	5–31	4–30	12–43
Time segment 0–15 s				
Mean	7.06	5.72	6.26	10.76
SD	2.37	2.25	2.20	2.14
Range	2–12	1–11	2–12	6–15
Percentage of total score (%)	36	39	36	41
Time segment 16–30 s				
Mean	4.48	3.54	4.24	6.06
SD	2.04	1.94	1.69	2.13
Range	0–9	0–8	1–8	0–10
Percentage of total score (%)	23	24	24	23
Time segment 31–45 s				
Mean	4.12	2.98	3.48	4.94
SD	1.90	1.64	1.66	2.36
Range	0–9	0–7	0–7	0–13
percentage of total words (%)	21	20	20	19
Time segment 46–60 s				
Mean	3.72	2.44	3.46	4.20
SD	1.51	1.57	1.85	2.36
Range	0–7	0–7	0–8	0–10
Percentage of total score (%)	19	17	20	16

Note. Total score = total number of acceptable words generated in a 60-s trial; Time segment = number of acceptable words generated in a 15-s time segment within a 60-s trial.

level in this article is indicated by $|t| > 1.96$. However, since many readers might be more familiar with p-values than t-values, rough estimates for p-values are also provided in model summaries (Supplementary Appendix B).

Results

For the sake of brevity, all full model summaries with effect size estimates and supplementary information on age bands are presented in Supplementary Appendix B.

Total score

Modeling the total score as a function of task type, participant gender, education, and age revealed main effects of task type, 8.47, $t = 8.96$, 95% CI [6.62, 10.33], gender, −5.06, $t = -3.78$, 95% CI [−7.69, −2.44] and education, 4.55, $t = 3.75$, 95% CI [2.18, 6.93]. Higher numbers of correct words were associated with semantic task type, higher education, or being female. Age was not significantly associated with task performance in our sample, −1.00, $t = -1.63$, 95% CI [−2.20, 0.20]. Interactions between the task type and education, −0.21, $t = -0.14$, 95% CI [−3.31, 2.88], task type and age, 0.53, $t = 0.67$, 95% CI [−1.03, 2.10] or task type and gender, 1.03, $t = 0.59$, 95% CI [−2.38, 4.44] were not significant. Descriptive statistics of raw scores in all four fluency tasks are presented in Table 3.

15-s segments

Modeling words produced during each 15-s segment revealed that frequency in producing acceptable words decreased during the task (Figure 1). This decrease was more substantial in semantic, as opposed to phonemic task type. Descriptive statistics of raw scores in all four fluency tasks are presented in Table 3.

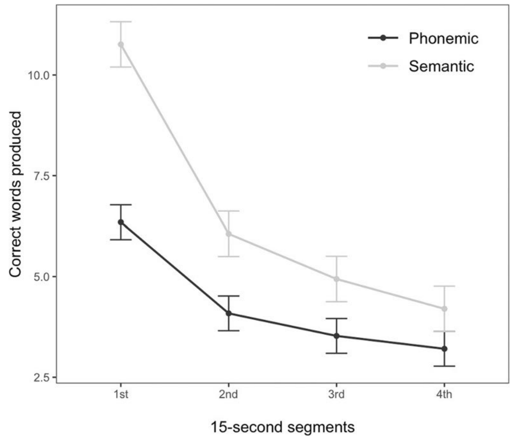


Figure 1. Predicted values of words produced in the four 15-s segments in phonemic and semantic verbal fluency tasks. Error bars represent 95% confidence intervals.

Errors

In PVF, roughly half of the participants generated errors in all three trials (/k/ 44% [$n = 22$]; /a/ 58% [$n = 29$]; /p/ 50% [$n = 25$]) with most errors being repetitions and categorical errors. In the SVF task, 28% of participants ($n = 14$) generated errors of which most were repetitions. The total number of errors and distribution of error types are presented in Table 4.

Task congruent clustering and switching

The number of switches, 6.22, $t = 15.39$, 95% CI [5.43, 7.02] (phonemic); 7.59, $t = 10.40$, 95% CI [6.11, 9.06] (semantic), and the mean cluster size, 1.72, $t = 4.14$, 95% CI [0.91, 2.54] (phonemic); 4.65, $t = 6.86$, 95% CI [3.29, 6.02] (semantic); were associated with the total score in both tasks. The semantic task model also included participant education as a predictor. The main effect of education, 2.98, $t = 2.87$, 95% CI [0.88, 5.08], was consistent with the observation in the total score model reported earlier in this chapter. Two- and three-level interactions of education, number of switches, and cluster size were not significant. Descriptive statistics of raw scores in all four fluency tasks are presented in Table 5, and all interactions are presented in Figure 2.

Task discrepant clustering

Participants generated task discrepant clusters in both task types. Numerically, most task discrepant clusters were generated under phoneme /k/ ($M = 2.98$, SD = 2.01), followed by /p/ ($M = 1.7$, SD = 1.17). Least task discrepant clusters in

Table 4. Number of participants who generate errors, total number of errors, distribution of error types and in semantic and phonemic verbal fluency tasks.

Errors	Phonemic /k/	%	Phonemic /a/	%	Phonemic /p/	%	Semantic animals	%
n with errors	22	44	29	58	25	50	14	28
Total no. of errors	39		44		35		20	
Mean	1.77		1.52		1.40		1.43	
SD	1.02		0.95		0.50		0.65	
Range	1–4		1–4		1–2		1–3	
Error type								
Repetition	26	67	25	57	27	77	19	95
Categorical	12	31	16	36	7	20	0	0
Nonword	1	3	3	7	1	3	1	5

Note. $N = 50$.

Table 5. Descriptive statistics of task congruent cluster size, number of task discrepant clusters and switches in semantic and phonemic verbal fluency tasks.

Variable	Phonemic /k/	Phonemic /a/	Phonemic /p/	Semantic animals
Task congruent cluster size[a]				
Mean	2.62	2.24	2.56	2.7
SD	0.83	0.34	0.78	0.37
Range	2.00–5.50	2.00–3.25	2.00–5.60	2.00–3.75
n with clusters[b]	46	42	47	50
%[c]	92	84	94	100
Number of task discrepant clusters[d]				
Mean	2.98	0.80	1.70	0.90
SD	2.01	1.18	1.17	1.15
Range	0–7	0–5	0–4	0–5
n with clusters[b]	46	22	43	26
%[c]	92	44	86	52
Switches[e]				
Mean	13.88	11.52	11.70	11.60
SD	5.95	3.90	3.97	3.07
Range	2–26	1–20	3–22	5–20

Note. $N = 50$ in all conditions.
[a]SVF semantic clusters, PVF phonemic clusters.
[b]Number of participants who generated clusters.
[c]Percentage of participants who generated clusters.
[d]SVF phonemic clusters, PVF semantic clusters.
[e]Number of switches calculated from task congruent clusters.

PVF were generated under vowel /a/ (M = 0.8, SD = 1.18) and this was in line with task discrepant cluster frequency in the semantic category (M = 0.9, SD = 1.15). In addition to the group mean values, it is worth noting that in the PVF category /k/ 92% ($n = 46$) of participants, in /p/ 86% ($n = 43$) of participants, and in /a/ 44% ($n = 22$) of participants generated semantic clusters. In SVF, 52% of the participants ($n = 26$) generated phonemic clusters. In PVF, the number of semantic clusters was a significant predictor for the total score, 1.42, $t = 6.60$, 95% CI [1.00, 1.84]. The interaction of the number of semantic clusters and trial order, 0.02, $t = 0.07$, 95% CI [−0.52, 0.56], was not a significant predictor for the use of task discrepant clustering nor was the main effect of trial order a significant predictor for the total score, −0.30, $t = -0.43$, 95% CI [−1.68, 1.08]. In the semantic task, the number of phonemic clusters did not predict the total score of 0.52, $t = 0.44$, 95% CI [−1.89, 2.93]. Descriptive statistics of raw scores in all four fluency tasks are presented in Table 5.

Discussion

This study describes a comprehensive analysis of phonemic and semantic verbal fluency tasks for clinical and research purposes. In addition to total scores, we demonstrate an analysis for temporal parameters, errors, and clustering and switching with strong inter-rater reliability in a sample group of 50 older healthy participants with the aim of providing a starting point for future studies. As discussed below, our results align with earlier literature, supporting the proposed method as a reliable starting point to analyze linguistic and cognitive processes underlying VF performance in varied clinical groups.

In our dataset of monolingual middle-aged and older adults, participants generated a higher total score in the semantic than in the phonemic tasks, in line with multiple normative datasets (for an overview, see Strauss et al., 2006). In the PVF, the trial total scores reflected the category size of word-initial phonemes in the Finnish language, as expected (Gollan et al., 2002). We evaluated education on a 2-tier scale and found a positive association to performance on both fluency types in line with Pereira et al. (2018). As hypothesized, we found no association between age and task performance in either task type. The lack of age effect is likely due to our sampling process resulting in the relatively small age range in our data (Ardila, 2020). While we found no significant age effect, the trends in our data shown in

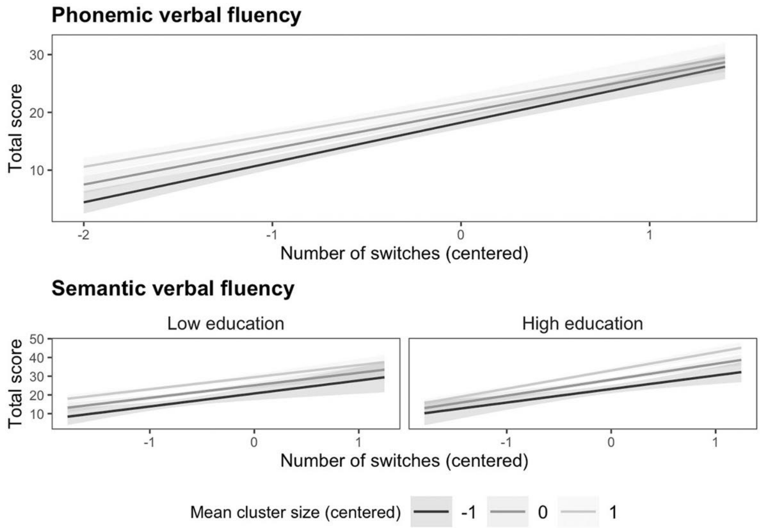


Figure 2. Estimates for the total score of phonemic and semantic tasks as a function of all predictor variables included in the models. For the sake of clarity, the continuous variable of centered mean cluster size is estimated on three levels. Error bars represent 95% confidence intervals. Note that due to the model selection procedure used, semantic but not phonemic model had education level as a predictor variable.

Supplementary Appendix B can be of value for clinical observations.

As expected, modeling task performance in 15-s segments revealed a performance decrease during both the task types. Following Crowe (1998), we infer that word retrieval becomes more effortful as time passes, reducing the number of words produced. Our results replicated results from Venegas and Mansur (2011), with performance in the first quartile being strongly associated with the overall score. In addition, the results corroborated with Kim et al. (2011), who suggest that 30-s total time for VF tasks can be a relevant approach to differentiate stroke patients with and without aphasia. Thus, it is tempting to suggest that it might be clinically efficient to screen patients using a very short VF task (i.e., 15 s) under specific circumstances. However, due to the limited sample size and focus on older adults across these studies, future studies with larger datasets and varied populations are needed to verify the relevance of this finding for clinical use.

Participants generated errors following error profiles described by Crowe (1998) and Gollan et al. (2011), with the most common error type being repetitions. Contrary to findings in Crowe (1998), we found more errors in PVF than in SVF. We did not randomize the order of trials but presented the semantic trial first, followed by phonemic trials. Thus, some errors in PVF could have been due to active semantic retrieval strategies in the first phonemic trial in PVF. However, no errors stemming from semantic strategies were detected. We conclude that generating errors with an emphasis on repetitions is a part of a verbal retrieval process in a healthy older monolingual population across tasks (Crowe, 1998; Gollan et al., 2011). A more significant number of errors, or a different distribution of error types than found here, can be an atypical finding, and as Thiele et al. (2016) point out, can yield insight into various processes underlying task performance in clinical populations.

We found both task congruent clustering and switching to be productive strategies in phonemic and semantic VF tasks following Troyer et al. (1997), but no interactions between task congruent mean cluster size and switching were detected. Thus, the efficacy of switching as a strategy was not dependent on cluster size, nor was the efficacy of cluster size dependent on the number of switches in either task type. Here we point out that a deficit in the use of one of the strategies (clustering or switching), can lead to more extensive use of the other strategy.

Task discrepant clustering was common in phonemic trials. Depending on the trial, 44–98% of the participants generated task discrepant clusters. Our results align with Abwender et al. (2001), supporting the importance of including task discrepant clusters in VF task analysis. Notably, the number of semantic clusters predicted the total score in PVF, but the number of phonemic clusters in SVF did not. We think there are three possible explanations. First, as we had data from only one SVF trial (compared to three PVF trials), it is possible that we did not have sufficient statistical power to detect small effects. However, we are confident that effects large enough to have practical significance would have been observed in our sample. Second, our results support the notion that automatic semantic activation plays an essential role in both semantic and phonemic tasks, resulting in a more pronounced use of semantic clusters in phonemic tasks (Sung et al., 2013). Third, the use of semantic clustering in a phonemic task can be an additional, intentional strategy resulting from participants reaching out to their hierarchical semantic memory in an effortful phonemic task (Abwender et al., 2001). In the future, exploring the temporal distribution of task discrepant clusters in PVF could provide evidence if semantic clustering is used throughout the task suggesting automatic semantic activation or if semantic strategies are used in the latter time segments suggesting a more intentional strategy use.

Here, we must consider if conducting the SVF trial before the PVF trials primed participants for semantic retrieval in PVF. If so, we would expect to see most semantic clusters in the first phonemic trial /k/. Statistical modeling of cluster frequency showed that the number of task discrepant clusters in phonemic trials does not differ between phonemic trials /k/, /a/, and /p/. Also, the proportion of participants who generated semantic clusters in phonemic trials lines up with phoneme category size rather than the order of trials presented. This suggests that the overall number of words available in a category can influence how semantic associations are activated in a phonemic VF task, reflected in our data as the number of participants who generated task discrepant clusters. Even with no difference between phoneme trials in the number of task discrepant clusters generated, we can not exclude the possibility of task order impacting task performance. Thus, we recommend randomizing tasks when implementing multiple VF tasks for research to minimize the possible effect task order might have on the performance.

In this study, we tracked the use of both cluster types, task congruent and task discrepant clusters, separately to simplify the coding for clinical use. Thus, our approach did not include switches stemming from task discrepant clustering. Our results show that both cluster types are evident and significant factors in VF task performance. We acknowledge the challenges including switching from both cluster types in the analysis but suggest that in the future, a combination of the two clustering analyses would be essential in determining the most reliable analyzing method for switching, especially in research settings.

Some limitations of the current study include screening patients' cognitive health via a self-reporting questionnaire and interview without standardized methods and a small sample size with a narrow age range. Based on rough estimates of statistical power, the sample size was sufficient in all models except clustering models in the semantic task. The statistical power was lower in these models due to only one semantic trial vs. three trials in the phonemic task. However, these preliminary findings on the semantic task can be valuable as guidelines to future studies, and even with limited sample size, we are confident that any effects of the size to have clinical relevance would be observed with this sample size and analyzing method. In future studies, standardized methods for cognitive screening should be included for reliable sampling. Exploring the suggested analysis in larger data samples of healthy participants with a broader age range as well as across languages and in varied clinical groups is needed to solidify our findings.

In the following, we further discuss the proposed method. The following aspects relate but are not limited to the Finnish language. Considerations for other languages should always be made in relation to the language and culture in question (e.g., for Spanish, see Olabarrieta-Landa et al., 2017).

In the administration for PVF tasks, we used the word "letter" instead of "sound." This was a deliberate, language-specific choice as the Finnish language has a strong letter-phoneme correspondence, and the use of the word "letter" is more common than the word "sound" in everyday language. However, there were some instances in the data where participants generated words that had the correct word-initial sound but that are spelled differently (e.g., */panaani/* with unvoiced /p/ vs. correct spelling */banaani/* with voiced /b/; [banana]), a common occurrence in spoken Finnish with a dialectical variation. As we did not evaluate spelling fluency these words were considered deviations from the task condition only when participants indicated that they had produced the word in error (e.g., "No, that begins with a different letter"). To eliminate potential confusion for the participants and simplify the analysis, we recommend using the word "sound" for PVF instructions in all languages following Olabarrieta-Landa et al. (2017) when assessing phonemic fluency. Using the word "sound" will also eliminate potential errors due to lack of spelling knowledge and reduce potential errors and lower total scores stemming from lower education. In our task instructions, we used the phrase "individual words" to discourage participants from producing multiple numerals in phonemic trials. This proved to be an effective strategy as very few participants included a string of numerals in PVF. Thus, we argue that the instruction is precise enough for VF task purposes without a restriction for sequential numerals.

Our clustering analysis consisted of naturally occurring clusters for each participant in both VF task types. In PVF, this meant specific rules to include all phonemic clustering strategies participants utilized in their output. In line with earlier literature, the inter-rater reliability for phonemic clustering was excellent between two raters (Becker & Salles, 2016; Ross, 2003; Troyer et al., 1997). For the PVF task, we did not allow lexical categories (verbs, adjectives, particles) as a basis for clustering due to interlanguage challenges in scoring demonstrated by Rosselli et al. (2002). However, during analysis, we did observe some participants using lexical categories as a productive word search strategy (e.g., *kiirehtiä* [to hurry], *keittää* [to boil], *kutittaa* [to tickle]). In future studies, it could be worthwhile to investigate if lexical categories could be included as an individual productive strategy for PVF utilizing language-specific guidelines for scoring.

In SVF, naturally occurring strategies allowed us to analyze subcategories in an area of expertise (e.g., birds of prey, aquatic birds), clustering based on geographical semantics (giraffe, monkey), and visual semantics (snake, eel). It also facilitated the use of context to determine the intended category for clustering analysis (e.g., forest animals, animals typical to Lapland; Becker & Salles, 2016). We chose this approach to include all language and culture-specific as well as individual clustering strategies resulting in a precise analysis of clustering and switching. Our inter-rater analysis indicated good reliability between raters and was deemed sufficiently objective, but the margin of error between raters was more prominent in the semantic than in phonemic clustering. Computational approaches aim to create objective semantic variables to reduce subjectivity (Kim et al., 2019; Tröger et al., 2019), but manual analysis of semantic clustering strategies includes a subjective semantic component (Tröger et al., 2019). This variability due to subjective semantic interpretation is important to bear in mind while applying these rules in clinical and research settings. To minimize this variability in clinical and research settings where analysis is done manually, we have included a sample protocol and instructions on training raters in Supplementary Appendix A, as suggested by Ross (2003).

In conclusion, this study provides a starting point for a comprehensive analysis of VF performance. Future studies applying the suggested method in varied clinical groups will further test and solidify the method. Currently, we are working on implementing these analyses to investigate lexical processes underlying performance in VF tasks in bilingualism, aphasia, and Alzheimer's Disease. A consistent approach in administration, scoring, and analyzing VF data across studies is needed to enable systematic insight into cognitive processes underlying or possibly hindering optimal performance in different populations. We are hopeful that our research will be beneficial in determining specific and straightforward scoring rules for phonemic and semantic VF tasks.

Acknowledgments

We would like to thank Pirjo Korpilahti and Professor Marja-Liisa Helasvuo for their scientific advice, valuable comments, and support. We would also like to thank research assistants at the Department of Psychology and Speech-Language Pathology, University of Turku, for their assistance in data collection and analysis. Finally, we are grateful to the participants of this study for the generous donation of their time and effort in taking part in this research.

Disclosure statement

The authors do not have conflicts of interest regarding this research study.

Funding

This work was supported in part by the Alfred Kordelin Foundation under a grant awarded to the first author and University of Turku Graduate School wages awarded to the second and third authors, as well as an anonymous endowed fund to the University of Turku, Department of Speech-Language Pathology.

ORCID

Nana Lehtinen http://orcid.org/0000-0002-3606-1302
Ida Luotonen http://orcid.org/0000-0003-3310-3180
Anna Kautto http://orcid.org/0000-0003-4885-9167

Data availability statement

The data supporting the findings of this study are openly available in Verbal Fluency at https://osf.io/kh8f3/.

References

Abwender, D. A., Swan, J. G., Bowerman, J. T., & Connolly, S. W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment*, *8*(3), 323–338. https://doi.org/10.1177/107319110100800308

Ardila, A. (2020). A cross-linguistic comparison of category verbal fluency test (animals): A systematic review. *Archives of Clinical Neuropsychology*, *35*(2), 213–225. https://doi.org/10.1093/arclin/acz060

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barry, D., Bates, M. E., & Labouvie, E. (2008). FAS and CFL forms of verbal fluency differ in difficulty: A meta-analytic study. *Applied Neuropsychology*, *15*(2), 97–106. https://doi.org/10.1080/09084280802083863

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Becker, N., & Salles, J. F. d. (2016). Methodological criteria for scoring clustering and switching in verbal fluency tasks. *Psico-USF*, *21*(3), 445–457. https://doi.org/10.1590/1413-82712016210301

Ben-Shachar, M. S., Makowski, D., & Lüdecke, D. (2020). *Compute and interpret indices of effect size*. CRAN. https://github.com/easystats/effectsize

Borkowski, J. G., Benton, A. L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia*, *5*(2), 135–140. https://doi.org/10.1016/0028-3932(67)90015-2

Bose, A., Wood, R., & Kiran, S. (2017). Semantic fluency in aphasia: Clustering and switching in the course of 1 minute: Semantic fluency in aphasia. *International Journal of Language & Communication Disorders*, *52*(3), 334–345. https://doi.org/10.1111/1460-6984.12276

Cavaco, S., Gonçalves, A., Pinto, C., Almeida, E., Gomes, F., Moreira, I., Fernandes, J., & Teixeira-Pinto, A. (2013). Semantic fluency and phonemic fluency: Regression-based norms for the Portuguese population. *Archives of Clinical Neuropsychology*, *28*(3), 262–271. https://doi.org/10.1093/arclin/act001

Crowe, S. F. (1998). Decrease in performance on the verbal fluency test as a function of time: Evaluation in a young healthy sample. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 391–401. https://doi.org/10.1076/jcen.20.3.391.810

Fernaeus, S.-E., & Almkvist, O. (1998). Word production: Dissociation of two retrieval modes of semantic memory across time. *Journal of Clinical and Experimental Neuropsychology*, *20*(2), 137–143. https://doi.org/10.1076/jcen.20.2.137.1170

Fernaeus, S.-E., Östberg, P., Hellström, Å., & Wahlund, L.-O. (2008). Cut the coda: Early fluency intervals predict diagnoses. *Cortex*, *44*(2), 161–169. https://doi.org/10.1016/j.cortex.2006.04.002

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, *16*(4), 562–576. https://doi.org/10.1037/0894-4105.16.4.562

Gollan, T. H., Sandoval, T., & Salmon, D. P. (2011). Cross-language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological Science*, *22*(9), 1155–1164. https://doi.org/10.1177/0956797611417002

Goral, M. (2004). First-language decline in healthy aging: Implications for attrition in bilingualism. *Journal of Neurolinguistics*, *17*(1), 31–52. https://doi.org/10.1016/S0911-6044(03)00052-6

Harrison, J. E., Buxton, P., Husain, M., & Wise, R. (2000). Short test of semantic and phonological fluency: Normal performance, validity and test-retest reliability. *The British Journal of Clinical Psychology*, *39*(2), 181–191. https://doi.org/10.1348/014466500163202

Helasvuo, M.-L. (2008). Aspects of the structure of Finnish. In *Research in Logopedics, Speech and Language Therapy in Finland. Communication disorders across languages* (pp. 9–18). Multilingual Matters.

Jewsbury, P. A., & Bowden, S. C. (2017). Construct validity of fluency and implications for the factorial structure of memory. *Journal of Psychoeducational Assessment*, *35*(5), 460–481. https://doi.org/10.1177/0734282916648041

Johns, B. T., Taler, V., Pisoni, D. B., Farlow, M. R., Hake, A. M., Kareken, D. A., Unverzagt, F. W., & Jones, M. N. (2018). Cognitive modeling as an interface between brain and behavior: Measuring the semantic decline in mild cognitive impairment. *Canadian Journal of Experimental Psychology*, *72*(2), 117–126. https://doi.org/10.1037/cep0000132

Kertesz, A. (1982). *The western aphasia battery*. Psychological Corporation.

Kielitoimiston sanakirja. (2021). *Kielitoimiston sanakirja*. https://www.kielitoimistonsanakirja.fi

Kim, H., Kim, J., Kim, D. Y., & Heo, J. (2011). Differentiating between aphasic and nonaphasic stroke patients using semantic verbal fluency measures with administration time of 30 seconds. *European Neurology*, *65*(2), 113–117. https://doi.org/10.1159/000324036

Kim, N., Kim, J.-H., Wolters, M. K., MacPherson, S. E., & Park, J. C. (2019). Automatic scoring of semantic fluency. *Frontiers in Psychology*, *10*, 1020. https://doi.org/10.3389/fpsyg.2019.01020

Kleiman, E. (2017). *EMAtools: Data management tools for real-time monitoring/ecological momentary assessment data*. CRAN. https://CRAN.R-project.org/package=EMAtools

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lanting, S., Haugrud, N., & Crossley, M. (2009). The effect of age and sex on clustering and switching during speeded verbal fluency tasks. *Journal of the International Neuropsychological Society*, *15*(2), 196–204. https://doi.org/10.1017/S1355617709090237

Leskinen, H. (1989). Tietoja sananalkuisten grafeemien ja grafeemikombinaatioiden yleisyyssuhteista. *Virittäjä*, *93*(3), 401. https://journal.fi/virittaja/article/view/38309

Long, J. A. (2020). *jtools: Analysis and presentation of social scientific data*. CRAN. https://cran.r-project.org/package=jtools

Lüdecke, D. (2018). *sjPlot—data visualization for statistics in social science*. Zenodo. https://doi.org/10.5281/ZENODO.1308157

Lüdecke, D., Makowski, D., & Waggoner, P. (2020). *Performance: Assessment of regression models performance*. CRAN. https://CRAN.R-project.org/package=performance

Mardani, N., Jalilevand, N., Ebrahimipour, M., & Kamali, M. (2019). Clustering and switching strategies in verbal fluency tasks: Comparison between amyotrophic lateral sclerosis (als) and healthy controls. *Journal of Rehabilitation Sciences & Research*, *6*(1), 21–26. https://doi.org/10.30476/jrsr.2019.44718

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., Mellits, E. D., & Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*, *39*(9), 1159–1159. https://doi.org/10.1212/wnl.39.9.1159

Morais, A. S., Olsson, H., & Schooler, L. J. (2013). Mapping the structure of semantic memory. *Cognitive Science*, *37*(1), 125–145. https://doi.org/10.1111/cogs.12013

Oberg, G., & Ramírez, M. (2006). Cross-linguistic meta-analysis of phonological fluency: Normal performance across cultures. *International Journal of Psychology*, *41*(5), 342–347. https://doi.org/10.1080/00207590500345872

Olabarrieta-Landa, L., Torre, E. L., López-Mugartza, J. C., Bialystok, E., & Arango-Lasprilla, J. C. (2017). Verbal fluency tests: Developing a new model of administration and scoring for Spanish language. *NeuroRehabilitation*, *41*(2), 539–565. https://doi.org/10.3233/NRE-162102

Patra, A., Bose, A., & Marinis, T. (2020). Performance difference in verbal fluency in bilingual and monolingual speakers. *Bilingualism: Language and Cognition*, *23*(1), 204–218. https://doi.org/10.1017/S1366728918001098

Pekkala, S., Albert, M. L., Spiro, A., III, & Erkinjuntti, T. (2008). Perseveration in Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *25*(2), 109–114. https://doi.org/10.1159/000112476

Pekkala, S., Goral, M., Hyun, J., Obler, L. K., Erkinjuntti, T., & Albert, M. L. (2009). Semantic verbal fluency in two contrasting languages. *Clinical Linguistics & Phonetics*, *23*(6), 431–445. https://doi.org/10.1080/02699200902839800

Pereira, A. H., Gonçalves, A. B., Holz, M., Gonçalves, H. A., Kochhann, R., Joanette, Y., Zimmermann, N., & Fonseca, R. P. (2018). Influence of age and education on the processing of clustering and switching in verbal fluency tasks. *Dementia & Neuropsychologia*, *12*(4), 360–367. https://doi.org/10.1590/1980-57642018dn12-040004

Quaranta, D., Caprara, A., Piccininni, C., Vita, M. G., Gainotti, G., & Marra, C. (2016). Standardization, clinical validation, and typicality norms of a new test assessing semantic verbal fluency. *Archives of Clinical Neuropsychology*, *31*(5), 434–445. https://doi.org/10.1093/arclin/acw034

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2020). *Psych: Procedures for psychological, psychometric, and personality research*. CRAN. https://CRAN.R-project.org/package=psych

Roberts, P. M., & Dorze, G. L. (1997). Semantic organization, strategy use, and productivity in bilingual semantic verbal fluency. *Brain and Language*, *59*(3), 412–449. https://doi.org/10.1006/brln.1997.1753

Ross, T. P. (2003). The reliability of cluster and switch scores for the Controlled Oral Word Association Test. *Archives of Clinical Neuropsychology*, *18*(2), 153–164. https://doi.org/10.1093/arclin/18.2.153

Rosselli, M., Ardila, A., Salvatierra, J., Marquez, M., Luis, M., & Weekes, V. A. (2002). A cross-linguistic comparison of verbal fluency tests. *The International Journal of Neuroscience*, *112*(6), 759–776. https://doi.org/10.1080/00207450290025752

Roy-O'Reilly, M., & McCullough, L. D. (2018). Age and sex are critical factors in ischemic stroke pathology. *Endocrinology*, *159*(8), 3120–3131. https://doi.org/10.1210/en.2018-00465

Santos Nogueira, D., Azevedo Reis, E., & Vieira, A. (2016). Verbal Fluency Tasks: Effects of age, gender, and education. *Folia Phoniatrica et Logopaedica*, *68*(3), 124–133. https://doi.org/10.1159/000450640

Scheuringer, A., Wittig, R., & Pletzer, B. (2017). Sex differences in verbal fluency: The role of strategies and instructions. *Cognitive Processing*, *18*(4), 407–417. https://doi.org/10.1007/s10339-017-0801-1

Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). The Guilford Press.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, 772. https://doi.org/10.3389/fpsyg.2014.00772

Strauss, E., Sherman, E. M. S., Spreen, O., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). Oxford University Press.

Sung, K., Gordon, B., Yang, S., & Schretlen, D. J. (2013). Evidence of semantic clustering in letter-cued word retrieval. *Journal of Clinical and Experimental Neuropsychology*, *35*(10), 1015–1023. https://doi.org/10.1080/13803395.2013.845141

Suomi, K., Toivanen, J., & Ylitalo, R. (2006). *Fonetiikan ja suomen äänneopin perusteet*. Gaudeamus.

Taler, V., Johns, B. T., & Jones, M. N. (2020). A large-scale semantic analysis of verbal fluency across the aging spectrum: Data from the Canadian longitudinal study on aging. *The Journals of Gerontology:*

Series B, Psychological Sciences and Social Sciences, *75*(9), e221–e230. https://doi.org/10.1093/geronb/gbz003

Tallberg, I. M., Ivachova, E., Jones Tinghag, K., & Östberg, P. (2008). Swedish norms for word fluency tests: FAS, animals and verbs. *Scandinavian Journal of Psychology*, *49*(5), 479–485. https://doi.org/10.1111/j.1467-9450.2008.00653.x

Thiele, K., Quinting, J. M., & Stenneken, P. (2016). New ways to analyze word generation performance in brain injury: A systematic review and meta-analysis of additional performance measures. *Journal of Clinical and Experimental Neuropsychology*, *38*(7), 764–781. https://doi.org/10.1080/13803395.2016.1163327

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology*, *14*(2), 167–177.

Tröger, J., Linz, N., König, A., Robert, P., Alexandersson, J., Peter, J., & Kray, J. (2019). Exploitation vs. exploration-computational temporal and semantic analysis explains semantic verbal fluency impairment in Alzheimer's disease. *Neuropsychologia*, *131*, 53–61. https://doi.org/10.1016/j.neuropsychologia.2019.05.007

Troyer, A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology*, *22*(3), 370–378. https://doi.org/10.1076/1380-3395(200006)22:3;1-V;FT370

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138

Tyysteri, L. (2015). *Aamiaiskahvilasta ötökkätarjontaan. Suomen kirjoitetun yleiskielen morfosyntaktisten yhdyssanarakenteiden produktiivisuus* [Unpublished doctoral dissertation]. University of Turku.

Venegas, M. J., & Mansur, L. L. (2011). Verbal fluency: Effect of time on item generation. *Dementia & Neuropsychologia*, *5*(2), 104–107. https://doi.org/10.1590/S1980-57642011DN05020008

Vicente, S. G., Benito-Sánchez, I., Barbosa, F., Gaspar, N., Dores, A. R., Rivera, D., & Arango-Lasprilla, J. C. (2021). Normative data for verbal fluency and object naming tests in a sample of European Portuguese adult population. *Applied Neuropsychology: Adult*. Advance online publication. https://doi.org/10.1080/23279095.2020.1868472

Weiss, E. M., Ragland, J. D., Brensinger, C. M., Bilker, W. B., Deisenhammer, E. A., & Delazer, M. (2006). Sex differences in clustering and switching in verbal fluency tasks. *Journal of the International Neuropsychological Society*, *12*(4), 502–509. https://doi.org/10.1017/S1355617706060656

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal fluency: Language or executive function measure? *Applied Neuropsychology: Adult*, *23*(1), 29–34. https://doi.org/10.1080/23279095.2015.1004574

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wickham, H. (2020). *tidyr: Tidy messy data*. CRAN. https://CRAN.R-project.org/package=tidyr

Wickham, H., François, R., Henry, L., & Müller, K. (2019). *dplyr: A grammar of data manipulation*. CRAN. https://CRAN.R-project.org/package=dplyr

Appendix A. Instruction Manual for Administration and Scoring

Verbal Fluency Tasks

Administration

Verbal Fluency tasks should be administered in a quiet environment (such as home- or clinical setting). Responses should always be recorded for later verification and scoring. We recommend using a timer with a gentle audio cue every 15 seconds to make transcribing and scoring more straightforward. We also recommend pausing for possible clarification requests from the participant before moving on to the category in question. Time should be recorded from where the interlocutor finishes giving the instruction for the category ("You may begin"). Task instructions are as follows:

Category	*Finnish*	*English*
Semantic	Tässä tehtävässä pyydän sinua luettelemaan yhden minuutin aikana mahdollisimman monta yksittäistä tiettyyn kategoriaan kuuluvaa sanaa. Esimerkkinä: jos pyytäisin sinua luettelemaan hedelmiä voisit sanoa appelsiini, omena, luumu jne. Yritä sanoa mahdollisimman monta yksittäistä sanaa.	In this task I ask you to name as many individual words belonging to the same category as possible within one minute. For example, if I asked you to name different fruits, you could say orange, apple, plum, and so on. Try to name as many individual words as you can.
Animals	Tämä kategoria on ELÄIMET. Voit aloittaa.	This category is ANIMALS. You may begin.
Phonemic	Tässä tehtävässä pyydän sinua luettelemaan yhden minuutin aikana mahdollisimman monta yksittäistä tietyllä kirjaimella alkavaa sanaa. Nämä voivat olla mitä tahansa sanoja paitsi erisnimiä. Ainoastaan sanan ensimmäisellä kirjaimella on merkitystä. Esimerkkinä: jos pyytäisin tuottamaan sanoja T-kirjaimella, niin voisit sanoa, tuuli, taulu jne. Yritä sanoa mahdollisimman monta yksittäistä sanaa.	In this next category, I ask you to name as many words beginning with the same letter as possible within one minute. These words can be any kind of words, except proper names. The only thing that matters is the first letter of the word. For example, if I asked you to name words beginning with the letter T, you could say time, task, tunnel, and so on. Try to name as many individual words as you can.
/k/	Tämä kirjain on K. Voit aloittaa.	This letter is K. You may begin.
/a/	Tämä kirjain on A. Voit aloittaa.	This letter is A. You may begin.
/p/	Tämä kirjain on P. Voit aloittaa.	This letter is P. You may begin.

Note. When implementing multiple VF tasks for research or clinical purposes, we recommend randomizing task order to minimize the possible effect task order might have on the performance. We also recommend replacing the word *kirjain* [letter] with the word *äänne* [sound] in phonemic task instructions.

Scoring

Total Score

The total score is the sum of acceptable words generated during a 60-second trial, following the scoring instructions stated below. A word is counted into a trial if the participant begins producing it within the trial.

All novel words carrying semantically distinctive features are scored as separate items. Examples:

- Variations of the same animal: *lehmä, vasikka* [cow, calf]
- Compound words: *koripallo* [basketball]
- Slang words: *kundi* [lad]
- Dialectal variations
- Established loan words: *klisee* [cliché]
- Derivational words: *kirja* [book]*, kirjasto* [library]
- Numerals: kolme [three], kolmetoista [thirteen]

Note. If a participant generates multiple successive numerals in a PVF trial, results should be interpreted cautiously as numerals can be highly automatized.

Synonyms are considered individual items representing versatility of vocabulary. Examples:

- *naali, napakettu* [arctic fox], *aamiainen, aamupala* [breakfast]

Inflections are scored as errors (repetition). Example:

- *lintu* [bird,], *lintuja* [some birds]

Homonyms are scored as separate items when the participant indicates a semantic distinction. Example:

- *kuusi, kuusi* [spruce, six]

For semantic categories, specific rules might have to be considered. Example:

- Animal category:
  - Insects are accepted
  - Fantasy and imaginary animals are not accepted but scored as categorical errors (item out of category).

Errors

Errors are excluded from the total score. For example, if 25 words in total were generated, of which 5 are errors, the total score is 20.

Errors are classified into into three categories:

- Repetitions
- Categorical errors
    - Items out of category (e.g., *plant* under the semantic trial *animals* or wrong initial phoneme under phonemic trials)
    - Rule-break errors (i.e., proper names)
- Paraphasias and nonwords

For multilingual populations, add:

- Language intrusions (i.e., words in a different language than the target language, excluding established loan words).

Time Segments

The trial can be divided into time segments (e.g. four 15-second segments: 1–15 s, 16–30 s, 31–45 s, 46–60 s or two 30-second segments 1–30 s, 31–60). The score for each segment is calculated following the protocol described above. A word is counted to the time segment where the participant begins producing it.

Clustering

A cluster consists of <u>two or more successively generated words</u> that belong to the same semantic or phonemic category or subcategory following the rules stated below. Clusters are calculated following the rules of the smallest possible cluster to track all clusters participants generate. General and specific rules for phonemic and semantic clustering in both task types are described below. In the following, numbering refers to sample protocols (see below).

If a word connects sequential clusters, it belongs to the first cluster and is not counted to both clusters. Example:

- *cow, sheep, pig, <u>chicken</u>* (four-word cluster) -- *crow, sparrow* (two-word cluster). *Chicken* is a farm animal *and* a bird connecting sequential clusters.

Errors, regardless of error type, are <u>not</u> counted into clusters with the following exceptions:

- Non-sequential repetitions. Example:
  - *<u>rabbit</u>, fox, wolf* (three-word cluster) - *guinea pig, hamster, <u>rabbit</u>* (three-word cluster)
- Proper names that reflect the surrounding PVF phonemic clustering strategy. Example:
  - *aamu, aava, Aapeli, aamiainen* [morning, open, Aapeli, breakfast] (four-word cluster)

1 Phonemic Clustering

A phonemic cluster consists of two or more successively generated words that share any of the characteristics stated below. Phonemic clusters in semantic and phonemic VF tasks follow the same rules.

1.1 Word Initial Phonemes

In phonemic and semantic fluency, words beginning with the same <u>two</u> phonemes always form a cluster. Example:

- *<u>ka</u>la* [fish], *<u>ka</u>mina* [stove] (two word cluster)

In addition, in semantic fluency, three or more consecutive words with the same word initial phoneme form a phonemic cluster regardless if there are one or two shared initial phonemes. Examples:

- *<u>k</u>issa* [cat], *<u>ko</u>libri* [hummingbird], *<u>ka</u>meli* [camel] (three-word cluster)
- *<u>k</u>ili* [kid], <u>kä</u>ärme [snake], *<u>kä</u>rpänen* [fly], *<u>ko</u>ira* [dog] (four-word cluster)

1.2 Rhyming Words

Words with identical phonemes in two or more word-final sequences, typically final stressed vowels and all following sounds, form a cluster. Example:

- *naakka* [jackdaw], *harakka* [magpie] (two-word cluster)

Compound words ending in the same word are considered rhyming words. Example:

- *koulu* [school], *keskikoulu* [middle school], *kansakoulu* [grammar school]

1.3 First and Last Phonemes

Structurally similar words that differ only by one sound (regardless of spelling) form a cluster. Example:

- */ale/, /ase/.* (two-word cluster)

1.4 Consonant Structure

Structurally similar words that differ only by their vowel sounds form a cluster. Example:

- */pilli/, /pulla/* (two-word cluster)

1.5 Homonyms

Identical words with different meanings, as indicated by the participant, are counted as a cluster. Example:

- *kuusi [six], kuusi [spruce]* (two-word cluster)

2 Semantic Clustering

A semantic cluster consists of two or more successively generated words that belong in the same semantic subcategory. Under the semantic condition, these are typically taxonomic subcategories of animals (e.g., birds or big cats), but they can also be formed by environmental semantic connections (e.g., farm animals), geographical semantics (e.g., African animals), or visual semantics (e.g., snake, eel). In the phonemic trials, semantic categories can be based on semantic connections such as cooking utensils, clothes, or beverages, as long as they follow the rules stated below. Semantic clusters in semantic and phonemic VF tasks follow the same rules.

Categories are based on naturally occurring clusters for each participant. Thus, semantic clusters are sometimes more overarching categories (e.g., birds in or items found in the kitchen) and sometimes more detailed categories (e.g., birds of prey, water birds and cooking utensils, silverware).

2.1 Strong Cluster: Three or More Words

Three or more words belonging to the same subgroup always form a cluster. Examples:

- *bear, wolf, lynx* (three-word cluster / predators) -- *rabbit* (non-clustering single word / prey*)*.

- *crow, sparrow, duck* (three-word cluster/birds) -- *falcon, eagle, vulture* (three-word cluster/birds of prey).

2.2 Weak Cluster: Two Words

Two words form a cluster when there are no words from a more overarching category surrounding the two words. Example:

- *monkey* (non-clustering single word) -- *falcon, eagle* (two-word cluster) -- *cat* (non-clustering single word).

If two words belonging to the same subgroup are surrounded by single words belonging to the same overarching category, these are counted as one cluster. Examples:

- *lion, tiger, antelope* (three-word cluster)
- *lion, tiger, hippo, antelope* (four-word cluster).
- *hippo, lion, tiger, antelope, elephant* (five-word cluster).

If two weak clusters from the same overarching category are found one after another, these are counted as separate clusters. Example:

- *lion, tiger* (two-word cluster) -- *antelope, gazelle* (two-word cluster).

If two words belonging to the same subgroup are surrounded by single words that belong to another subgroup of the same overarching category, there are two subgroups. Example:

- *gazelle* (non-clustering single word) -- *lion, tiger* (two-word cluster) -- *antelope* (non-clustering single word).

2.3 Word Associations

Associations do not form semantic clusters. Examples:

- *sammakko* [frog] (non-clustering single word) -- *kärpänen* [fly] (non-clustering single word).
- *kallio* [rock] (non-clustering single word) -- *kova* [hard] (non-clustering single word).

Antonyms are counted as associations. Example:

- *kylmä* [cold] (non-clustering single word) -- *kuuma* [hot] (non-clustering single word).

To determine if words form a cluster or an association, one can add more words to the category. If more words could be added, they form a cluster. If words can not be added, the connection is considered a word association. Examples:

- *poimia* [pick] (non-clustering single word) -- *puolukka* [lingonberry] (non-clustering single word).
- *pyykki* [laundry], *pulveri* [detergent] (two-word cluster) as, for example, *huuhteluaine* [softener] and *pesukone* [washing machine] could semantically be added to the category “laundry”.

2.4. Special Rules for Semantic Clustering in Phonemic VF Tasks

Runoff sentences and alliteration are considered as associations and are not calculated as semantic clusters. Example:

- *kukko* [a rooster] (non-clustering single word) -- *kiekui* [crowed] (non-clustering single word) -- *katolla* [on the roof] (non-clustering single word) .

Adjectives or verbs do not form a cluster per se, but when they share semantic properties, they do. Examples:

- *kirmata* [frolic] -- *kaivaa* [dig] -- *kinata* [argue] (verbs are not semantically linked, non-clustering single words).
- *paloitella* [cut into pieces], *paistaa* [fry], *pilkkoa* [chop] (verbs relate to cooking, three-word cluster).

Particles form clusters as they carry no semantic meaning themselves but share similar features as function words. Example:

- *koska* [because], *kun* [when] (two-word cluster).

Switching

A transition between two words that do not belong in the same cluster, including single word transitions, is a switch. Examples:

- dog -- cow -- eagle (three single word clusters, <u>two switches</u>)
- *kala* [fish], *kamina* [stove], *kammata* [comb] -- *kiperä* [tricky], *kirmata* [frolic] (one three word cluster, one two word cluster, <u>one switch</u>)

Sample Protocols and Instructions on Training Raters

To ensure consistent clustering across trials it is beneficial to code multiple practice trials following the clustering rules stated above before rating actual data. We recommend that raters compare and discuss their results to solve potential conflicts. The following example trials can be used as reference material.

Example Trial of Task Congruent Clustering in Phonemic Category /k/

Task congruent (phonemic) clusters		Number of clusters	Cluster size (min. size 2)	Clustering rule applied
kissa	[cat]			non-clustering single word
komea *koiras*	[handsome] [male animal]	1	2	1.1 Word initial phonemes
kaamea	[horrendous]			Non-clustering single word
kurkkia	[peek]			Non-clustering single word
kynä	[pen]			Non-clustering single word
kumi	[eraser]			Non-clustering single word
kaunis	[beautiful]			Non-clustering single word
komea *korea*	[handsome] [decorated]	1	2	1.1 Word initial phonemes 1.3 First and last phonemes *NB. komea* [handsome] is a non-sequential repetition. It is scored 0 for the total score but is counted in the cluster.
kauhea *kamala* *karu* *karkki* *karhea* *kauha*	[awful] [horrible] [bare] [candy] [rough] [ladle]	1	6	1.1 Word initial phonemes
kippo *kuppi*	[scoop] [cup]	1	2	1.4 Consonant structure *NB.* Words connecting two clusters *kuppi* [cup] is counted to the first cluster
kurki	[crane]			Non-clustering single word
kirjosieppo	[flycatcher]			Non-clustering single word

Number of task congruent clusters for this trial is 4.

Mean cluster size for this trial is 3.

Example Trial of Task Discrepant Clustering in Phonemic Category /k/

Task discrepant (semantic) clusters		Number of clusters	Cluster size (min. size 2)	Clustering rule applied
kissa	[cat]			Non-clustering single word
komea	[handsome]			Non-clustering single word
koiras	[male animal]			Non-clustering single word
kaamea	[horrendous]			Non-clustering single word
kurkkia	[peek]			Non-clustering single word
kynä *kumi*	[copy] [eraser]	1	2	2.2 Weak cluster: 'relates to writing'.
kaunis *komea* *korea*	[beautiful] [handsome] [decorated]	1	3	2.4 Special rules for semantic clustering in PVF; adjectives sharing semantic properties: 'relates to looks'
kauhea *kamala* *karu*	[awful] [horrible] [bare]	1	a. 3 b. 2	2.4. Special rules for semantic clustering in PVF; adjectives sharing semantic properties: 'carry similar connotations' Two acceptable options: a. 2.1 Strong cluster: *kauhea* [awful], *kamala* [horrible], *karu* [bare] b. 2.2 Weak cluster: *kauhea* [awful], *kamala* [horrible] two-word cluster; Non-clustering single word: *karu* [bare]
karkki	[candy]			Non-clustering single word
karhea	[rough]			Non-clustering single word
kauha *kippo* *kuppi*	[ladle] [scoop] [cup]	1	3	2.1 Strong cluster: 'items in the kitchen'
kurki *kirjosieppo*	[crane] [flycatcher]	1	2	2.2 Weak cluster: 'birds'

Number of task discrepant clusters for this trial is 5.

Mean cluster size for this trial is a. 2.6 or b. 2.4.

Example Trial Of Task Congruent Clustering In Semantic Category 'Animals'

Task congruent (semantic) clusters		Number of clusters	Cluster size (min. size 2)	Clustering rule applied
koira	[dog]	1	2	2.2 Weak cluster: 'pets'.
kissa	[cat]			
hevonen	[horse]			Non-clustering single word
hämähäkki	[spider]	1	2	2.2 Weak cluster: 'insect'
kärpänen	[fly]			
käärme	[snake]	1	2	2.2 Weak cluster: visual semantic connection 'long, slithering animal'
ankerias	[eel]			
hevonen	[horse]	1	5	2.1 Strong cluster:
lehmä	[cow]			'farm animal' including a variation of the same animal (cow, calf) with last word (chicken) connecting sequential clusters (1.3)
vasikka	[calf]			
sika	[pig]			
kana	[chicken]			*NB. hevonen* [horse] is a non-sequential repetition. It is scored 0 for the total score but is counted in the cluster.
kirjosieppo	[flycatcher]			Non-clustering single word
kuikka	[loon]	1	3	2.1 Strong cluster: 'aquatic bird'
pelikaani	[pelican]			
joutsen	[swan]			
varpunen	[sparrow]			Non-clustering single word
hamsteri	[hamster]	1	a. 4	2.1 Strong cluster
rotta	[rat]			Two acceptable options:
hiiri	[mouse]			a. 4 -word cluster 'rodent' including *hamsteri* [hamster], *rotta* [rat], *hiiri* [mouse], *orava* [squirrel]
orava	[squirrel]		b. 3	b. 3 -word cluster 'pet' including *hamsteri* [hamster], *rotta* [rat], *hiiri* [mouse]; Non-clustering single word *orava* [squirrel].

Number of task congruent clusters for this trial is 4.

Mean cluster size for this trial is a. 3.0 or b. 2.8.

Example Trial Of Task Discrepant Clustering In Semantic Category ‘Animals’

Task discrepant (phonemic) clusters		Number of clusters	Cluster size (min. size 2)	Clustering rule applied
koira	[dog]			Non-clustering single word
kissa	[cat]			Non-clustering single word
hevonen	[horse]			Non-clustering single word
hämähäkki	[spider]			Non-clustering single word
kärpänen *käärme*	[fly] [snake]	1	2	2.1 Word initial phonemes
ankerias	[eel]			Non-clustering single word
hevonen	[horse]			Non-clustering single word
lehmä	[cow]			Non-clustering single word
vasikka	[calf]			Non-clustering single word
sika	[pig]			Non-clustering single word
kana *kirjosieppo* *kuikka*	[chicken] [flycatcher] [loon]	1	3	1.1. Word initial phonemes: Three consecutive words with the same word initial phoneme in semantic fluency
pelikaani	[pelican]			Non-clustering single word
joutsen	[swan]			Non-clustering single word
varpunen	[sparrow[			Non-clustering single word
hamsteri	[hamster]			Non-clustering single word
rotta	[rat]			Non-clustering single word
hiiri	[mouse]			Non-clustering single word
orava	[squirrel]			Non-clustering single word

Number of task discrepant clusters for this trial is 2.

Mean cluster size for this trial is 2.5.

Appendix B.

Model Summaries with Effect Size Estimates of Semantic and Phonemic Fluency Tasks and Supplementary Information on Age Bands

Overall Performance in Verbal Fluency Tasks.

Predictors	Total score				
	Estimates	95% CI	*t*	*p*	*Cohen's d*
(Intercept)	18.87	17.44–20.29	25.94	**<0.001**	
Task type [semantic]	8.47	6.62–10.33	8.96	**<0.001**	0.91
Education (centered)	4.55	2.18–6.93	3.75	**<0.001**	1.11
Age (centered)	-1.00	-2.20–0.20	-1.63	0.104	0.48
Gender [male]	-5.06	-7.69– -2.44	-3.78	**<0.001**	-1.11
Task type [semantic] * Education (centered)	-0.21	-3.31–2.88	-0.14	0.892	-0.04
Task type [semantic] * Age (centered)	0.53	-1.03–2.10	0.67	0.502	0.20
Task type [semantic] * Gender [male]	1.03	-2.38–4.44	0.59	0.553	0.18
Random Effects					
σ^2	12.79				
τ_{00} ID	13.83				
τ_{11} ID.tasktypesem	13.52				
ρ_{01} ID	-0.42				
ICC	0.53				
N_{ID}	50				

Observations	200
Marginal R^2 / Conditional R^2	0.479 / 0.754

Performance in All Four Verbal Fluency Tasks, Phonemic Task /a/ as a Reference Level.

	Total score				
Predictors	Estimates	95% CI	*t*	*p*	*Cohen's d*
(Intercept)	14.68	13.07–16.29	17.90	**<0.001**	
Task [animal]	11.28	9.91–12.65	16.13	**<0.001**	2.66
Task [k]	4.70	3.33–6.07	6.72	**<0.001**	1.11
Task [p]	2.76	1.39–4.13	3.95	**<0.001**	0.65
Random Effects					
σ^2	12.23				
τ_{00} ID	21.39				
ICC	0.64				
N_{ID}	50				
Observations	200				
Marginal R^2 / Conditional R^2	0.341 / 0.760				

Descriptive Statistics of Total Scores for Semantic and Phonemic Verbal Fluency Tasks in Three Age Bands

Task type	49-59[a] *N=17*	60-69[a] *N=23*	70-79[a] *N=10*	All age groups N=50
Semantic				
Mean (*SD*)	26.29 (6.91)	26.61 (5.69)	23.90 (4.41)	25.96 (5.90)

Range	17–43	12–37	16–29	12–43
Phonemic				
Mean (*SD*)	18.37 (6.68)	17.29 (5.82)	14.83 (4.79)	17.17
Range	4–35	6–33	5–24	4–35

[a] Group age range in years

Performance in 15-Second Segments, First Segment (1–15 Sec) and Phonemic Task Type Used as Reference Levels.

Predictors	**Words produced in 15-second segments**				
	Estimates	95% CI	*t*	*p*	Cohen's d
(Intercept)	6.35	5.92–6.78	28.84	**<0.001**	
Time segment [16–30 sec]	-2.26	-2.63– -1.89	-12.13	**<0.001**	-0.92
Time segment [31–45 sec]	-2.82	-3.19– -2.45	-15.14	**<0.001**	-1.15
Time segment [46–60 sec]	-3.14	-3.51– -2.77	-16.86	**<0.001**	-1.28
Task type [Semantic]	4.41	3.83–5.00	14.82	**<0.001**	1.87
Time segment [16–30 sec] * Task type [Semantic]	-2.44	-3.17– -1.71	-6.55	**<0.001**	-0.50
Time segment [31–45 sec] * Task type [Semantic]	-3.00	-3.73– -2.27	-8.05	**<0.001**	-0.61
Time segment [46–60 sec] * Task type [Semantic]	-3.42	-4.15– -2.69	-9.18	**<0.001**	-0.70

Random Effects

σ^2	2.60
τ_{00} ID	1.55
τ_{11} ID.tasktypesem	0.96
ρ_{01} ID	-0.41
ICC	0.37
N ID	50
Observations	800
Marginal R^2 / Conditional R^2	0.469 / 0.667

Task Congruent Clusters and Switches Predicting Total Score in Phonemic and Semantic Verbal Fluency Tasks.

	Total score: Phonemic				**Total score: Semantic**				
Predictors	Esti-mates	95% CI	*t*	*p*	Esti-mates	95% CI	*t*	*p*	*d*
(Intercept)	19.96	19.09–20.83	44.95	**<0.001**	26.62	25.57–27.67	51.19	**<0.001**	
Switches (centered)	6.22	5.43–7.02	15.39	**<0.001**					2.81
Cluster size (centered)	1.72	0.91–2.54	4.14	**<0.001**					0.82
Switches (centered) * Cluster size (centered)	-0.67	-1.36–0.02	-1.91	0.057					-0.38
Education (centered)					2.98	0.88–5.08	2.87	**0.006**	0.37
Switches (centered)					7.59	6.11–9.06	10.40	**<0.001**	2.88

Cluster size (centered)		4.65	3.29–6.02	6.86	**<0.001**	1.45
Education (centered) * Switches (centered)		1.79	-1.15–4.74	1.23	0.226	0.30
Education (centered) * Cluster size (centered)		0.58	-2.15–3.32	0.43	0.669	-0.01
Switches (centered)* Cluster size (centered)		0.50	-1.01–2.01	0.67	0.507	0.07
(Education (centered) * Switches (centered) * Cluster size (centered)		1.49	-1.54–4.51	0.99	0.326	0.12
Random Effects						
σ^2	4.63					
τ_{00}	6.38_{ID}					
ICC	0.58					
N	49_{ID}					
Observations	135	50				
Marginal R^2 / Conditional R^2	0.600 / 0.832	0.881 / 0.861				

Number of Task Discrepant Clusters Predicting Total Scores in Phonemic and Semantic Tasks

	Total score: Phonemic				**Total score: Semantic**				
Predictors	Estimates	95% CI	*t*	*p*	Estimates	95% CI	*t*	*p*	*d*
(Intercept)	15.74	12.63–18.84	9.92	**<0.001**	26.11	21.26–30.95	11.13	**<0.001**	
Number of semantic clusters	1.30	0.42–2.18	2.89	**0.004**					0.69
Trial order	-0.30	-1.68–1.08	-0.43	0.669					-0.10
Number of semantic clusters * Trial order	0.02	-0.52–0.56	0.07	0.941					0.02
Number of phonemic clusters					0.52	-1.89–2.93	0.44	0.662	0.18
Random Effects									
σ^2	7.53								
τ_{00}	16.88 ID								
ICC	0.69								
N	50 ID								
Observations	111				26				
Marginal R^2 / Conditional R^2	0.165 / 0.742				0.008 / -0.033				

Lehtinen, N., Kautto, A., & Renvall, K. (2024)
Frequent native language use supports phonemic and semantic verbal fluency in L1 and L2: An extended analysis of verbal fluency task performance in an L1 language attrition population.
International Journal of Bilingualism

II

Article

The International Journal of Bilingualism

International Journal of Bilingualism
2024, Vol. 28(5) 884–906

DOI: 10.1177/13670069231193727
journals.sagepub.com/home/ijb
Sage

Frequent native language use supports phonemic and semantic verbal fluency in L1 and L2: An extended analysis of verbal fluency task performance in an L1 language attrition population

Nana Lehtinen
University of Turku, Finland; University of California, Berkeley, USA

Anna Kautto and
Kati Renvall
University of Turku, Finland

Abstract

Purpose: Verbal fluency (VF) tasks are an efficient tool for exploring lexical retrieval. Attriters generally produce fewer words in semantic categories than monolinguals, but the processes underlying attriters' lexical retrieval remain unclear. Phonemic tasks are scarcely applied in language attrition studies. We aim to identify processes underlying attriters' phonemic verbal fluency (PVF) and semantic verbal fluency (SVF) performance by extending the analysis beyond total scores in the first (L1) and second (L2) language.
Design: We modeled total scores and temporal aspects of task performance in L1 and L2 for the attriters, with L1 use and length of residence (LoR) as predictors. We analyzed the number and types of errors between languages and compared L1 task performance with a matched monolingual group.
Data and analysis: Attriters' ($N=38$) phonemic and semantic task performance in L1 (Finnish) and L2 (English) were modeled, and results were contrasted to a matched monolingual control group ($N=50$).
Findings: Attriters demonstrated strong proficiency in L1 and similar lexical retrieval processes in L1 and L2 after 20 years of immersive L2 exposure. Frequent L1 use supported overall performance but slowed performance down in both languages. Compared with monolinguals, attriters show a disadvantage in SVF, but not due to slower initiation. Instead, attriters rely more on rapid L1 retrieval than monolinguals.
Originality: Prior research on VF tasks in language attrition populations has mainly focused on SVF total scores in L1. Our investigation explores underlying mechanisms of L1 and L2 task

Corresponding author:
Nana Lehtinen, Department of Psychology and Speech-Language Pathology, University of Turku, Turku 20014, Finland.
Email: nanleh@utu.fi

performance in PVF and SVF within a language attrition population and contrasts L1 performance to a matched monolingual group performance.
Implications: This study demonstrates the importance of investigating L2 alongside L1 in language attrition studies for a holistic approach to language processing strategies. Our findings enhance understanding of the processes underlying VF task performance, emphasizing the significance of VF task early-stage performance.

Keywords
Language attrition, verbal fluency task, bilingualism, semantic verbal fluency, phonemic verbal fluency

Introduction

Within the broad field of bilingualism, the gradual, non-pathological change in the native language (L1) due to reduced use and exposure is referred to as L1 language attrition (Schmid, 2011b; Schmid & Köpke, 2017). *Language attrition* is a complex phenomenon that affects all aspects of language and communication. It can be observed as a series of subtle but consistent changes in L1 that are not always directly observable in language performance. L1 attrition primarily occurs during online processing and may lead to reduced fluency, speaker insecurities, reduced accuracy, and overall simplification in L1 use, as well as interference from L2 (e.g., the integration of L2 elements into L1) (Jarvis, 2019, Chapter 21, pp. 241–245; Schmid, 2011b; Schmid & Köpke, 2009). The most investigated domain for language attrition is the lexicon. Fluency and fluency-related phenomena are studied by measuring speech rate, hesitations, filled pauses, repetitions, and self-repairs in free speech. Verbal fluency (VF) tasks are often used for a more controlled approach to lexical fluency (Jarvis, 2019, Chapter 21, pp. 248–249; Schmid & Köpke, 2009). Observed changes in online task performance are not necessarily representations of systematic grammatical restructuring, erosion, or loss of language competence (Sharwood Smith, 2019, Chapter 8, p. 85). However, language attriters tend to apply categorical rules or features with optionality and exhibit greater variation in their L1 performance than monolinguals (Schmid & Cherciov, 2019, Chapter 23, p. 267; Schmid & Dusseldorp, 2010). For a comprehensive overview of language attrition studies and findings, see Schmid and Köpke (2019, pp. 509–541).

Building on previous language attrition research, this study examines the L1 of immigrants who share a language history characterized by late sequential bilingualism, immersion in an L2 environment, and minimal exposure to their native L1 over an extended period (Schmid, 2011b, 2019; Schmid & Jarvis, 2014; Schmid & Köpke, 2017). In this study, we refer to participants who are late sequential bilinguals fully immersed in their second language as “attriters” to differentiate them from other bilingual groups (the term bilingualism is used for the sake of brevity while encompassing multilingualism). We use the more general term “bilinguals” to include individuals with various language backgrounds and sociolinguistic contexts. Nonetheless, we acknowledge the controversy of using the term attriter to refer to a participant group for which evidence of attrition has not been detected (Kasparian & Steinhauer, 2017; Schmid & Köpke, 2017).

Many language attrition studies focus on L1 performance and compare performance between monolingual speakers and L1 attriters (Schmid, 2019; Schmid & Dusseldorp, 2010; Schmid & Yılmaz, 2018). It is widely recognized that languages in the bilingual brain interact continuously in a bidirectional process, altering the cognitive and neural dynamics of both languages and resulting in shifts in individual language dominance patterns over time (Grosjean, 2013; Gurunandan et al., 2023; Laine & Lehtonen, 2018; Linck & Kroll, 2019, Chapter 9, p. 97; Treffers-Daller, 2019). These shifts are influenced by sociolinguistic parameters such as language exposure and use and can, to some extent, be observed in language performance (Treffers-Daller, 2019; Yılmaz, 2019). The language history of language attrition populations typically involves a rapid decline in L1 input coupled with full immersion in L2 upon immigration. Previous studies show that within language attrition populations, some attriters perform at native levels, while others exhibit noticeable changes in L1 performance (Schmid & Yilmaz, 2018). Examining L2 performance alongside L1 can provide insights into individual factors contributing to L1 attrition or retention, as well as L2 acquisition (Bylund & Ramírez-Galan, 2016; Gurunandan et al., 2023; Laine & Lehtonen, 2018; Runnqvist & Costa, 2012; Schmid & Köpke, 2017; Schmid & Yilmaz, 2018).

To investigate the impact of attriters' shared language history on task performance, many language attrition studies investigate the influence of two variables: length of residence (LoR, i.e., time deprived of native L1 exposure and extensive exposure to L2) and the frequency of L1 use in daily life. It has been shown that LoR predicts task performance when the minimum LoR is less than 10 years, with the effect stabilizing after this period (Opitz, 2011; Schmid, 2011a, 2019). This stabilization has been partially attributed to the initial years of immersion which typically coincide with a period of rapid L2 learning in an immersive environment (Linck & Kroll, 2019, Chapter 9, p. 95). However, while the added prevalence of L2 partly triggers L1 attrition, it is not directly linked to L2 acquisition or level of L2 proficiency (Yılmaz, 2019, Chapter 26, p. 307). Unfortunately, we are not aware of language attrition studies that directly investigate the impact of LoR on L2 performance.

The frequency of L1 use has not been found to systematically impact L1 performance (for an overview, see Schmid, 2019, Chapter 25, pp. 294–295). However, L1 use in different life domains can have different implications for changes in L1 performance. On one hand, frequent use of L1 in a professional setting has been found to have a positive impact on low-frequency word retrieval (Yilmaz & Schmid, 2012), verbal fluency, general proficiency (c-test), grammaticality judgment task, and lexical diversity (Schmid & Dusseldorp, 2010). On the other hand, frequent L1 use in informal interactions with peers (attriters) has been associated with increased variability in the phonemic domain (De Leeuw et al., 2010). Although LoR and frequent L1 use do not systematically predict attrition as separate variables (Schmid, 2019; Schmid & Dusseldorp, 2010; Schmid & Yilmaz, 2018), it has been shown that the impact of LoR can become significant with very little L1 use (de Bot & Clyne, 1994).

Consequently, it has been suggested that attriters who have spent a long time in an L2 environment and use L1 less often exhibit signs of attrition at an individual level. In contrast, speakers who have spent extended time in an L2 environment and who use L1 frequently with peers reflect a contact-induced language change in the community on a larger scale (Schmid, 2011a; Schmid & Köpke, 2017). In the current study, we investigate the impact of LoR and frequency of L1 use on verbal fluency task performance separately, as well as their interactions, to reliably include both extralinguistic variables.

Verbal fluency tasks

VF tasks are widely applied across various populations and languages (Goral, 2004; Olabarrieta-Landa et al., 2017). In a VF task, participants are given a criterion and asked to generate words within a predetermined time frame (typically 60 seconds) (Strauss et al., 2006). The common VF task types are phonemic (PVF) and semantic (SVF).

In a PVF task, participants generate words starting with a specific letter or phoneme, and three trials are typically included for reliable fluency assessment (Oberg & Ramírez, 2006; Strauss et al., 2006; Tombaugh et al., 1999). In studies conducted in English, the letter combination /f/, /a/, and /s/ is frequently used. The letters are initially selected based on their grouping as "easy letters" (Borkowski et al., 1967; Ross, 2003; Strauss et al., 2006), and the frequent use of these letters in the literature allows comparison to other datasets. For a more straightforward letter selection, a language-specific "high-frequency dictionary approach" can be used. In this approach, letters are selected based on their word-initial frequency in the target language, such as /k/, /a/, and /p/ in Finnish (Lehtinen et al., 2021; Mardani et al., 2020; Oberg & Ramírez, 2006; Schmid, 2011a; Tombaugh et al., 1999).

In SVF tasks, participants generate words within a specific semantic category. Unlike PVF tasks, SVF tasks typically focus on one category. Combining data from semantic categories is complex. As unique life experiences influence semantic memory organization, varied demographic and cultural settings result in diverse category sizes and content between populations (e.g., "items in a supermarket" can evoke varied semantic representations in different demographic groups) (Abwender et al., 2001; Olabarrieta-Landa et al., 2017; Roberts & Le Dorze, 1997; Rosselli et al., 2002; Strauss et al., 2006; Troyer, 2000). The category "animals" has been shown to be relatively neutral regarding culture and language (e.g., Pekkala et al., 2009). The category is widely used in linguistic- and clinical studies and in neuropsychological test batteries across languages (for an overview, see Strauss et al., 2006).

PVF and SVF tasks are commonly used to study various aspects of bilingualism, such as lexical access, vocabulary knowledge, dominance pattern of languages, the role of executive functions, and cross-linguistic fluency strategies (Luo et al., 2010; Marsh et al., 2019; Patra et al., 2020; Roberts & Le Dorze, 1997; Rosselli et al., 2002; Schmid & Köpke, 2009). PVF tasks engage strategic cognitive organization, initiation, inhibition, and maintenance of effort as participants conduct a non-routine search for words based on specific lexical representation (i.e., first letter) without the support of the hierarchical organization of semantic memory (Barry et al., 2008; Santos Nogueira et al., 2016; Strauss et al., 2006). In contrast, the SVF task requires a systematic lexical-semantic search—a relatively automatic process that resembles the everyday use of language (e.g., generating a shopping list). SVF task performance mainly relies on semantic categorization, hierarchical mental lexicon, and memory organization (Luo et al., 2010; Patra et al., 2020; Strauss et al., 2006).

This study in the context of existing literature

VF task performance is typically evaluated by the total number of correct responses (Strauss et al., 2006). A growing body of cross-disciplinary research aims to enhance the analytical power of VF tasks through more detailed data analyses (e.g., Becker & de Salles, 2016; Thiele et al., 2016). Based on these studies, Lehtinen et al. (2021) generated guidelines for a systematic approach to VF task administration and analysis, including scoring and calculating total scores, error analysis, investigating temporal parameters (e.g., words generated in 10-, 15-, or 30-second segments within the total time), and exploring clustering and switching strategies. In this study, we analyze total scores, errors, and temporal parameters in PVF and SVF performance in a Finnish-English language attrition population following these guidelines. Next, we will briefly review previous literature on VF task analysis in bilingual contexts before describing the present study in detail.

Total scores. PVF task performance measured by total score has been found to vary from a comparable performance between mono- and bilinguals (Rosselli et al., 2000, 2002; Soltani et al., 2021) to bilinguals performing better than monolinguals (Patra et al., 2020). Similar PVF performance between mono- and bilinguals has been interpreted as less language interference in PVF compared with SVF (Rosselli et al., 2000, 2002). Stronger PVF performance has been linked to superior executive functions related to inhibiting language interference while switching between languages (Ljunberg et al., 2013; Luo et al., 2010; Marsh et al., 2019; Patra et al., 2020; Sandoval et al., 2010). However, PVF tasks are scarcely applied in language attrition studies (Jarvis, 2019, Chapter 21, p. 249; Schmid & Köpke, 2009).

In SVF tasks, bilingual populations consistently generate fewer words than monolinguals. This disadvantage in total scores has been attributed to language interference, weaker connections between lexical representations, or smaller vocabulary (Gollan et al., 2002; Rosselli et al., 2000; Sandoval et al., 2010; Schmid & Köpke, 2017; Soltani et al., 2021; Yılmaz & Schmid, 2018). Studies on language attrition populations line up with these findings, with attriters performing more poorly than monolinguals in SVF, though often with a relatively small effect size (e.g., Badstübner, 2011; Dostert, 2009; Opitz, 2011; Schmid & Dusseldorp, 2010; Schmid & Jarvis, 2014; Schmid & Keijzer, 2009; Schmid & Köpke, 2009). It has been shown that SVF total scores have limited predictive power in profiling individual speakers as attriters or monolinguals, and total scores are largely unrelated to LoR frequency of L1 use (Schmid, 2011a; Schmid & Jarvis, 2014; Schmid & Köpke, 2009). Therefore, a more comprehensive approach to VF data analyses is called for.

Errors. Errors are a part of VF task performance in all populations (Crowe, 1998; Sandoval et al., 2010). In bilingual environments, language intrusions are of particular interest (Gollan et al., 2002) as cross-language intrusions are directly connected to language interference (Gollan et al., 2011). Their presence can also indicate language dominance with more frequent language intrusion errors generated in the non-dominant language (Sandoval et al., 2010). In a language attrition population, Badstübner (2011) found that attriters generated more errors than monolinguals in SVF and described these errors as direct L2 intrusions and incorrect lexical items (such as partial recall of an L1 word). They detected no significant difference in the number of errors between the groups in PVF.

Temporal parameters. Analyzing the temporal parameters of VF tasks involves calculating the number of correct words produced within shorter time segments of the total time (e.g., four 15-second time segments within the 60 seconds). Typically, most words are generated in the early segments of the task across all populations. Approximately half of the words are produced in the first 15 seconds, facilitated by a semi-automatic rapid retrieval process that relies on semantic memory. As time progresses, a more effortful retrieval strategy is employed, engaging strategic executive processes, such as monitoring performance to avoid repetitions. This results in slower word-finding with fewer and more infrequent words (Fernaeus et al., 2008; Lehtinen et al., 2021; Sandoval et al., 2010; Venegas & Mansur, 2011). Thus, group differences in the early segments of the task suggest variations in language knowledge and lexical retrieval, while differences in later segments indicate discrepancies in executive control (Fernaeus & Almkvist, 1998; Fernaeus et al., 2008; Gurunandan et al., 2023; Luo et al., 2010).

Studies have consistently shown that bilinguals generate fewer words than monolinguals during the initial stages of a VF task, but the difference between the groups tends to diminish as the trial progresses (Luo et al., 2010; Sandoval et al., 2010). This pattern is interpreted to reflect more language interference in the early stages of the task, where bilinguals may produce high-frequency words they know in both languages (e.g., “cat”) rather than low-frequency words they may only know in one language (e.g., “bobcat”) (Luo et al., 2010; Sandoval et al., 2010). Interestingly, similar neural activation patterns have been observed during L1 and L2 performance, suggesting a common bilingual effect in both languages (Gurunandan et al., 2023). Regarding language attrition studies, attriters have been shown to retrieve words more slowly than monolinguals in L1 SVF (Jarvis, 2019, Chapter 21, p. 243; Schmid & Jarvis, 2014; Schmid & Köpke, 2009). This finding has been linked to an increased load of inhibiting L2 competitors (Schmid & Jarvis, 2014; Yilmaz & Schmid, 2018) consistent with previous studies on bilingual VF performance (Gollan et al., 2011; Luo et al., 2010; Sandoval et al., 2010). However, highly automatic language skills are considered to be more resistant to language attrition (Goral, 2004; Segalowitz, 1991).

The present study

In the present study, we examine the processes that affect and potentially hinder optimal performance in L1 and L2 PVF and SVF tasks among a group of L1 language attriters who self-report as balanced bilinguals with a slight preference for L2. We conduct a systematic analysis for total scores, errors, and temporal parameters guided by the methodology proposed by Lehtinen et al. (2021). We also investigate how the LoR and frequency of L1 use affect task performance in both L1 and L2. To detect differences between attriters and monolingual speakers, we compare the L1 performance of attriters with a matched L1 monolingual group previously studied by Lehtinen et al. (2021).

Research questions

The research questions and hypotheses for the current study are as follows:

1. How do language attriters perform in PVF and SVF tasks for their first (L1) and second (L2) languages, and to what extent do LoR and frequency of L1 use affect their performance on the measures listed below?
2. How is the performance of language attriters in PVF and SVF tasks in their first language (L1) compared with that of a matched monolingual group across the measures listed below?

Measures:

a. Total scores
b. Errors: number, frequency, and distribution of error types
c. Temporal parameters: performance change during the task (measured as the number of words generated in four 15 seconds segments)
d. Ability to generate words rapidly in the early stages of the task (measured as the proportion of words generated within the initial 15-second interval of a task relative to the overall word count)

We will be referring to the different aspects of these research questions by referencing their numbers (RQ1 and RQ2) and letters (a, b, c, d) to specify which aspects of the data analysis is relevant.

Hypotheses

As stated above, while both, PVF and SVF task types engage a wide array of cognitive skills, PVF tasks emphasize executive skills, and SVF tasks rely more on semantic categorization, hierarchical mental lexicon, and memory organization. Based on literature, we expect more variation in the processes required for SVF than PVF in our population, i.e., we expect differences between languages and groups to be more pronounced in SVF than in PVF. For RQ1, we anticipate that attriters will show comparable L1 and L2 performance or stronger performance in their self-reported dominant language (L2) in total scores. We predict shorter LoR and more frequent L1 use will positively affect L1 total scores in PVF and SVF but that these variables will not be strong predictors of performance independently. As participants are healthy, neurotypical adults, we anticipate minimal errors across languages. We hypothesize that attriters will produce more cross-language intrusion errors in their self-reported less dominant language (L1) and that a longer LoR and less frequent L1 use will result in an increased number of errors in L1 but not L2. Regarding temporal parameters, we hypothesize a similar performance profile across languages, and that attriters will more successfully employ rapid retrieval strategies in their self-reported stronger language (L2), particularly in SVF. We also predict that a shorter LoR and more frequent L1 use will facilitate rapid retrieval in L1 but that these variables will not be strong independent predictors of performance.

As for RQ2 on performance between attriters and monolinguals, we predict that attriters will achieve lower total scores in SVF than monolinguals but demonstrate comparable performance in PVF. We expect minimal errors overall, with a lower number of error-free trials in the attriter group. We anticipate that attriters will retrieve words more slowly than monolinguals in the early stages of the task, especially in SVF.

Method

Participants

Data from two healthy participant groups, attriters, and monolingual controls, were analyzed. The University of Turku Ethics committee approved all experimental procedures, and all participants provided informed consent before participating in the study. Exclusion criteria for both groups included (history of) cardiovascular, neurological, psychiatric, developmental language or speech disorders, toxic substance abuse, severe hearing loss, and age over 80 years.

Attriters

The attriter group ($N = 38$) consisted of first-generation Finnish immigrants living in Northern California. Participants were recruited through local Finnish groups and associations in Northern California and through introductions from other participants.

The group's age was $M = 60.90$ ($SD = 8.42$, range 45–79). Education was measured at two tiers: no academic degree ($n = 16$) and academic degree ($n = 22$). All participants spoke Finnish as their first language (L1) and English as their second language (L2). They had immigrated after puberty (age at immigration $M = 26.68$, $SD = 7.38$, range 9–48) between 1948 and 1998. All participants had migrated of their own volition as adults or as older children with their families. Participants who arrived with their parents had chosen to stay in the United States as adults.

Participants had lived in an L2 environment for at least 20 years, and the majority had used L2 as their primary professional language during this time (*What has been the primary working language in your work career? "Only English"* 65.8% $n = 25$, *"both but mostly English"* 26.3% $n = 10$, *"both, equally"* 5.3% $n = 2$ *"both, but mostly Finnish"* 2.6%, $n = 1$ *"only Finnish"* 0% $n = 0$*)*.

At the time of the interview, all attriters identified themselves as bilinguals (*Do you see yourself as bilingual? "no"* $n = 0$, *"yes"* $n = 38$). The majority reported that their language balance had shifted while living in the United States (*Has the balance between your Finnish and English changed during the time you have lived in the USA? "no"* 13.5% $n = 5$, *"yes"* 86.5%, $n = 32$).

Most attriters reported that they had experienced a decline in their L1 proficiency (L1) (*Do you think your Finnish language proficiency has changed since you moved to the USA? "Yes, I think it has become worse"* 84.2% $n=32$, *"no"* 13.2% $n=5$, *"yes, I think it has become better"* 2.6% $n=1$). In terms of proficiency, the majority perceived their proficiency in L1 and L2 as equal *(Do you feel that you are equally proficient in Finnish and English? "No, more proficient in English"* 27%, $n=10$, *"yes"* 48.6%, $n=18$, *"no, more proficient in Finnish"* 24.3%, $n=9$). Most attriters were equally comfortable in speaking L1 and L2 with a higher number of attriters indicating that they felt more comfortable in speaking L2 than L1 *(Do you feel more comfortable speaking Finnish or English? "English"* 26.3% $n=10$, *"Finnish,"* 5.3% $n=2$, *"no preference"* 68.4% $n=26$).

When asked about language dominance, attriters reported the following distribution on a 5-point Likert scale (*Language dominance [i.e., the language you are "best" at]? "1. English"* 31.6% $n=12$, *"2. mostly English"* 13.2% $n=5$, 3. *both Finnish and English* 23.7% $n=9$, 4. *mostly Finnish* 21.1% $n=8$, "5. *Finnish"* 10.5% $n=4$). This suggests that, on average, the perceived dominance in the attriter group was balanced bilingualism with a slight preference for English (L2) ($M=2.66$ [$SD=1.40$, range 1–5]).

Monolingual group

The monolingual group ($N=50$) consisted of Finnish speakers who had always lived in Finland and used Finnish as their only language in their everyday life. Data from the monolingual group for a similar analysis were published previously (Lehtinen et al., 2021) and are referenced here for group comparison. The group's average age was $M=62.58$ ($SD=7.59$, range 49–79). All participants identified as monolinguals, but due to language policy in Finland, all individuals were exposed to at least one foreign language in a school setting at a young age. In addition, passive exposure to foreign languages, notably Swedish and English, through media is assumed to be substantial in Finland, as television and movies are typically presented in their original language with Finnish subtitles.

Matched demographics

Monolingual participants were recruited to correspond to the age, education, and gender distribution of the attriter group to minimize demographic variables' potential effect. Statistical analyses verified no significant differences between the groups for age (attriters $M=60.90$, monolinguals $M=62.60$) $z=-0.97$, $p=.332$, education, $\chi(1)=0.49$, $p=.27$, or gender, $\chi(1)=0.43$, $p=.51$.

Data

Data were extracted from a larger dataset consisting of five verbal fluency tasks (L1, L2), a WUG-task (Crystal, 2015) (L1), speech samples elicited via a film retelling task (L1, L2), and free speech samples (L1). In this study, we focused on the participants' performance in four VF tasks in L1 and L2: three phonemic categories in L1 (Finnish, /k/, /a/, /p/) and L2 (English, /f/, /a/, /s/) and one concrete semantic category, "animals".

Background information was collected via the Language Attrition Test Battery Sociolinguistic Questionnaire (SQ), introduced and discussed in detail in Schmid (2011b), Schmid and Cherciov (2019, Chapter 23, pp. 267–276), and Schmid and Dusseldorp (2010), and available in multiple languages on www.languageattrition.org. The first author translated the SQ into Finnish (Supplemental Appendix A). An abridged questionnaire for the control group was used to verify language use in daily life (Supplemental Appendix B). Participants were given the option to fill out the questionnaire before or during the session.

All tasks were completed in a quiet environment (such as participants' home or clinical setting) in one sitting in Northern CA, USA for attriters, and Finland for the control group. Tasks were completed in L1 and L2 within a single session. All tasks were completed in one language and after a short break in the other. The order of languages was randomized, and participants were aware that the interlocutor was bilingual. Tasks were presented in randomized order within each language, with VF tasks presented in the following order: concrete category "animals," (abstract category "emotions"), phonemes "/k/, /a/, /p/" for L1 and "/f/, /a/, /s/" for L2.

This study focuses on three phonemic and one semantic VF tasks. For the phonemic task in L1, we followed the high-frequency dictionary approach (Lehtinen et al., 2021; Mardani et al., 2020; Oberg & Ramírez, 2006; Schmid, 2011a) selecting the two most frequent word-initial consonants of Finnish /k/ (15242 words) and /p/ (10640 words) and the most frequent word-initial vowel /a/ (4361 words) (Kielitoimiston sanakirja, 2021; Leskinen, 1989). For L2 (English), we selected phonemes /f/ (6939 words), /a/ (10360 words), and /s/ (19236 words) (Merriam-Webster, n.d.) based on their frequent use in literature (e.g., Strauss et al., 2006). For the SVF, task we included one semantic category, the culturally and linguistically relatively neutral, concrete category, "animals" (e.g., Pekkala et al., 2009), to limit the complexity of semantic analysis. The administration of the five VF tasks typically lasted for 5–10 minutes; the whole session was completed within 2 hours.

As extralinguistic variables, we examined the LoR in the L2 dominant environment in years ($M = 34.24$ years, $SD = 10.83$, range 20–70) and frequency of L1 use. For frequency of L1 use, we asked the participants to estimate their language use on a 5-point Likert scale, with the majority reporting weekly use of L1 *(How often do you speak Finnish? "rarely"* $n = 1$, *"few times a year"* $n = 1$, *"monthly"* $n = 7$, *"weekly"* $n = 22$, *"daily"* $n = 7$). We aimed to investigate L1 use for different domains, especially leisure versus professional use. However, only three participants reported using L1 for work regularly, making it unreliable to analyze the impact of professional use of L1.

Procedure

All VF tasks were administered and scored following the procedure outlined in Lehtinen et al. (2021). Briefly, participants were asked to produce as many different words as possible in 60 seconds following the given criteria, with the only restriction being for proper names. Responses were transcribed, and acceptable words were calculated for total scores, with semantically distinctive words calculated as separate items. Errors were excluded from the total score and classified into four categories: repetitions, categorical errors, nonwords, and language intrusions. Utterances in Finglish, a macaronic mixture of Finnish and English (e.g., *päntsit "pants" [standard Finnish housut]* see also Virtaranta, 1992), were scored as nonwords. Temporal parameters were analyzed by calculating the number of acceptable words in four 15-second segments (i.e., 0–15, 16–30, 31–45, and 46–60). An example of scored trials is shown in Supplemental Appendix C (L1 Table 1, L2 Table 2).

We utilized R software (R Core Team, 2019) for statistical modeling and data visualization with packages dplyr (Wickham et al., 2019), lme4 (Bates et al., 2015), sjPlot (Lüdecke, 2018), ggplot2 (Wickham, 2016), arsenal, and ggeffects (Lüdecke, 2018). The analysis script is available at https://osf.io/fue3k/?view_only=6b6762f07e2243d6b8548c0992dce9f1. Continuous predictors used as predictors were scaled and centered to sample means in all models.

To address RQ1.a. on overall performance in the attriters group, we used a linear mixed-effects model to examine total scores as a function of task language (L1/L2), task type (PVF/SVF), LoR, and frequency of L1 use and their interactions. We included participant intercept and slope for task language as a random factor to account for individual differences in proficiency between Finnish and English. We selected predictors for the final model using a model comparison with Bayesian information criterion (BIC) values, choosing the simplest model when the two models did not differ significantly in BIC. To address RQ1b on the number of errors between L1 and L2, we used the Wilcoxon Rank-Sum Test. Due to the limited number of errors in the dataset, we excluded the impact of extralinguistic variables on errors from the analysis. To answer RQ1c on temporal patterns, we fitted a model to examine the attriters' performance changes over time during the 60 seconds. The model considered the number of correct words generated during 15-second time intervals and included task language, time sequence (0–15, 16–30, 31–45, or 46–60), task type, LoR, and L1 use as predictors. Participant intercept with a slope for task type was used as a random factor to account for individual-level differences. To analyze the ability to generate words rapidly in the early stages of the task (RQ1d), we modeled the ratio of correct words generated in the first segment to the total score. We used task language (L1/L2), LoR, and L1 use as predictors, and included participant intercept with a slope for task language as a random factor.

For RQ2a on the overall performance between attriter and monolingual groups, we used a linear mixed-effects model to model total scores as a function of task type (PVF/SVF) and participant group. Participant intercept with a slope for task type was included as a random factor to account for individual-level variation. The number of errors between groups in L1 (RQ2b) was analyzed using the Wilcoxon Rank-Sum Test. For group comparisons between temporal performance profiles between attriters and monolinguals (RQ2c), we fitted a model to investigate performance changes during the task, with the number of correct words generated during 15-second intervals as the outcome variable, and participant group, time sequence, and task type as predictors. We used the same model selection procedure as previously described. To compare the ability to generate words rapidly in the early stages of the task between attriters and monolinguals (RQ2d), we modeled the ratio of correct words generated in the first segment to the total score as the outcome variable, with task type and participant group as predictors, and included participant intercept with a slope for task type as a random factor. The model selection followed the procedure described earlier.

Results

For brevity, we have presented descriptive statistics and complete model summaries in Supplemental Appendix C, with statistics for key findings summarized in the following.

Research question 1

Our first research question was: How do language attriters perform in PVF and SVF tasks for their first (L1) and second (L2) languages, and to what extent do LoR and frequency of L1 use affect performance?

Total scores. To analyze total scores in VF tasks, the model selection procedure based on BIC values (L1 use, LoR, or both, combined with task language and fluency type) suggested the model with L1 use but no LoR as the best fit for the data (Supplemental Appendix C, Table 3 and Model Summary 1). Our analysis showed statistically significant main effects of task language, fluency type, and L1 use, indicating that more words were generated in L1 than in L2 and in semantic compared with phonemic tasks. Frequent L1 use was associated with better overall performance in the tasks across fluency types and task languages (1.46, 95% CI = [0.36, 2.55], $t = 2.62$, $p = .009$).

We found no significant interactions between L1 use and task type and/or task languages. VF performance for attriters in L1 and L2 is visualized in Figure 1.

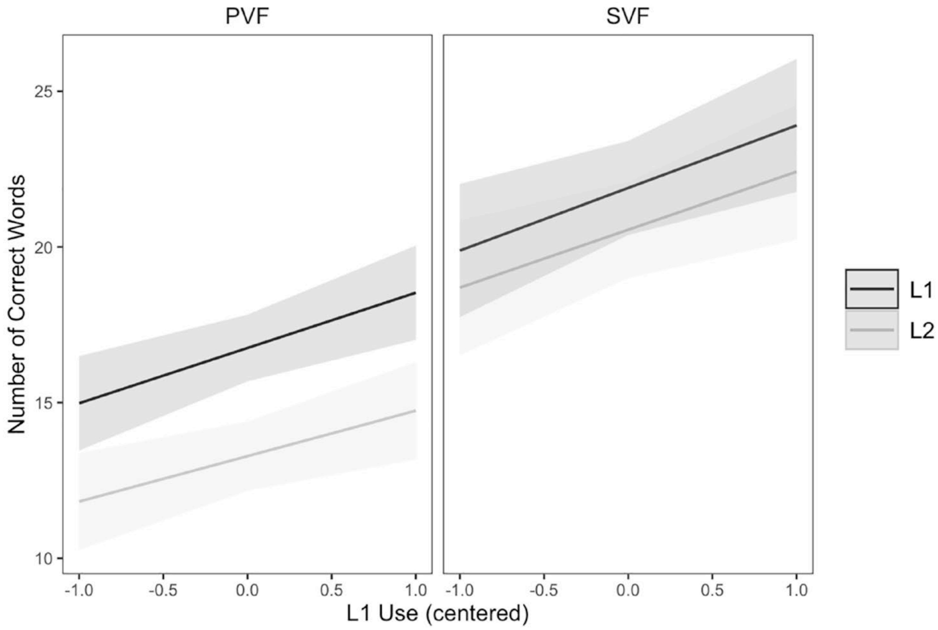


Figure 1. Predicted values of number of correct words in the attriter group in one semantic and three phonemic tasks (Combined) in L1 and L2 with frequency of L1 use as a predictor.

The number of errors, frequency, and distribution of error types. Table 1 presents the descriptive statistics for errors in L1 and L2 for the attriter group. Overall, the number of errors was minimal, which led us to exclude the impact of extralinguistic variables on errors from our analysis.

Table 1. Number of attriters who generate errors, total number of errors, and distribution of error types in one semantic and three phonemic verbal fluency tasks in L1 and L2.

Errors	L1				L2			
	/k/	/a/	/p/	/animals/	/f/	/a/	s	/animals/
n with errors	27 (71%)	22 (58%)	26 (68%)	15 (39%)	22 (59%)	25 (68%)	17 (46%)	17 (46%)
Total nr. of errors	43	40	48	27	41	36	20	26
Mean	1.13	1.05	1.26	0.71	1.11	0.97	0.54	0.70
SD	1.14	1.21	1.25	1.27	1.26	0.96	0.73	1.02
Range	0–5	0–4	0–4	0–6	0–4	0–3	0–3	0–4
Error type								
Repetition	21 (49%)	27 (68%)	34 (71%)	18 (66%)	26 (63%)	16 (44%)	10 (50%)	18 (69%)
Categorical	12 (28%)	12 (30%)	5 (10%)	5 (19%)	10 (24%)	15 (42%)	7 (35%)	6 (23%)
Nonword	10 (23%)	1 (3%)	9 (19%)	0 (0%)	4 (10%)	4 (11%)	2 (10%)	0 (0%)
Intrusion	0 (0%)	0 (0%)	0 (0%)	4 (15%)	1 (2%)	1 (3%)	1 (5%)	2 (9%)

Note. L1 $N = 38$; L2 $N = 37$.

As expected, the most common error type was repetition. There was no significant difference in the number of errors between L1 and L2 in PVF ($W = 5573$, $p = .1007$) or SVF ($W = 731$, $p = .741$) in the attriters group.

Performance change during the task. To examine the temporal profile and its association with task attributes (SVF / PVF and L1 / L2), we employed a model to predict the total number of acceptable words produced during four 15-second time windows, with L1 use and LoR as predictors. Using the model comparison procedure described earlier, we determined that a model with L1 use but without LoR was the most parsimonious (Supplemental Appendix C, Table 3 and Model Summary 2). The selected model showed that more words were generated at the beginning of the task than in the later segments in both L1 and L2, with no significant difference between languages, as illustrated in Figure 2. As in RQ1a (Model Summary 1) a main effect of L1 use was detected. However, there were no significant interactions between L1 use and temporal parameters, suggesting that the observed pattern was similar across languages despite the L1 use frequency.

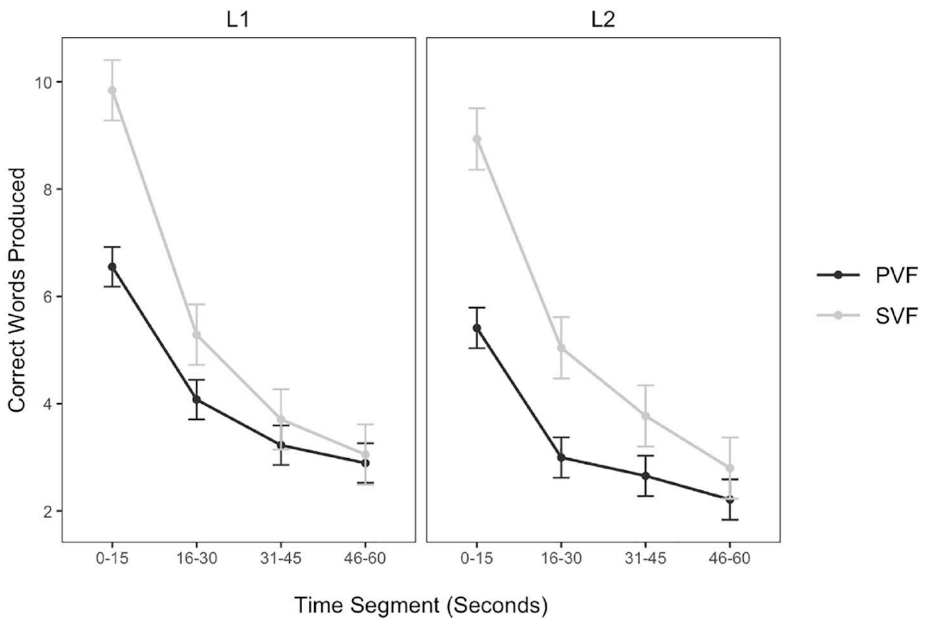


Figure 2. Predicted values of correct words in the four 15-second time segments in one semantic and three phonemic verbal fluency tasks in the attriter group in L1 and L2. Error bars represent 95% confidence intervals.

Ability to generate words rapidly in the early stages of the task. To further examine performance variation in the attriters group, we examined the quotient of acceptable words produced during the first time segment in all tasks as a function of task language and L1 use (Supplemental Appendix C, Model Summary 3). Our analysis revealed that attriters generated a smaller quotient in L1 (42%) than in L2 (44%) during the first time segment, although the difference between languages was not statistically significant. Moreover, our model showed that attriters who reported using L1 more frequently in their everyday life produced a smaller quotient in both languages than those who reported using L1 less frequently (−0.02, 95% CI = [−0.04, −0.01], $t = -2.52$, $p = .012$), illustrated in Figure 3.

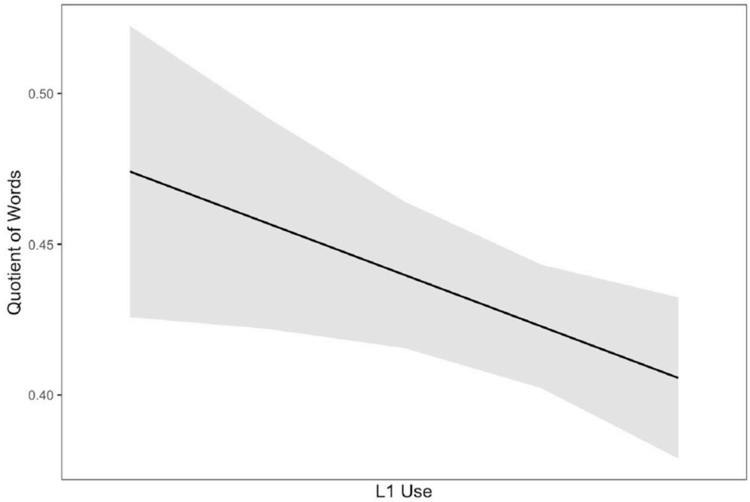


Figure 3. Predicted values of the quotient of correct words in the first 15-second time segment in the attriters group in L1 and L2 (Combined) in all task types (combined) with frequency of L1 use as predictor.

Research question 2

Our second research question was: How is the performance of language attriters in PVF and SVF tasks in their first language (L1) compared with that of a matched monolingual group?

Total scores. For group comparison for overall performance measured in total scores, the results indicated that monolinguals outperformed attriters in the L1 SVF task (3.65, 95% CI = [1.40, 5.91], $t = 3.19$, $p = .002$). The performance of both groups was similar in the PVF task (Supplemental Appendix C, Tables 3 and 4, Model Summary 4). Results are shown in Figure 4.

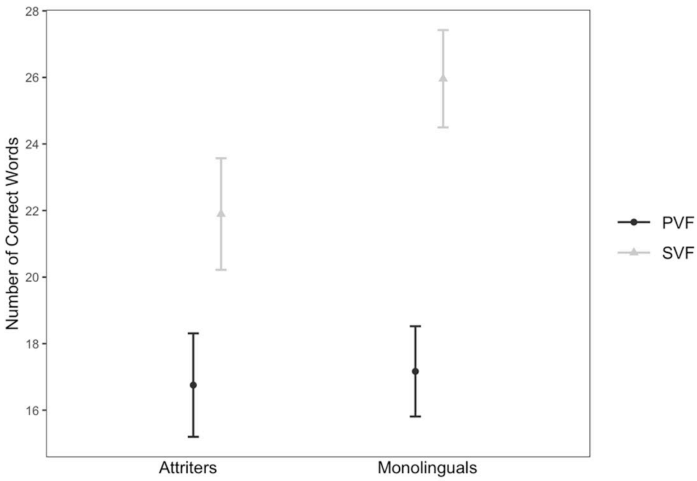


Figure 4. Comparison of predicted values of number of correct words in one semantic and three phonemic verbal fluency tasks (Combined) in L1 between the attriter and monolingual groups using participant group as a predictor.

The number of errors, frequency, and distribution of error types. For the monolingual group, results published previously in Lehtinen et al. (2021) are referenced here. In the PVF task, approximately half of the monolingual participants generated errors in all three trials (/k/ 44% [$n=22$]; /a/ 58% [$n=29$]; /p/ 50% [$n=25$]). Compared with monolinguals, a higher percentage of attriters generated errors in L1 PVF trials /k/ (71% [$n=27$]) and /p/ (68% [$n=26$]), and the percentage was similar in trial /a/ (58% [$n=22$]). Attriters generated significantly more errors than monolinguals in PVF (monolinguals $Md=1$, attriters $Md=1$, $W=9937.5$, $p=.016$). The most common error type was repetitions across tasks in both groups, followed by nonword errors in the attriter group. In SVF more attriters (39% [$n=15$]) than monolinguals (28% [$n=14$]) generated errors. While attriters generated more errors than monolinguals numerically (27 vs. 20), this difference was not statistically significant (monolinguals $Md=0$, attriters $Md=0$, $W=1062.5$, $p=.255$).

Performance change during the task. To investigate the temporal patterns of task performance and their relationship to task type in attriter and monolingual groups, we modeled the total number of acceptable words as a function of the number of acceptable words generated during four 15-second time segments (Supplemental Appendix C, Tables 3 and 4, Model Summary 5). Our analysis showed no significant differences between the two groups in the distribution of words across the four time segments, as demonstrated in Figure 5.

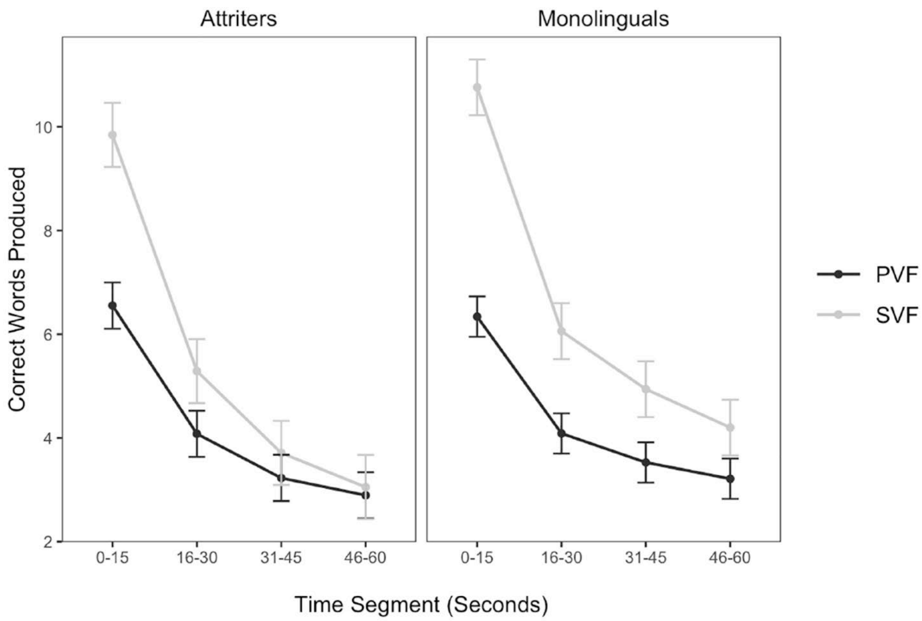


Figure 5. Predicted values of words produced in the four 15-second time segments in one semantic and three phonemic verbal fluency tasks in the attriter and monolingual groups. Error bars represent 95% confidence intervals.

Ability to generate words rapidly in the early stages of the task. We also investigated the performance of attriters and monolinguals during the initial 15-second of all tasks by modeling the quotient of correct words produced as a function of the group (Supplemental Appendix C, Model Summary 6). The analysis revealed that the quotient was higher in attriters (43%) than in monolinguals (41%)

(−0.02, 95% CI = [−0.04, 0.00], $t = -2.29$, $p = .022$), suggesting that attriters were able to produce more correct words in the initial stages of the task.

Discussion

We set out to explore processes that underlie performance in verbal fluency tasks in Finnish-English mature immersed bilinguals (i.e., language attriters), who self-identify as balanced bilinguals with a slight preference for L2. We analyzed PVF and SVF data in both their first (L1) and second (L2) languages, focusing on total scores, errors, and temporal parameters, and assessed the impact of immersion duration (LoR), frequency of L1 use, and their interactions on task performance. In addition, we contrasted the attriters' performance in L1 verbal fluency tasks to that of monolinguals to identify potential language attrition markers.

Our key findings show that attriters generated more acceptable words in their first language than in their second language across fluency types and that frequent L1 use supported performance in both languages (RQ1a, Model Summary 1, Figure 1). Attriters made very few errors in either language, with the most typical error type being repetitions (RQ1b). The temporal performance profile was similar for both languages (RQ1c, Model Summary 2, Figure 2), and those attriters who used L1 more frequently generated a smaller quotient of words in the initial stage of VF tasks in both languages than those who used L1 less frequently (RQ1d, Model Summary 3, Figure 3).

Compared with monolingual L1 performance, attriters generated fewer acceptable words in the L1 semantic task than monolinguals, and groups performed similarly in the phonemic task (RQ2a, Model Summary 4, Figure 4). Attriters made more errors than monolinguals in the phonological task, but the number of errors was comparable in the semantic task between groups. Qualitatively fewer attriters than monolinguals generated error-free trials in both tasks, with the most typical error in both groups being repetitions (RQ2b). The temporal performance profile in L1 did not differ between attriters and monolinguals (RQ2c, Model Summary 5, Figure 5). However, attriters generated more correct words in the initial stage of the task compared with monolinguals (RQ2d, Model Summary 6).

In the following, we discuss our findings in relation to earlier literature. We focus on performance between L1 and L2 in the attriter group before moving on to group comparison between the attriter and monolingual groups in L1. Finally, we outline suggestions for future research and address some limitations of this study.

RQ1. Performance between L1 and L2 within the attriter group

Attriters demonstrated strong L1 proficiency by generating higher total scores in L1 than in L2 across tasks. This points in the direction of L1 as the more dominant language, contrary to the attriters' self-reported language dominance (Roberts & Le Dorze, 1997; Rosselli et al., 2002). The small number of errors across tasks did not allow robust statistical analysis of errors, but some observations can be made. While the number of errors was similar across languages, qualitatively, attriters generated more language intrusions in L1 than in L2, suggesting transference from L2 to L1 but not vice versa. As attriters demonstrated stronger overall performance in L1, this contradicts our hypothesis of increased intrusion errors in the non-dominant language (Sandoval et al., 2010). Overall, by generating more correct words and allowing more language intrusions in L1 than L2, attriters demonstrate robust language proficiency but greater flexibility in L1 performance compared with L2, potentially as a marker of cross-linguistic influence of L2 (Schmid & Cherciov, 2019, Chapter 23, p. 267; Schmid & Dusseldorp, 2010; Sharwood Smith, 2019, Chapter 8, p. 85).

For temporal performance profiles, we anticipated similar performance across languages or faster initiation in the more dominant language among the attriters group (Fernaeus & Almkvist, 1998; Fernaeus et al., 2008; Gollan et al., 2011; Luo et al., 2010; Sandoval et al., 2010; Schmid & Jarvis, 2014; Yilmaz & Schmid, 2018). We found comparable temporal word distributions across languages and equally efficient lexical retrieval in the initial stage of both languages, indicating balanced language dominance.

Contrary to our hypothesis, there was no interaction effect for LoR and frequency of L1 use, but frequent L1 use supported overall performance and proportionally slowed down rapid retrieval in both languages. Previous research (e.g., Gurunandan et al., 2023) has shown that similar neural activation patterns occur during L1 and L2 VF performance, suggesting a common bilingual effect for both languages. Our results of similar effect of frequency of L1 use for both languages rather than a specific language may reflect the influence bilingualism has on overall cognitive and neural language processing. Thus, our findings highlight the importance of including both languages, L1 as well as L2 in language attrition studies to account for the general effect of bilingualism.

In addition to the impact of frequency of L1 use, we also aimed to investigate to what extent LoR affects VF performance in L1 and L2. However, our model selection procedure suggested a model without LoR to be the most parsimonious fit to our data and we were not able to directly investigate the effect of LoR on task performance. In a similar vein to the similar impact of frequency of L1 use for both languages, this omission might suggest that LoR was not a meaningful predictor for VF task performance, in either the first or the second language.

RQ2. Performance between the attriter and the monolingual group

Our findings between attriters and monolinguals mirror literature. Monolinguals outperformed attriters in SVF, and groups performed similarly in PVF (e.g., Badstübner, 2011; Dostert, 2009; Opitz, 2011; Schmid, 2011b, 2019; Schmid & Dusseldorp, 2010; Schmid & Jarvis, 2014; Schmid & Keijzer, 2009; Schmid & Köpke, 2009). As expected, there were fewer error-free trials in the attriter group, especially in trials /k/ and /p/, with emphasis on repetitions and nonword errors. Numerically attriters generated more errors in the PVF than monolinguals. Thus, attriters experienced some difficulty in rapid lexical retrievals compared with monolinguals, potentially due to language interference or contact-induced language change that manifested as nonword errors.

Contrary to our hypothesis, performance in the attriter group was not slowed down in the early stages of the task compared with monolinguals. Thus, attriters did not demonstrate markers of language interference in temporal analysis that would explain the differences in overall SVF task performance, as suggested by earlier studies (Rosselli et al., 2000, 2002; Sandoval et al., 2010; Schmid & Jarvis, 2014). Moreover, attriters generated a higher percentage of total words in the first 15-second segment than monolinguals. As word-finding is facilitated by semi-automatic rapid retrieval strategies in the early segments of the task (Fernaeus et al., 2008; Lehtinen et al., 2021; Sandoval et al., 2010; Venegas & Mansur, 2011), our results suggest that attriters relied more on semi-automatic rapid retrieval strategies in L1 than monolinguals, implying not only that these strategies are resistant to language attrition (Gollan et al., 2002; Segalowitz, 1991), but also that in the first 15-second time segment attriters may utilize rapid retrieval more efficiently than monolinguals. Interestingly, this is highlighted by the finding that while frequent L1 use slowed attriters down in the first 15-second segment, they still performed better than monolinguals in the first 15-second segment, with the overall performance resulting in lower total scores in SVF. Investigating the processes underlying these observations is beyond the scope of this study. In the following, we discuss suggestions for future studies and address the limitations of the present study.

Limitations of the present study, and suggestions for future research

As our sample was small, interpretations of the data are preliminary explorations. Future studies with larger datasets are needed to investigate the trends found in our data reliably. This is particularly relevant for error analysis. Due to the small amount of data, we were not able to conduct a robust statistical analysis or investigate the distribution profile of errors within tasks. In future studies, examining the distribution of errors over time may provide valuable insight into the performance differences in the initial stage between attriters and monolinguals.

Regarding error analysis, interpreting fine-grained data on error types from studies using different categorization systems can be challenging. For instance, our findings contrast with Badstübners' (2011) findings, which found a significant difference between monolinguals and attriters only in a SVF task. However, our analysis cannot be directly compared with Badstübners'. They examined combined data from three semantic categories and found that most incorrect lexical items in SVF resulted from L2 transfer in the "things in the kitchen" category (i.e., words related to everyday life). In addition to complexities in combining data from different semantic categories, in environments where L1 is used infrequently (or not at all), words related to everyday life are highly activated in L2. Consequently, Badstübners' (2011) category related to everyday life objects might have prompted more language interference from L2 than the less frequently used category of "animals" in our study.

In addition, we categorized non-standard utterances as nonword errors. These included terms that could be classified as Finglish (a macaronic blend of Finnish and English, as described in Virtaranta, 1992), which might also be considered L2 language intrusions, indicators of substantial language interference, or signs of language evolution within the attriters group in L1. As such, a distinct category for this type of error could have been informative, especially in future studies conducted with larger datasets. We conducted our analysis following Lehtinen et al. (2021) and remain optimistic that this qualitative data analysis can serve as a foundation for future research.

When interpreting our results, it is important to consider the potential impact of shared cognates on language interference. The languages in this study, L1 (Finnish) and L2 (English) have very few cognates. As Schmid and Jarvis (2014) demonstrate, there may be less competition between high-frequency words in Finnish and English compared with language pairs with a larger shared vocabulary. Thus, it is possible that in our dataset, the effect of language interference on rapid lexical retrieval was more subtle, and our analysis only detected it as a within-group variation effect linked to the frequency of L1 use and not at the group level.

Our analysis consisted of two groups: attriters and monolinguals. A comprehensive analysis of the monolingual group VF task performance was previously reported by Lehtinen et al. (2021). They found no effect for age but detected a positive effect of education and gender on VF task total scores. For the present study, the monolingual and attriter groups were matched for age, gender, and education. Although we are confident that closely matched groups effectively control for age, gender, and education in group comparisons, it is possible that these variables influenced the performance in the attriter group in ways not accounted for in our analyses. Controlling for education and gender within the attriter group, especially for the impact of L1 use on rapid retrieval, could have strengthened our data interpretation. We recommend controlling for age and education in future studies to ensure a more comprehensive interpretation of the results.

Limitations of this study include measuring L1 use only as an overall measure of the frequency of use in everyday life. Based on earlier research (e.g., Schmid & Dusseldorp, 2010), our goal was to explore the impact of L1 use for professional purposes on VF performance. Unfortunately, our dataset did not allow for such an analysis. A more detailed approach to L1 use might help explain variation in task performance within the attriter group. However, assessing the influence of extralinguistic variables on observed phenomena is complex. Language history reports on L1 use typically involve subjective self-reports spanning several years and varying external circumstances that dictate the use of L1 in everyday life. These reports should be considered as the best available approximations, but their comparability between individuals is not straightforward (Bylund & Ramírez-Galan, 2016; Köpke & Schmid, 2004). Furthermore, LoR and L1 use are interconnected in many ways. While the time elapsed from leaving the L1 country is measurable, assessing the level of deprivation of L1 during that time is more challenging, especially in datasets dating to modern times (Schmid, 2019). Despite the limitations, exploring the impact of shared language history in language attrition populations is valuable beyond research interests. Gaining insight into how these factors affect L1 performance can motivate and support individuals in preserving their native language and cultural identity.

Conclusions

The present study showed that attriters, who self-report as balanced bilinguals with a slight preference for L2, demonstrate strong proficiency in L1 with similar lexical retrieval strategies in L1 and L2. Our analysis suggests balanced bilingualism with a subtle emphasis for L1 at the group level after 20 years of immersive exposure to L2, partly contrary to participant self-reports. We showed that frequent L1 use supports overall VF task performance but proportionally slows performance down in the initial stage of the task in both languages without significantly impacting overall performance. Compared with monolinguals, attriters demonstrate an overall disadvantage in SVF, but this disadvantage is not due to a slower initiation profile, as hypothesized. In contrast, attriters rely on rapid retrieval in L1 more than monolinguals. These findings add to our understanding of how attriters and monolinguals approach verbal fluency tasks and highlight the potential importance of early task performance in VF task analysis.

Our findings support the notion of two-way interaction in cognitive language processing, acquisition, attrition and dominance shifts in a bilingual environment. In the future, analyzing processes underlying VF performance in-depth in L1 and L2 using techniques like clustering and switching analysis (Lehtinen et al., 2021; Troyer, 2000; Troyer et al., 1997) could increase our understanding of lexical retrieval strategies in language attrition populations and the role of frequent L1 use in bilingual language processing.

Acknowledgements

The authors thank MPhil Ida Luotonen for her contributions to data analysis and overall support, and Professor Marja-Liisa Helasvuo and Dr Raymond Bertram for scientific advice and valuable comments and suggestions on earlier versions of this manuscript. They would also like to thank Professor Monika Schmid for inspiration and support during this study. Finally, they thank PhD Lotta Weckström for her efforts in recruiting participants, and express their gratitude to the participants of this study for the generous donation of their time and effort in participating in this research.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by The Turku University Foundation, The Finlandia Foundation, The Migration Institute of Finland, The University of Turku Graduate School, and The Alfred Kordelin Foundation.

ORCID iDs

Nana Lehtinen https://orcid.org/0000-0002-3606-1302

Anna Kautto https://orcid.org/0000-0003-4885-9167

Supplemental material

Supplemental material for this article is available online.

References

Abwender, D. A., Swan. J. G., Bowerman, J. T., & Connolly, S. W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment*, *8*(3), 323–338. https://doi.org/10.1177/107319110100800308

Badstübner, T. (2011). *L1 attrition: German immigrants in the U.S*. [Doctoral thesis, The University of Arizona].

Barry, D., Bates, M. E., & Labouvie, E. (2008). FAS and CFL forms of verbal fluency differ in difficulty: A meta-analytic study. *Applied Neuropsychology*, *15*(2), 97–106. https://doi.org/10.1080/09084280802083863

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Becker, N., & de Salles, J. F. (2016). Methodological criteria for scoring clustering and switching in verbal fluency tasks. *Psico-USF*, *21*(3), 445–457. https://doi.org/10.1590/1413-82712016210301

Borkowski, J. G., Benton, A. L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia*, *5*(2), 135–140. https://doi.org/10.1016/0028-3932(67)90015-2

Bylund, E., & Ramírez-Galan, P. (2016). Language aptitude in first language attrition: A study on late Spanish-Swedish bilinguals. *Applied Linguistics*, *37*(5), 621–638. https://doi.org/10.1093/applin/amu055

Crowe, S. F. (1998). Decrease in performance on the verbal fluency test as a function of time: Evaluation in a young healthy sample. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 391–401. https://doi.org/10.1076/jcen.20.3.391.810

Crystal, D. (2015). *A dictionary of linguistics and phonetics*. http://search.credoreference.com/content/title/bkdictling

de Bot, K., & Clyne, M. (1994). A 16-year longitudinal study of language attrition in Dutch immigrants in Australia. *Journal of Multilingual and Multicultural Development*, *15*(1), 17–28. https://doi.org/10.1080/01434632.1994.9994554

De Leeuw, E., Schmid, M. S., & Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition*, *13*(1), 33–40. https://doi.org/10.1017/S1366728909990289

Dostert, S. C. (2009). *Multilingualism, L1 attrition and the concept of "native speaker"* [Doctoral thesis, Heinrich Heine University Düsseldorf].

Fernaeus, S.-E., & Almkvist, O. (1998). Word production: Dissociation of two retrieval modes of semantic memory across time. *Journal of Clinical and Experimental Neuropsychology*, *20*(2), 137–143. https://doi.org/10.1076/jcen.20.2.137.1170

Fernaeus, S.-E., Östberg, P., Hellström, Å., & Wahlund, L.-O. (2008). Cut the coda: Early fluency intervals predict diagnoses. *Cortex*, *44*(2), 161–169. https://doi.org/10.1016/j.cortex.2006.04.002

Gollan, T. H., Montoya, R. I., & Werner, G. A. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, *16*(4), 562–576. https://doi.org/10.1037/0894-4105.16.4.562

Gollan, T. H., Sandoval, T., & Salmon, D. P. (2011). Cross-language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological Science*, *22*(9), 1155–1164. https://doi.org/10.1177/0956797611417002

Goral, M. (2004). First-language decline in healthy aging: Implications for attrition in bilingualism. *Journal of Neurolinguistics*, *17*(1), 31–52. https://doi.org/10.1016/S0911-6044(03)00052-6

Grosjean, F. (2013). Bilingualism: A short introduction. In F. Grosjean & P. Li (Eds.), *The psycholinguistics of bilingualism* (Vol.5, pp. 5–25). John Wiley.

Gurunandan, K., Carreiras, M., & Paz-Alonso, P. M. (2023). Verbal production dynamics and plasticity: Functional contributions of language and executive control systems. *Cerebral Cortex*, *33*, 740–753. https://doi.org/10.1093/cercor/bhac097

Jarvis, S. (2019). Lexical attrition. In M. S. Schmid, & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 240–250). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.20

Kasparian, K., & Steinhauer, K. (2017). On missed opportunities and convenient "truths." *Linguistic Approaches to Bilingualism*, *7*(6), 709–714. https://doi.org/10.1075/lab.00010.kas

Kielitoimiston sanakirja. (2021, October 28). [Dictionary of standard Finnish compiled in the institute for the languages of Finland]. https://www.kielitoimistonsanakirja.fi

Köpke, B., & Schmid, M. S. (2004). Language attrition: The next phase. In M. S. Schmid, B. Köpke, M. Keijzer, & L. Weilemar (Eds.), *Studies in bilingualism* (Vol.28, p. 1). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.28.02kop

Laine, M., & Lehtonen, M. (2018). Cognitive consequences of bilingualism: Where to go from here? *Language, Cognition and Neuroscience*, *33*(9), 1205–1212. https://doi.org/10.1080/23273798.2018.1462498

Lehtinen, N., Luotonen, I., & Kautto, A. (2021). Systematic administration and analysis of verbal fluency tasks: Preliminary evidence for reliable exploration of processes underlying task performance. *Applied Neuropsychology: Adult*. Advance online publication. https://doi.org/10.1080/23279095.2021.1973471

Leskinen, H. (1989). Tietoja sananalkuisten grafeemien ja grafeemikombinaatioiden yleisyyssuhteista [About the frequency of word-initial graphems and grapheme combinations]. *Virittäjä*, *93*(3), 401–419. https://journal.fi/virittaja/article/view/38309

Linck, J. A., & Kroll, J. F. (2019). Memory retrieval and language attrition: Language loss or manifestations of a dynamic system? In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 87–97). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.8

Ljunberg, J. K., Hansson, P., Andrés, P., Josefsson, M., & Nilsson, L.-G. (2013). A longitudinal study of memory advantages in bilinguals. *PLOS ONE*, *8*(9), e73029. https://doi.org/10.1371/journal.pone.0073029

Lüdecke, D. (2018). *sjPlot—Data visualization for statistics in social science*. Zenodo. https://doi.org/10.5281/ZENODO.1308157

Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition*, *114*(1), 29–41. https://doi.org/10.1016/j.cognition.2009.08.014

Mardani, N., Pourjafari, M., Irandegani, M. A., Ahmadi, N., & Baghban, K. (2020). The effect of bilingualism on the processing of clustering and switching in verbal fluency tasks. *Journal of Rehabilitation Sciences & Research*, *7*(3), 114–117. https://doi.org/10.30476/jrsr.2020.86911.1096

Marsh, J. E., Hansson, P., Sörman, D. E., & Ljungberg, J. K. (2019). Executive processes underpin the bilingual advantage on phonemic fluency: Evidence from analyses of switching and clustering. *Frontiers in Psychology*, *10*, Article 1355. https://doi.org/10.3389/fpsyg.2019.01355

Merriam Webster. (n.d.). *[Official SCRABBLE® players dictionary]*. Retrieved August 30, 2023, from https://scrabble.merriam.com

Oberg, G., & Ramírez, M. (2006). Cross-linguistic meta-analysis of phonological fluency: Normal performance across cultures. *International Journal of Psychology*, *41*(5), 342–347. https://doi.org/10.1080/00207590500345872

Olabarrieta-Landa, L., Torre, E. L., López-Mugartza, J. C., Bialystok, E., & Arango-Lasprilla, J. C. (2017). Verbal fluency tests: Developing a new model of administration and scoring for Spanish language. *NeuroRehabilitation*, *41*(2), 539–565. https://doi.org/10.3233/NRE-162102

Opitz, C. (2011). *First language attrition and second language acquisition in a second language environment* [Doctoral thesis, Trinity College, University of Dublin]. http://hdl.handle.net/2262/78154

Patra, A., Bose, A., & Marinis, T. (2020). Performance difference in verbal fluency in bilingual and monolingual speakers. *Bilingualism: Language and Cognition*, *23*(1), 204–218. https://doi.org/10.1017/S1366728918001098

Pekkala, S., Goral, M., Hyun, J., Obler, L. K., Erkinjuntti, T., & Albert, M. L. (2009). Semantic verbal fluency in two contrasting languages. *Clinical Linguistics & Phonetics*, *23*(6), 431–445. https://doi.org/10.1080/02699200902839800

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Roberts, P. M., & Le Dorze, G. (1997). Semantic organization, strategy use, and productivity in bilingual semantic verbal fluency. *Brain and Language*, *59*(3), 412–449. https://doi.org/10.1006/brln.1997.1753

Ross, T. P. (2003). The reliability of cluster and switch scores for the Controlled Oral Word Association Test. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *18*(2), 153–164.

Rosselli, M., Ardila, A., Araujo, K., Weekes, V. A., Caracciolo, V., Padilla, M., & Ostrosky-Solí, F. (2000). Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. *Applied Neuropsychology*, *7*(1), 17–24. https://doi.org/10.1207/S15324826AN0701_3

Rosselli, M., Ardila, A., Salvatierra, J., Marquez, M., Luis, M., & Weekes, V. A. (2002). A cross-linguistic comparison of verbal fluency tests. *International Journal of Neuroscience*, *112*(6), 759–776. https://doi.org/10.1080/00207450290025752

Runnqvist, E., & Costa, A. (2012). Is retrieval-induced forgetting behind the bilingual disadvantage in word production? *Bilingualism: Language and Cognition*, *15*(2), 365–377. https://doi.org/10.1017/S1366728911000034

Sandoval, T. C., Gollan, T. H., Ferreira, V. S., & Salmon, D. P. (2010). What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition*, *13*(2), 231–252. https://doi.org/10.1017/S1366728909990514

Santos Nogueira, D., Azevedo Reis, E., & Vieira, A. (2016). Verbal fluency tasks: Effects of age, gender, and education. *Folia Phoniatrica et Logopaedica*, *68*(3), 124–133. https://doi.org/10.1159/000450640

Schmid, M. S. (2011a). Contact x time: External factors and variability in L1 attrition. In M. S. Schmid & W. Lowie (Eds.), *Studies in bilingualism* (Vol.43, pp. 155–176). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.43.11sch

Schmid, M. S. (2011b). *Language attrition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511852046

Schmid, M. S. (2019). The impact of frequency of use and length of residence on l1 attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 287–303). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.24

Schmid, M. S., & Cherciov, M. (2019). Introduction to extralinguistic factors in language attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 265–276). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.22

Schmid, M. S., & Dusseldorp, E. (2010). Quantitative analyses in a multivariate study of language attrition: The impact of extralinguistic factors. *Second Language Research*, *26*(1), 125–160. https://doi.org/10.1177/0267658309337641

Schmid, M. S., & Jarvis, S. (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism: Language and Cognition*, *17*(4), 729–748. https://doi.org/10.1017/S1366728913000771

Schmid, M. S., & Keijzer, M. (2009). First language attrition and reversion among older migrants. *International Journal of the Sociology of Language*, *2009*(200), 83–101. https://doi.org/10.1515/IJSL.2009.046

Schmid, M. S., & Köpke, B. (2009). L1 attrition and the mental lexicon. In A. Pavlenko (Ed.), *The bilingual mental lexicon: Interdisciplinary approaches* (pp. 209–238). Multilingual Matters.

Schmid, M. S., & Köpke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism*, *7*(6), 637–667. https://doi.org/10.1075/lab.17058.sch

Schmid, M. S., & Köpke, B. (Eds.). (2019). *The Oxford handbook of language attrition* (1st ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.001.0001

Schmid, M. S., & Yılmaz, G. (2018). Predictors of language dominance: An integrated analysis of first language attrition and second language acquisition in late bilinguals. *Frontiers in Psychology*, *9*, Article 1306. https://doi.org/10.3389/fpsyg.2018.01306

Segalowitz, N. (1991). Does advanced skill in a second language reduce automaticity in the first language? *Language Learning*, *41*(1), 59–83. https://doi.org/10.1111/j.1467-1770.1991.tb00676.x

Sharwood Smith, M. (2019). Language attrition as a special case of processing change: A wider cognitive perspective. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 72–87). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.7

Soltani, M., Moradi, N., Rezaei, H., Hosseini, M., & Jasemi, E. (2021). Comparison of verbal fluency in monolingual and bilingual elderly in Iran. *Applied Neuropsychology: Adult*, *28*(1), 80–87. https://doi.org/10.1080/23279095.2019.1594234

Strauss, E., Sherman, E. M. S., Spreen, O., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). Oxford University Press.

Thiele, K., Quinting, J. M., & Stenneken, P. (2016). New ways to analyze word generation performance in brain injury: A systematic review and meta-analysis of additional performance measures. *Journal of Clinical and Experimental Neuropsychology*, *38*(7), 764–781. https://doi.org/10.1080/13803395.2016.1163327

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *14*(2), 167–177.

Treffers-Daller, J. (2019). What defines language dominance in bilinguals? *Annual Review of Linguistics*, *5*(1), 375–393. https://doi.org/10.1146/annurev-linguistics-011817-04555

Troyer, A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology*, *22*(3), 370–378. https://doi.org/10.1076/1380-3395(200006)22:3;1-V; FT370

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138

Venegas, M. J., & Mansur, L. L. (2011). Verbal fluency: Effect of time on item generation. *Dementia & Neuropsychologia*, *5*(2), 104–107. https://doi.org/10.1590/S1980-57642011DN05020008

Virtaranta, P. (1992). *Amerikansuomen sanakirja. A dictionary of American Finnish*. Siirtolaisuusinstituutti, Institute of Migration. https://urn.fi/URN:ISBN:951-9266-43-7

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., & Müller, K. (2019). *dplyr: A grammar of data manipulation*. https://CRAN.R-project.org/package=dplyr

Yılmaz, G. (2019). L1 attrition, L2 development, and integration. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 303–313). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.25

Yılmaz, G., & Schmid, M. S. (2018). First language attrition and bilingualism: Adult speakers. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Studies in bilingualism* (Vol.54, pp. 225–250). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.54.11yil

Yilmaz, G., & Schmid, M. S. (2012). L1 accessibility among Turkish-Dutch bilinguals. *The Mental Lexicon*, *7*(3), 249–274. https://doi.org/10.1075/ml.7.3.01yil

Author biographies

Nana Lehtinen is a Speech-Language Pathologist and a doctoral student at the University of Turku. Her research focuses on native language change and attrition in immigrant communities, particularly in the Finnish language in California, with the aim of providing insight into how immigrant settings influence the evolution of native languages. Additionally, she teaches Finnish at UC Berkeley, CA, USA enriching students' language and cultural understanding.

Anna Kautto is a researcher at Åbo Akademi University and the University of Turku, examining how linguistic skills intersect with non-linguistic information processing.

Dr. Kati Renvall, a Speech-Language Pathologist with a Ph.D. in psychology and holds the position of Senior Lecturer at the University of Turku. She also serves as a Docent at the University of Helsinki, Finland, and holds the title of Honorary Senior Research Fellow at Macquarie University, Australia. Her research primarily focuses on cognitive neuropsychology of language, aphasia, word-finding difficulties (anomia), and the assessment and treatment of adult neurogenic language disorders.

Appendix A

Language Attrition Test Battery Sociolinguistic Questionnaire

Sociolinguistic questionnaire, subjects

Identification code:

Date:

With this questionnaire I would like to get an impression of the personal background and language use of Finnish emigrants in USA. It consists of 71 items. It is important to note that not all items may apply to you personally. Should you think that a certain item does not apply to you (for example when you are asked about the language use of your children and you don't have any children), you may cross out the number in front of that particular question and move on to the next.

It is important that you answer these questions on your own, because I am interested in *your* language use. If you don't understand a certain question, please do not hesitate to ask me. There are no right or wrong answers!

1) What is your date of birth? ……………………month…………………year

2) Gender (at birth) ❑ 1. female ❑ 2. male ❑ 3. other ❑ 4. no reply

3) Where were you born:
Country :…………………………………………..……………………………
County: ……………………………………………………………………………
Ahvenanmaa, Etelä-Karjala, Etelä-Pohjanmaa, Etelä-Savo, Kanta-Häme, Keski-Pohjanmaa, Keski-Suomi, Kymenlaakso, Lappi, Pirkanmaa, Pohjanmaa, Pohjois-Karjala, Pohjois-Pohjanmaa, Pohjois-Savo, Päijät- Häme, Satakunta, Uusimaa, Varsinais-Suomi

Town/Village……………………………………………………………………………

4) What nationality do you have?
❑ 1. Finland
❑ 2. Finland and USA
❑ 3. USA
❑ 4. other, namely………………………………………………………………
❑ 5. no answer

5) Would you say that you spoke a variety of Finnish while you lived in the Finland or a dialect?
❑ standard Finnish
❑ a dialect, namely: ……………………………………………...

6) When did you come to USA (month/year) …………………………………………

7) Why did you emigrate and why to USA in particular?
❑ 1. job
❑ 2. partner's job
❑ 3. partner
❑ 4. other, namely: ……………………………………………………………

8) Apart from USA, have you ever lived in a country other than Finland for more than 6 months)?
❑ 1. no
❑ 2. less than 1 year: town…………………………country……………….…………..
❑ 3. 1 year or more: town…………………………..country……………………………

9) What language(s) did you acquire before starting school?
❑ 1. Finnish
❑ 2. Finnish and other, namely………………………………………………………….
❑ 3. other……………………………………………………………………….

10) Did you attend any English classes before coming to USA? (this has to be in an educational environment, like a school or some similar institution):
❑ 1. no
❑ 2. yes, less than 1 year
❑ 3. yes, more than 1 year

11) What is the highest level of education you have completed?
❑ 1. primary school, grades:…………………………………………………………….
❑ 2. secondary school, level: ……………………………………………………………
❑ 3. higher education, namely: ………………………………………………………….
❑ 4. University, degree: ………………………………………………………………….

12) What language was studies conducted in?
❑ 1. only English
❑ 2. both Finnish and English, but mostly English
❑ 3. both Finnish and English, equally
❑ 4. both Finnish and English, but mostly Finnish
❑ 5. only Finnish
❑ 6. other or no answer

13) Have you pursued further education while living in USA (this does not have to be language-related) ?
❑ 1. yes, for (number of years):……………………………………………………………
❑ 2. no

14) What language or languages did you learn professionally or at school?
……………………………………………………………………………………………….

15) What language or languages did you learn outside of school or work)?
…………………………………………………………………………………….…………

16) What has been the primary working language in your work career?

❑ 1. only English
❑ 2. both Finnish and English, but mostly English
❑ 3. both Finnish and English, equally
❑ 4. both Finnish and English, but mostly Finnish
❑ 5. only Finnish
❑ 6. other or no answer

17) What is your current profession? If you are retired, could you please indicate your last profession before retirement?

………………………………………………………………………………………………

18) Have you ever been diagnosed with a disorder that can influence speech or language processing, such as e.g. dyslexia?

❑ 1. no
❑ 2. yes, namely……………………………………………………………………………

19) Have you ever been diagnosed with an injury or illness that can influence your speech or language processing, such as e.g. stroke or neurological illness?

❑ 1. no
❑ 2. yes, namely……………………………………………………………………………

20) Do you have a hearing aid or have you discussed getting one with your doctor?

❑ 1. no
❑ 2. yes ……………………………………………………………………………………

21) Have you ever been back to Finland since leaving for USA?

❑ 1. never
❑ 2. seldom
❑ 3. regularly, 1-2 times a year
❑ 4. regularly, 3-5 times a year
❑ 5. regularly, over 5 times a year

22) If you have indicated that you have been back to Finland, could you please indicate what the reason or reasons for your visit were (you may tick more than one box)?

❑ 1. urgent family matters (such as a wedding or a funeral)
❑ 2. work
❑ 3. to visit without a particular reason
❑ 4. for another reason

23) Are you in contact with relatives and friends in Finland?

❑ 1. no
❑ 2. sporadically
❑ 3. couple of times per year
❑ 4. monthly
❑ 5. weekly

24) If yes, how do you keep in touch with those relatives and friends in Finland? (you can tick more than one box)
❑ 1. letters / post
❑ 2. telephone
❑ 3. text messages / instant messages
❑ 4. e-mail
❑ 5. social media, namely……………………………………………………………………
❑ 6. another way, namely: ……………………………………..……………………………

25) What language or languages do you mostly use to keep in touch with relatives and friends in Finland?
❑ only English
❑ both Finnish and English, but mostly English
❑ both Finnish and English, equally
❑ both Finnish and English, but mostly Finnish
❑ only Finnish
❑ other / no answer

26) Do you ever go to church in USA?
❑ 1. no, never
❑ 2. yes, sometimes
❑ 3. yes, regularly

27) If you have indicated you go to church, could you please indicate in which language the services are held?
❑ 1. English
❑ 2. Finnish
❑ 3. English and Finnish
❑ 4. other / no answer

28) In general, how would you rate your <u>English</u> language proficiency <u>before</u> you moved to USA?
❑ 1. none
❑ 2. bad
❑ 3. good enough to get by
❑ 4. good
❑ 5. very good

29) In general, how would you rate your <u>English</u> language proficiency <u>at present</u>?
❑ 1. none
❑ 2. bad
❑ 3. good enough to get by
❑ 4. good
❑ 5. very good

30) In general, how would you rate your Finnish language proficiency before you moved to USA?
❑ 1. none
❑ 2. bad
❑ 3. good enough to get by
❑ 4. good
❑ 5. very good

31) In general, how would you rate your Finnish language proficiency at present?
❑ 1. none
❑ 2. bad
❑ 3. good enough to get by
❑ 4. good
❑ 5. very good

32) How often do you speak Finnish?
❑ 1. rarely
❑ 2. few times a year
❑ 3. monthly
❑ 4. weekly
❑ 5. daily

33) Do you consider it important to maintain your Finnish?
❑ 1. no, unimportant
❑ 2. relatively unimportant
❑ 3. rather important
❑ 4. important /
❑ 5. yes, very important

34) Do you feel more at home with Finnish or with American culture?
❑ 1. with American culture
❑ 2. with both, but more with American culture
❑ 3. with both cultures, equally
❑ 4. with both, but more with Finnish culture
❑ 5. with Finnish culture

35) Do you feel more comfortable speaking Finnish or English?
❑ 1. English
❑ 2. Finnish
❑ 3. no preference

Could you elaborate on your answer: why do you feel more comfortable speaking either Finnish or English or why don't you have any preference?

…………………………………………………………………………………………

36) Have you ever been a member of a Finnish club or association while living in USA?
❑ 1. no
❑ 2. yes, name of organization……………………………………………………………
Member during years: ……………………….……..…… Hours per week…………….
37) Are you currently a member of a Finnish club or association?
❑ 1. no
❑ 2. yes, name of organization…………………………….………… Hours per week………
38) What is domestic situation at the moment?
❑ 1. I live alone
❑ 2. I have lived with a partner for a long time, but currently live alone
❑ 3. I live with a partner

39) With what language(s) was your (ex)partner brought up?
❑ 1. Finnish
❑ 2. English
❑ 3. other, namely: ………………………………………………………………………….

40) What language or languages do you mostly use when talking to your (ex)partner?
❑ 1. only English
❑ 2. both Finnish and English, but mostly English
❑ 3. both Finnish and English, without preference
❑ 4. both Finnish and English, but mostly Finnish
❑ 5. only Finnish
❑ 6. other or no answer

41) What language or languages does your (ex)partner mostly use when talking to you?
❑ 1. only English
❑ 2. both Finnish and English, but mostly English
❑ 3. both Finnish and English, without preference
❑ 4. both Finnish and English, but mostly Finnish
❑ 5. only Finnish
❑ 6. other or no answer

42) Do you have children?
❑ 1. no (move on to question 53)
❑ 2. yes

43) Do you consider it important that your children can speak and understand Finnish?
❑ 1. unimportant
❑ 2. relatively unimportant
❑ 3. not very important
❑ 4. important
❑ 5. very important

44) What language or languages do you mostly use when talking to your children?
- ❑ 1. only English
- ❑ 2. both Finnish and English, but mostly English
- ❑ 3. both Finnish and English, without preference
- ❑ 4. both Finnish and English, but mostly Finnish
- ❑ 5. only Finnish
- ❑ 6. other or no answer

45) What language or languages do your children mostly use when talking to you?
- ❑ 1. only English
- ❑ 2. both Finnish and English, but mostly English
- ❑ 3. both Finnish and English, without preference
- ❑ 4. both Finnish and English, but mostly Finnish
- ❑ 5. only Finnish
- ❑ 6. other or no answer

46) Do you encourage your children to speak Finnish?
- ❑ 1. no, never
- ❑ 2. seldom
- ❑ 3. sometimes
- ❑ 4. often
- ❑ 5. yes, all the time

47) Do / did your children ever take part in a Finnish club or associaltion (e.g. Suomikoulu)?
- ❑ 1. no
- ❑ 2. yes

48) Did /do you ever correct your children's Finnish?
- ❑ 1. no, never
- ❑ 2. seldom
- ❑ 3. sometimes
- ❑ 4. often
- ❑ 5. yes, all the time

49) If your children do not speak or understand Finnish, do you regret that?
- ❑ 1. not at all
- ❑ 2. not much
- ❑ 3. a bit
- ❑ 4. yes
- ❑ 5. yes, very much
- ❑ 6. no answer

50) Do you have grandchildren?
- ❑ 1. no (move on to question 53)
- ❑ 2. yes

51) What language or languages do you mostly use when talking to your grandchildren?
❑ 1. only English
❑ 2. both Finnish and English, but mostly English
❑ 3. both Finnish and English, without preference
❑ 4. both Finnish and English, but mostly Finnish
❑ 5. only Finnish
❑ 6. other or no answer

52) What language or languages do your grandchildren mostly use when talking to you?
❑ 1. only English
❑ 2. both Finnish and English, but mostly English
❑ 3. both Finnish and English, without preference
❑ 4. both Finnish and English, but mostly Finnish
❑ 5. only Finnish
❑ 6. other or no answer

53) Do you think Finnish plays an important role in the relationship between your direct family members?
❑ 1. no, Finnish language does not have any significance in our family
❑ 2. not much
❑ 3. probably
❑ 4. yes, has significance
❑ 5. yes, Finnish has an important role in our family
❑ 6. no answer

54) What is the mother tongue of the majority of your friends and acquaintances in the USA?
❑ 1. English
❑ 2. Finnish
❑ 3. bilingual (Finnish/English)
❑ 4. another language / no answer

55) Could you, in the following tables, please indicate to what extent you use Finnish (table 1) and English (table 2) in the domains provided? You may simply tick the box. If a certain domain is not applicable to you (for example, if you don't have any pets), you may leave the box empty.

	I speak Finnish					
		always	almost always	some-times	rarely	very rarely
1	With relatives	1	2	3	4	5
2	With friends	1	2	3	4	5
3	At work	1	2	3	4	5
4	In church	1	2	3	4	5
5	Shopping, running errands	1	2	3	4	5
6	Recreational activities	1	2	3	4	5

	I speak English					
		always	almost always	some-times	rarely	very rarely
1	With relatives	1	2	3	4	5
2	With friends	1	2	3	4	5
3	At work	1	2	3	4	5
4	In church	1	2	3	4	5
5	Shopping, running errands	1	2	3	4	5
6	Recreational activities	1	2	3	4	5

56) Could you, in the following table, please indicate which language you use more intuitively or instinctively in the following situations. Tick the language that you think is the right answer first. I am hoping to learn about your first reaction, so do not ponder for too long.

	Language choise					
		English	mostly English	both, Finnish and English	mostly Finnish	Finnish
1	Language in which you think	1	2	3	4	5
2	Language in which you dream	1	2	3	4	5
3	Language in which you count / do maths	1	2	3	4	5
4	Language in which you pray	1	2	3	4	5
5	Dominant language (i.e. language you are "best" at)	1	2	3	4	5
6	Language in which you have the largest vocabulary	1	2	3	4	5
7	Language in which you have no pronunciation difficulties	1	2	3	4	5
8	Language which you are able to understand / use intuitively	1	2	3	4	5
9	Language in which you are familiar with various dialects, slang	1	2	3	4	5
10	Language in which you have an intuitive feeling what is "correct" and "incorrect"	1	2	3	4	5
11	Language in which you can understand and make jokes	1	2	3	4	5
12	Language in which you swear	1	2	3	4	5
13	Language to which you have the strongest emotional ties	1	2	3	4	5
14	Language used most on a daily basis	1	2	3	4	5
15	Language which is your native language.	1	2	3	4	5
16	Language of which other speakers consider you a native speaker.	1	2	3	4	5
17	Language you use when talking to pets	1	2	3	4	5

57) Do you ever get homesick in the sense of missing Finland?
❑ 1. no
❑ 2. yes, what I then miss most is/are: ……………………………………………………
❑ 3. no answer

58) Do you listen to Finnish songs?
❑ 1. no
❑ 2. yes
❑ 3. I would like to, but I have no access to them

59) Do you watch Finnish television programmes?
❑ 1. no
❑ 2. yes
❑ 3. I would like to, but I have no access to them

60) Do you listen to Finnish radio programmes?
❑ 1. no
❑ 2. yes
❑ 3. I would like to, but I have no access to them

61) Do you read newspapers, books or magazines written in Finnish?
❑ 1. no
❑ 2. yes
❑ 3. I would like to, but I have no access to them

62) Do you write in Finnish?
❑ 1. no
❑ 2. yes, but prefer writing in English
❑ 3. yes

63) Do you think your <u>Finnish language</u> proficiency has changed since you moved to USA?
❑ 1. yes, I think it has become worse
❑ 2. no
❑ 3. yes, I think it has become better

64) Do you think you use more or less Finnish since you moved to USA?
❑ 1. yes, I think I use less Finnish
❑ 2. no, I don't think I use more or less Finnish now
❑ 3. yes, I think I use more Finnish

65) Do you ever feel uncomfortable when speaking Finnish with a Finnish person who has not spent a considerable amount of time in an English-speaking country?
❑ 1. no, never
❑ 2. yes, sometimes

66) Do you ever feel uncomfortable when speaking Finnish with someone who, like you, has lived in USA for a long time?

❑ 1. no, never

❑ 2. yes, sometimes

67) Do you see yourself as bilingual?

❑ 1. no

❑ 2. yes

❑ 3. other, namely……………………………………………………………………

❑ 4. no answer

68) Do you feel that you are equally proficient in Finnish and English?

❑ 1. no, I'm more proficient in English

❑ 2. yes

❑ 3. no, I'm more proficient in Finnish

❑ 4. I don't know, because:……………………………………………………………..

69) Has the balance between your Finnish and English changed during the time you have lived in the USA?

❑ 1. no

❑ 2. yes, please elaborate

………………………………………………………………………………………

………………………………………………………………………………………

70) Do you ever intend to move back to Finland?

❑ 1. no, I don't intend to ever return to Finland

❑ 2. yes, I would eventually like to move back to Finland

❑ 3. I have never really given it much thought

❑ 4. no answer

Would you like to explain why you feel that way?:

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

71) You have come to the end of this questionnaire. Thank you for your time!
Is there anything you would like to add? Here you can add anything from language-related comments to remarks about the questionnaire or research itself.

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

Appendix B
An abridged questionnaire for the control group

Sosiolingvistinen kyselylomake, verrokkihenkilöt

Nimi:

Päivämäärä:

Tällä lomakkeella kerätään tutkimukseen osallistuvilta henkilöiltä tietoa suomen kielen käytöstä sekä taustatietoja. Lomake sisältää 31 kohtaa. Huomaathan että kaikki kohdat eivät koske jokaista tutkimukseen osallistuvaa henkilöä. Mikäli koet, että tietty kysymys ei koske sinua *(kuten esimerkiksi kysymykset lastenlasten kanssa käytettävästä kielestä eikä sinulla ole lapsenlapsia)*, voit jättää kysymyksen huomiotta ja siirtyä seuraavaan kysymykseen.

On tärkeää että vastaat jokaiseen kysymykseen itse, sillä tutkimuksessa kerätään tietoa juuri sinun kielestäsi ja sen käytöstä. Jos jokin kysymys on epäselvä tai vaikea ymmärtää, älä epäröi kysyä lisätietoa minulta. Huomaathan, että kysymyksiin ei ole olemassa oikeita tai vääriä vastauksia.

1) Syntymäaika ____________________

2) Sukupuoli (syntymässä) ❑ 1. nainen ❑ 2. mies ❑ 3. muu ❑ 4. ei vastausta

3) Syntymäpaikka

Maa: ____________________

Maakunta: ____________________

Ahvenanmaa, Etelä-Karjala, Etelä-Pohjanmaa, Etelä-Savo, Kanta-Häme, Keski-Pohjanmaa, Keski-Suomi, Kymenlaakso, Lappi, Pirkanmaa, Pohjanmaa, Pohjois-Karjala, Pohjois-Pohjanmaa, Pohjois-Savo, Päijät-Häme, Satakunta, Uusimaa, Varsinais-Suomi

Kaupunki/Kylä: ____________________

4) Puhutko yleiskieltä vai jotakin tiettyä murretta?
❑ 1. Yleiskieltä
❑ 2. Murretta, tarkemmin ____________________

5) Oletko koskaan asunut muussa maassa kuin Suomessa yli 6 kuukauden ajan
❑ 1. en
❑ 2. alle 1 vuosi: kaupunki__________maa__________
❑ 3. yli 1 vuosi: kaupunki__________maa__________

6) Mitä kieliä opit lapsena (ennen kouluikää)?
❑ 1. suomi
❑ 2. suomi ja jokin muu, mikä ____________________
❑ 3. jokin muu, mikä ____________________

7) Mikä on korkein kouluaste jonka olet suorittanut?
❑ 1. peruskoulu, luokat ____________________
❑ 2. toisen asteen tutkinto, taso ____________________
❑ 3. korkeakoulututkinto, tarkemmin ____________________
❑ 4. yliopistotutkinto, tarkemmin ____________________

8) Millä kielellä opiskelu toteutui?

❑ 1. suomeksi
❑ 2. sekä suomeksi että toisella kielellä, mutta pääasiassa suomeksi
❑ 3. kahdella kielellä tasavertaisesti
❑ 4. sekä suomeksi että toisella kielellä, mutta pääasiassa toisella kielellä
❑ 5. __________________________(kieli)
❑ 6. ei vastausta

9) Mikä on nykyinen ammattisi? Jos olet eläkkeellä, mikä oli ammatti josta jäit eläkkeelle?

____

10) Millä kielellä toimit työssäsi?

❑ 1. suomeksi
❑ 2. sekä suomeksi että toisella kielellä, mutta pääasiassa suomeksi
❑ 3. kahdella kielellä tasavertaisesti
❑ 4. sekä suomeksi että toisella kielellä, mutta pääasiassa toisella kielellä
❑ 5. __________________________(kieli)
❑ 6. ei vastausta

11) Onko sinulla todettu jokin kielellisiin toimintoihin vaikuttava häiriö kuten esim. lukihäiriö?

❑ 1. ei
❑ 2. kyllä Tarkemmin:____

12) Onko sinulla joskus todettu jokin vamma tai sairaus joka on vaikuttanut puhe- tai kielikykyysi (kuten aivoverenkiertohäiriö tai neurologinen sairaus)?

❑ 1. ei
❑ 2. kyllä Tarkemmin:____

13) Onko sinulla kuulokoje tai oletko keskustellut lääkärisi kanssa sellaisen hankkimisesta?

❑ 1. ei
❑ 2. kyllä, tarkemmin:____

14) Oletko oikea-vai vasenkätinen?

❑ 1. oikea
❑ 2. vasen
❑ 4. molempikätinen
❑ 4. ei vastausta

15) Oletko kaksikielinen?

❑ 1. ei
❑ 2. kyllä
❑ 4. muu
❑ 4. ei vastausta

16) Mitä kieliä käytät arjessasi?
- ❑ 1. suomea
- ❑ 2. sekä suomea että toista kieltä, mutta pääasiassa suomea
- ❑ 3. kahta kieltä tasavertaisesti
- ❑ 4. sekä suomea että toista kieltä, mutta pääasiassa toista kieltä
- ❑ 5. ________________________(kieli)
- ❑ 6. ei vastausta

17) Asumismuoto
- ❑ 1. Asun yksin
- ❑ 2. Olen asunut pitkään kumppanin kanssa, mutta asun tällä hetkellä yksin
- ❑ 3. Asun kumppanin kanssa

18) Mitä kieltä/kieliä (ex)kumppanisi puhui lapsuudenkodissaan?
- ❑ 1. suomea
- ❑ 2. muu, mikä: ____

19) Mitä kieltä/kieliä käytät puhuessasi (ex)kumppanisi kanssa?
- ❑ 1. suomea
- ❑ 2. sekä suomea että toista kieltä, mutta pääasiassa suomea
- ❑ 3. kahta kieltä tasavertaisesti
- ❑ 4. sekä suomea että toista kieltä, mutta pääasiassa toista kieltä
- ❑ 5. ________________________(kieli)
- ❑ 6. ei vastausta

20) Mitä kieltä/kieliä (ex)kumppanisi käyttää puhuessaan sinulle?
- ❑ 1. suomea
- ❑ 2. sekä suomea että toista kieltä, mutta pääasiassa suomea
- ❑ 3. kahta kieltä tasavertaisesti
- ❑ 4. sekä suomea että toista kieltä, mutta pääasiassa toista kieltä
- ❑ 5. ________________________(kieli)
- ❑ 6. ei vastausta

21) Onko sinulla lapsia?
- ❑ 1. ei
- ❑ 2. kyllä

22) Mitä kieltä / kieliä käytät puhuessasi lapsillesi?
- ❑ 1. suomea
- ❑ 2. sekä suomea että toista kieltä, mutta pääasiassa suomea
- ❑ 3. kahta kieltä tasavertaisesti
- ❑ 4. sekä suomea että toista kieltä, mutta pääasiassa toista kieltä
- ❑ 5. ________________________(kieli)
- ❑ 6. ei vastausta

23) Mitä kieltä/kieliä lapset käyttävät puhuessaan sinulle?
- ❑ 1. suomea
- ❑ 2. sekä suomea että toista kieltä, mutta pääasiassa suomea
- ❑ 3. kahta kieltä tasavertaisesti
- ❑ 4. sekä suomea että toista kieltä, mutta pääasiassa toista kieltä
- ❑ 5. ___________________________(kieli)
- ❑ 6. ei vastausta

24) Onko sinulla lastenlapsia?
- ❑ 1. ei
- ❑ 2. kyllä

25) Mitä kieltä / kieliä käytät pääsääntöisesti puhuessasi lapsenlapsillesi?
- ❑ 1. suomea
- ❑ 2. sekä suomea että toista kieltä, mutta pääasiassa suomea
- ❑ 3. kahta kieltä tasavertaisesti
- ❑ 4. sekä suomea että toista kieltä, mutta pääasiassa toista kieltä
- ❑ 5. ___________________________(kieli)
- ❑ 6. ei vastausta

26) Mitä kieltä/kieliä lastenlapset käyttävät pääsääntöisesti puhuessaan sinulle?
- ❑ 1. suomea
- ❑ 2. sekä suomea että toista kieltä, mutta pääasiassa suomea
- ❑ 3. kahta kieltä tasavertaisesti
- ❑ 4. sekä suomea että toista kieltä, mutta pääasiassa toista kieltä
- ❑ 5. ___________________________(kieli)
- ❑ 6. ei vastausta

27) Kuunteletko suomalaista musiikkia / suomalaisia lauluja?
- ❑ 1. en
- ❑ 2. kyllä
- ❑ 3. ei vastausta

28) Katsotko suomalaisia televisio-ohjelmia?
- ❑ 1. en
- ❑ 2. kyllä
- ❑ 3. ei vastausta

29) Kuunteletko suomalaista radiota/suomalaisia radio-ohjelmia?
- ❑ 1. en
- ❑ 2. kyllä
- ❑ 3. ei vastausta

30) Luetko suomalaisia kirjoja, sanomalehtiä tai aikakauslehtiä?
- ❑ 1. en
- ❑ 2. kyllä

❑ 3. ei vastausta

31) Olet tullut haastattelulomakkeen viimeiseen kohtaan. Kiitos yhteistyöstä! Onko jotain mitä haluaisit vielä lisätä? Allaolevaan tilaan voit kirjata lisätietoja tai kommentteja sekä tähän haastattelulomakkeeseen että tutkimukseen liittyen

____

____

____

____

Sosiolinguistic questionnaire, controls

Name:

Date:

This questionnaire is used to collect information on Finnish language use and background factors relevant to this study. It consists of 31 items. It is important to note that not all items may apply to you personally. Should you think that a certain item does not apply to you *(for example when you are asked about the language use of your children and you don't have any children)*, you may cross out the number in front of that particular question and move on to the next.

It is important that you answer these questions on your own, because I am interested in *your* language use. If you don't understand a certain question, please do not hesitate to ask me. There are no right or wrong answers!

1) Date of birth ____

2) Gender (at birth) ❑ 1. female ❑ 2. male ❑ 3. other ❑ 4. No answer

3) Where were you born:

Country:

____

County: ____

Ahvenanmaa, Etelä-Karjala, Etelä-Pohjanmaa, Etelä-Savo, Kanta-Häme, Keski-Pohjanmaa, Keski-Suomi, Kymenlaakso, Lappi, Pirkanmaa, Pohjanmaa, Pohjois-Karjala, Pohjois-Pohjanmaa, Pohjois-Savo, Päijät-Häme, Satakunta, Uusimaa, Varsinais-Suomi

Town/Village:____

4) Would you say you speak standard Finnish or a dialect?
❑ 1. Standard Finnish
❑ 2. A dialect, namely ____________________________________

5) Have you ever lived in a country other than Finland for more than 6 months?
❑ 1. No
❑ 2. Less than 1 year: town_______________country___________________
❑ 3. 1 year or more: town_______________country_____________________

6) What language(s) did you acquire before starting school?
❑ 1. Finnish
❑ 2. Finnish and other, namely ____________________________________
❑ 3. Other, namely ____

7) What is the highest level of education you have completed??
❑ 1. Primary school, grades ____________________________________
❑ 2. Secondary school, level ____________________________________
❑ 3. Higher education, namely ___________________________________
❑ 4. University Degree, namely __________________________________

8) What languages were studies conducted in?
❑ 1. Only Finnish
❑ 2. Finnish and another language but mostly Finnish
❑ 3. Finnish and another language, equally
❑ 4. Finnish and another language, but mostly in another language
❑ 5. Only _________________________(language)
❑ 6. No answer

9) What is your current profession? If you are retired, could you please indicate your last profession before retirement?

____

10) What has been the primary working language during your career?
❑ 1. Finnish
❑ 2. Finnish and another language but mostly Finnish
❑ 3. Finnish and another language, equally
❑ 4. Finnish and another language, but mostly in another language
❑ 5. Only _________________________(language)
❑ 6. No answer

11) Have you ever been diagnosed with a disorder that can influence speech and language processing, such as e.g. dyslexia?
❑ 1. No
❑ 2. Yes, namely____

12) Have you ever been diagnosed with an injury or illness that can influence your speech or language processing, such as e.g. stroke or neurological illness?
❑ 1. No
❑ 2. Yes, namely____

13) Do you have a hearing aid or have you discussed getting one with your doctor?
❑ 1. No
❑ 2. Yes ____

14) Are you right or left handed?
❑ 1. Right
❑ 2. Left
❑ 4. Both
❑ 4. No answer

15) Are you bilingual?
❑ 1. No
❑ 2. Yes
❑ 4. Other, namely____
❑ 4. No answer

16) What languages do you use in your everyday life?
- ❑ 1. Finnish
- ❑ 2. Finnish and another language but mostly Finnish
- ❑ 3. Finnish and another language, equally
- ❑ 4. Finnish and another language, but mostly in another language
- ❑ 5. Only ______________________________(language)
- ❑ 6. No answer

17) What is your domestic situation at the moment?
- ❑ 1. I live alone
- ❑ 2. I have lived with a partner for a long time, but currently live alone
- ❑ 3. I live with a partner

18) With what language(s) was your (ex)partner brought up?
- ❑ 1. Finnish
- ❑ 2. Other, namely ____

19) What language or languages do you mostly use when talking to your (ex)partner?
- ❑ 1. Finnish
- ❑ 2. Finnish and another language but mostly Finnish
- ❑ 3. Finnish and another language, equally
- ❑ 4. Finnish and another language, but mostly in another language
- ❑ 5. Only ______________________________(language)
- ❑ 6. No answer

20) What language or languages does your (ex)partner mostly use when talking to you?
- ❑ 1. Finnish
- ❑ 2. Finnish and another language but mostly Finnish
- ❑ 3. Finnish and another language, equally
- ❑ 4. Finnish and another language, but mostly in another language
- ❑ 5. Only ______________________________(language)
- ❑ 6. No answer

21) Do you have children?
- ❑ 1. No
- ❑ 2. Yes

22) What language or languages do you use when talking to your children?
- ❑ 1. Finnish
- ❑ 2. Finnish and another language but mostly Finnish
- ❑ 3. Finnish and another language, equally
- ❑ 4. Finnish and another language, but mostly in another language
- ❑ 5. Only ______________________________(language)
- ❑ 6. No answer

23) What language or languages do your children mostly use when talking to you?
- ❑ 1. Finnish
- ❑ 2. Finnish and another language but mostly Finnish
- ❑ 3. Finnish and another language, equally
- ❑ 4. Finnish and another language, but mostly in another language
- ❑ 5. Only _________________________(language)
- ❑ 6. No answer

24) Do you have grandchildren?
- ❑ 1. No
- ❑ 2. Yes

25) What language or languages do you use when talking to your grandchildren?
- ❑ 1. Finnish
- ❑ 2. Finnish and another language but mostly Finnish
- ❑ 3. Finnish and another language, equally
- ❑ 4. Finnish and another language, but mostly in another language
- ❑ 5. Only _________________________(language)
- ❑ 6. No answer

26) What language or languages do your grandchildren mostly use when talking to you
- ❑ 1. Finnish
- ❑ 2. Finnish and another language but mostly Finnish
- ❑ 3. Finnish and another language, equally
- ❑ 4. Finnish and another language, but mostly in another language
- ❑ 5. Only _________________________(language)
- ❑ 6. No answer

27) Do you listen to Finnish songs?
- ❑ 1. No
- ❑ 2. Yes
- ❑ 3. No answer

28) Do you watch Finnish television programmes?
- ❑ 1. No
- ❑ 2. Yes
- ❑ 3. No answer

29) Do you listen to Finnish radio programmes?
- ❑ 1. No
- ❑ 2. Yes
- ❑ 3. No answer

30) Do you read newspapers, books or magazines written in Finnish?
- ❑ 1. No
- ❑ 2. Yes

❑ 3. No answer

31) You have come to the end of this questionnaire. Thank you for your time!

Is there anything you would like to add? Here you can add anything from language-related comments to remarks about the questionnaire or research itself.

____

____

____

____

Appendix C

Example Trials, Descriptive Statistics and Model Summaries

Table 1. *Example of Three Phonemic and One Semantic trials in L1*

/k/	**/a/**	**/p/**	**Eläimet** (animals)
kulta	aalto	peli	kissa
kala	alistaa	pelata	koira
kiekko	aalloissa	pilkku	hiiri
kissa	auttaa	pissa	rotta
kello		piru	karhu
kalkkuna		perkele	susi
kilisee		paatti	hyeena
kulkunen			kirahvi
15s	**15s**	**15s**	**15s**
kirjava	antaa	passata	leijona
kertoo	annostaa	pois	tiikeri
katsoo	antoisa	pelko	norsu
kalastaa	antava	~~peli~~ (1)	elefantti
kolkottaa	astua	pirullinen	poro
			hirvi
			~~mountain lion~~ (4)
30s	**30s**	**30s**	**30s**
kilkuttaa	~~astua~~ (1)	parsa	lokki
kallistaa	astuva	perse	tiainen
~~killustaa~~ (3)	~~alistaa~~ (1)	pelottava	varis
~~fast~~		pelli	talitintti
		pellillinen	
45s	**45s**	**45s**	**45s**
korko	~~akvaattinen~~ (3)	~~pärssi~~ (3)	käärme
kestävä		~~pärsyinen~~ (3)	lohi
kissa		~~Pöyri~~ (2)	ahven
			silakka
60s	**60s**	**60s**	**60s**

Note. Error type are marked in brackets (1 - repetition, 2 - categorical, 3 - nonword, 4 - intrusion)

Table 2. *Example of Three Phonemic and One Semantic trials in L2*

/f/	/a/	/s/	Animals
frisky	apple	sea	dog
fallible	alley	season	cat
fish	air	sink	mouse
frankly	aerial	silly	elephant
	aisle	sissy	lion
			tiger
			seagull
			snake
			hippo
15s	**15s**	**15s**	**15s**
ferocious	ant	sassy	fish
finicky	aunt	sank	rat
fast	anxious	silence	whale
ferry	antsy	sincere	
fig			
30s	**30s**	**30s**	**30s**
fell	audacious	sew	giraffe
fall	audible	~~cynical~~ (2)	hyena
fag	anchor		moose
~~fast~~ (1)	~~anchor~~ (1)		reindeer
	~~alley~~ (1)		bear
45s	**45s**	**45s**	**45s**
foliage		soul	~~fish~~ (1)
fig		soulful	
furious		sail	
		sailing	
60s	**60s**	**60s**	**60s**

Note. Error type is marked in brackets (1 - repetition, 2 - categorical, 3 - nonword, 4 - intrusion)

Table 3. *Mean, Standard Deviation, and Range for Total Scores and Temporal Parameters in One Semantic and Three Phonemic Verbal Fluency Tasks in the Attriter Group in L1 and L2*

Variable	L1 PVF /k/	L1 PVF /a/	L1 PVF /p/	L1 SVF animals	L2 PVF /f/	L2 PVF /a/	L2 PVF /s/	L2 SVF animals
Total score 0 – 60 s								
Mean	18.92	14.90	16.45	21.90	14.05	10.60	15.32	20.60
SD	4.94	5.03	4.54	4.31	4.23	5.13	5.21	4.85

Range	11 – 33	6 – 24	7 – 27	13 – 33	7 – 27	1 – 20	6 – 28	9 – 32
Time segment 0 – 15 s								
Mean	7.30	5.90	6.47	9.84	5.78	4.68	5.81	8.95
SD	2.07	1.91	1.64	2.02	2.00	2.08	1.73	2.07
Range	3 – 12	2 – 10	3 – 10	5 – 14	2 – 11	1 – 9	3 – 10	4 – 14
% of total score	38.58	39.60	39.33	44.93	41.14	44.15	37.92	43.45
Time segment 16 – 30 s								
Mean	4.40	3.80	4.05	5.90	2.89	2.14	4.00	5.05
SD	1.65	1.68	1.77	2.05	1.66	1.40	1.92	1.79
Range	2 – 8	0 – 7	1 – 9	1 – 9	0 – 6	1 – 9	0 – 8	2 – 9
% of total score	23.26	25.50	24.62	26.94	20.57	20.19	26.11	24.51
Time segment 31 – 45 s								
Mean	3.42	2.87	3.40	3.71	2.97	2.11	2.92	3.78
SD	1.61	1.82	1.50	1.84	1.50	1.65	1.74	2.04
Range	0 – 7	0 – 8	1 – 6	0 – 8	0 – 6	0 – 7	0 – 7	0 – 9
% of total words	18.08	19.26	20.67	16.94	21.14	19.91	19.06	18.35
Time segment 46– 60 s								
Mean	3.82	2.34	2.53	3.05	2.41	1.68	2.60	2.81
SD	1.67	1.48	1.62	1.91	1.50	1.47	1.72	1.93
Range	1 – 9	0 – 5	0 – 6	0 – 8	0 – 6	0 – 6	0 – 8	0 – 8
% of total score	20.19	15.70	15.38	13.93	17.15	15.85	16.97	13.64

Note. L1 *N* = 38; L2 *N* = 37; SVF = Semantic Verbal Fluency; PVF = Phonemic Verbal Fluency; Total score = the total number of acceptable words generated in a 60 s trial; Time segment = the number of acceptable words generated in a 15 s time segment within a 60 s trial.

Table 4. *Mean, Standard Deviation, and Range for Total Scores and Temporal Parameters in One Semantic and Three Phonemic Verbal Fluency Tasks in the Monolingual Group in L1*

Variable	L1 PVF /k/	L1 PVF /a/	L1 PVF /p/	L1 SVF animals

Total score 0 – 60 s				
Mean	19.38	14.68	17.44	25.96
SD	6.12	5.65	5.50	5.90
Range	5 – 35	5 – 31	4 – 30	12 – 43
Time segment 0 – 15 s				
Mean	7.06	5.72	6.26	10.76
SD	2.37	2.25	2.20	2.14
Range	2 – 12	1 – 11	2 – 12	6 – 15
% of total score	36	39	36	41
Time segment 16 – 30 s				
Mean	4.48	3.54	4.24	6.06
SD	2.04	1.94	1.69	2.13
Range	0 – 9	0 – 8	1 – 8	0 – 10
% of total score	23	24	24	23
Time segment 31 – 45 s				
Mean	4.12	2.98	3.48	4.94
SD	1.90	1.64	1.66	2.36
Range	0 – 9	0 – 7	0 – 7	0 – 13
% of total words	21	20	20	19
Time segment 46– 60 s				
Mean	3.72	2.44	3.46	4.20
SD	1.51	1.57	1.85	2.36
Range	0 – 7	0 – 7	0 – 8	0 – 10
% of total score	19	17	20	16

Note. Total score = the total number of acceptable words generated in a 60 s trial; Time segment = number of acceptable words generated in a 15 s time segment within a 60 s trial.

Model Summary 1. *Predicted Values of Number of Correct Words in the Attriter Group in One Semantic and Three Phonemic Tasks (Combined) L1 and L2 with Frequency of L1 Use as a Predictor.*

Predictor	Estimate	95% CI		*t*	*p*
		LL	*UL*		
(Intercept)	13.28	12.18	14.39	23.69	**<0.001**
TaskLanguage [L1]	3.47	2.34	4.61	6.02	**<0.001**
Task Type [semantic]	7.27	5.74	4.61	9.35	**<0.001**
Frequency of L1 Use c	1.46	0.36	2.55	2.62	**0.009**
TaskLanguage [L1] * Task Type [semantic]	-2.13	-4.28	0.02	-1.95	0.052
TaskLanguage [L1] * Frequency of L1 Use c	0.32	-0.81	1.44	0.55	0.580
Task Type [semantic] * Frequency of L1 Use c	0.40	-1.11	1.92	0.52	0.602
(TaskLanguage [L1] * Task Type [semantic]) * Frequency of L1 Use c	-0.17	-2.31	1.97	-0.16	0.877
Random Effects					
σ2	16.79				
τ00 ID	6.11				
τ11 ID.taskLanguageL1	1.21				
ρ01 ID	-0.29				
ICC	0.26				
N ID	38				
Observations	300				
Marginal R2 / Conditional R2	0.356 / 0.524				

Model Summary 2. *Predicted Values of Correct Words in the Four 15-Second Time Segments in One Semantic and Three Phonemic Verbal Fluency Tasks in the Attriter Group in L1 and L2.*

Predictor	Estimate	95% CI		*t*	*p*
		LL	*UL*		
(Intercept)	5.41	5.04	5.79	28.17	**<0.001**
TaskLanguage [L1]	1.14	0.68	1.60	4.86	**<0.001**
which15 [16_30s]	-2.41	-2.85	-1.98	-10.98	**<0.001**
which15 [31_45s]	-2.76	-3.19	-2.33	-12.54	**<0.001**
which15 [46_60s]	-3.20	-3.63	-2.77	-14.55	**<0.001**
Task Type [semantic]	3.52	2.91	4.13	11.33	**<0.001**
Frequency of L1 Use c	0.43	0.06	0.81	2.28	**0.022**
Task Language [L1] * which15 [16_30s]	-0.06	-0.67	0.55	-0.19	0.847
Task Language [L1] * which15 [31_45s]	-0.57	-1.17	0.04	-1.84	0.066
Task Language [L1] * which15 [46_60s]	-0.46	-1.06	0.15	-1.49	0.138
Task Language [L1] * Task Type [semantic]	-0.23	-1.09	0.62	-0.53	0.595
which15 [16_30s] * Task Type [semantic]	-1.48	-2.34	-0.61	-3.36	**0.001**
which15 [31_45s] * Task Type [semantic]	-2.40	-3.27	-1.54	-5.47	**<0.001**
which15 [46_60s] * Task Type [semantic]	-2.94	-3.80	-2.07	-6.68	**<0.001**
Task Language [L1] * Frequency of L1 Use c	0.25	-0.20	0.71	1.09	0.275

which15 [16_30s] * Frequency of L1 Use c	0.06	-0.37	0.49	0.28	0.779
which15 [31_45s] * Frequency of L1 Use c	-0.10	-0.53	0.33	-0.45	0.650
which15 [46_60s] * Frequency of L1 Use c	-0.24	-0.67	0.18	-1.12	0.265
Task Type [semantic] * Frequency of L1 Use c	-0.19	-0.79	0.41	-0.62	0.533
(Task Language [L1] which15 [16_30s]) Task Type [semantic]	-0.60	-1.81	0.61	-0.98	0.330
(Task Language [L1] which15 [31_45s]) Task Type [semantic]	-0.40	-1.62	0.81	-0.65	0.513
(Task Language [L1] which15 [46_60s]) Task Type [semantic]	-0.20	-1.41	1.02	-0.32	0.751
(Task Language [L1] which15 [16_30s]) Frequency of L1 Use c	-0.28	-0.89	0.32	-0.92	0.356
(Task Language [L1] which15 [31_45s]) Frequency of L1 Use c	-0.35	-0.96	0.25	-1.15	0.250
(Task Language [L1] which15 [46_60s]) Frequency of L1 Use c	-0.06	-0.66	0.54	-0.20	0.840
(Task Language [L1] * Task Type [semantic]) * Frequency of L1 Use c	-0.05	-0.90	0.81	-0.11	0.913
(which15 [16_30s] * Task Type [semantic]) * Frequency of L1 Use c	0.21	-0.64	1.06	0.48	0.630

(which15 [31_45s] * Task Type [semantic]) * Frequency of L1 Use c	0.55	-0.30	1.40	1.26	0.207
(which15 [46_60s] * Task Type [semantic]) * Frequency of L1 Use c	0.41	-0.44	1.26	0.95	0.344
(Task Language [L1] which15 [16_30s] Task Type [semantic]) * Frequency of L1 Use c	0.43	-0.77	1.64	0.70	0.482
(Task Language [L1] which15 [31_45s] Task Type [semantic]) * Frequency of L1 Use c	0.01	-1.20	1.22	0.02	0.987
(Task Language [L1] which15 [46_60s] Task Type [semantic]) * Frequency of L1 Use c	-0.42	-1.63	0.78	-0.69	0.493
Random Effects					
σ2	2.68				
τ00 ID	0.48				
τ11 ID.taskLanguagefi	0.27				
ρ01 ID	-0.41				
ICC	0.15				
N ID	38				
Observations	1200				
Marginal R2 / Conditional R2	0.542 / 0.609				

Model summary 3. *Predicted Values of the Quotient of Correct Words in the First 15-Second Time Segment in the Attriters Group in L1 and L2 (Combined) in All Task Types (Combined) with Frequency of L1 Use as Predictor*

Predictor	Estimate	95% CI		*t*	*p*
		LL	*UL*		
(Intercept)	0.44	0.39	0.48	20.36	**<0.001**
Task Language [L1]	-0.02	-0.08	0.48	-0.69	0.493
Frequency of L1 use c	-0.02	-0.04	-0.01	-2.52	**0.012**
Task Language [L1] * Frequency of L1 use c	0.02	-0.01	0.04	1.37	0.172
Random Effects					
σ2	0.01				
τ00 taskLanguage	0.00				
ICC	0.03				
N taskLanguage	2				
Observations	300				
Marginal R2 / Conditional R2	0.029 / 0.057				

Model Summary 4. *Predicted Values of Number of Correct Words in One Semantic and Three Phonemic Verbal Fluency Tasks (Combined) in L1 in the Attriter Group and Monolingual Group.*

Predictor	Estimate	95% CI		*t*	*p*
		LL	*UL*		
(Intercept)	16.75	15.19	18.31	21.12	**<0.001**
Task Type [semantic]	5.14	3.44	6.84	5.95	**<0.001**
Group [monolingual]	0.41	-1.66	2.48	0.39	0.696

Task Type [semantic] * Group [monolingual]	3.65	1.40	5.91	3.19	**0.002**
Random Effects					
σ2	12.39				
τ00 ID	19.79				
τ11 ID.flu_typesem	11.88				
ρ01 ID	-0.53				
ICC	0.60				
N ID	88				
Observations	352				
Marginal R2 / Conditional R2	0.259 / 0.705				

Model summary 5. *Predicted Values of Words Produced in the Four 15 s Time Segments in One Semantic and Three Phonemic Verbal Fluency Tasks in the Attriter and Monolingual Groups*

Predictor	Estimate	95% CI		*t*	*p*
		LL	*UL*		
(Intercept)	6.55	6.11	7.00	28.85	**<0.001**
which15 [16_30s]	-2.47	-2.90	-2.04	-11.32	**<0.001**
which15 [31_45s]	-3.32	-3.75	-2.90	-15.21	**<0.001**
which15 [46_60s]	-3.66	-4.09	-3.23	-16.74	**<0.001**
Group [Monolingual]	-0.21	-0.80	0.38	-0.71	0.481
Task Type [semantic]	3.29	2.68	3.90	10.64	**<0.001**
which15 [16_30s] * Group [Monolingual]	0.22	-0.35	0.79	0.76	0.447

which15 [31_45s] * Group [Monolingual]	0.51	-0.06	1.08	0.76	0.078
which15 [46_60s] * Group [Monolingual]	0.53	-0.04	1.10	1.83	0.067
which15 [16_30s] * Task Type [semantic]	-2.08	-2.94	-1.22	-4.76	**<0.001**
which15 [31_45s] * Task Type [semantic]	-2.81	-3.66	-1.95	-6.42	**<0.001**
which15 [46_60s] * Task Type [semantic]	-3.13	-3.99	-2.27	-7.16	**<0.001**
Group [Monolingual] * Task Type [semantic]	1.13	0.33	1.93	2.76	**0.006**
(which15 [16_30s] Group [Monolingual]) Task Type [semantic]	-0.37	-1.51	0.77	-0.63	0.526
(which15 [31_45s] Group [Monolingual]) Task Type [semantic]	-0.20	-1.34	0.94	-0.34	0.731
(which15 [46_60s] Group [Monolingual]) Task Type [semantic]	-0.30	-1.44	0.84	-0.52	0.603
Random Effects					
σ2	2.72				
τ00 ID	1.05				
ICC	0.28				
N ID	88				
Observations	1408				
Marginal R2 / Conditional R2	0.492 / 0.634				

Model summary 6. *Predicted Values of the Quotient of Correct Words in the First 15-Second Time Segment All Task Types (Combined) in the Attriter and Monolingual groups*

Predictor	Estimate	95% CI		*t*	*p*
		LL	*UL*		
(Intercept)	0.43	0.37	0.48	15.92	**<0.001**
Group [Monolingual]	-0.02	-0.04	0.00	-2.29	**0.022**
Random Effects					
σ2	0.01				
τ00 flu_type	0.00				
ICC	0.12				
N flu_type	2				
Observations	352				
Marginal R2 / Conditional R2	0.013 / 0.136				

Lehtinen, N., Kautto, A., & Renvall, K. (2024)

Efficacy of clustering and switching strategies in verbal fluency tasks in a Finnish-English language attrition population.

The Language Learning Journal

III

THE LANGUAGE LEARNING JOURNAL
2024, VOL. 52, NO. 2, 218–231
https://doi.org/10.1080/09571736.2023.2294060

Efficacy of clustering and switching strategies in verbal fluency tasks in a Finnish-English language attrition population

Nana Lehtinen [a,b], Anna Kautto [a] and Kati Renvall [a,c]

[a]Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland; [b]Department of Scandinavian, University of California, Berkeley, CA, USA; [c]School of Psychological Sciences, Macquarie University, Sydney, Australia

ABSTRACT

Verbal Fluency (VF) task total scores are widely used in language attrition studies, but they do not provide insight into the processes underlying optimal performance. We analyse the efficacy of clustering (subcategories within a category) and switching (shifting between these subcategories) strategies in phonemic (PVF) and semantic (SVF) tasks. First, we focus on L1 Finnish and L2 English performance among attriters ($N = 38$). Our analyses suggest similar processes underlying performance in both languages. These processes seem to remain unaffected by immersion time in the L2 environment (LoR) and frequency of L1 use, highlighting the importance of including L2 data alongside comparisons to L1 monolingual populations to account for a broad bilingual effect in language attrition studies.

Second, we compare attriters' and monolinguals' ($N = 50$) performance in L1. Our findings suggest that attriters rely on clustering in PVF more systematically than monolinguals, and they struggle to initiate a search for a new subcategory or return to a previous category (switching) after depleting a cluster in PVF and SVF tasks. Thus, our analysis demonstrates a difference in processing strategies between the groups that could potentially contribute to similar total scores in PVF and lower total scores in the attriter group in SVF.

ARTICLE HISTORY
Received 27 November 2023
Accepted 7 December 2023

KEYWORDS
Verbal fluency task; clustering; switching; language attrition

Introduction

Language attrition is a complex phenomenon characterised by subtle alterations in the native language of individuals who are immersed in another language, particularly late sequential bi- or multilinguals. Participants in language attrition studies are typically immigrants who share a history of leaving their native country as young adults after achieving mature L1 proficiency and who have spent a prolonged time in an immersive L2 environment with little exposure to their native language (Schmid 2011b; Schmid 2019; Schmid and Jarvis 2014; Schmid and Köpke 2017). The participants in this study align with the description above. In order to underscore the shared language background of the participants, the terms 'attrition' and 'attriter,' which are the conventional labels for this population, will be used even if the presence of language attrition has not specifically been identified within this group (Kasparian and Steinhauer 2017; Schmid and Köpke 2017).

Many language attrition studies focus on L1 performance of attriters, comparing it solely to that of monolingual speakers of the same L1 (Schmid 2019; Schmid and Dusseldorp 2010; Yilmaz and

CONTACT Nana Lehtinen nanleh@utu.fi

Schmid 2018). However, languages constantly interact in a bidirectional process that influences the cognitive and neural processes of both languages of a bilingual individual (Abutalebi and Green 2007; Grosjean 2013; Gurunandan, Garreiras, and Paz-Alonso 2022; Laine and Lehtonen 2018; Linck and Kroll 2019; Treffers-Daller 2019). Thus, performances in both languages and individuals' exposure to them should be taken into account. In this study, we explore performance in L1 and L2 verbal fluency tasks and the impact of extralinguistic variables on both languages by a group of Finnish immigrants who have resided in an English language environment for over 20 years.

Verbal fluency tasks

Verbal fluency (VF) tasks are widely used to investigate lexical processing in diverse bilingual populations, including those experiencing language attrition. These tasks are simple and quick to administer. During a VF task, the participant is prompted to generate as many items as possible following a specific condition within a set time frame. Typically, the tasks are scored based on the total number of words successfully produced, referred to as total scores. The two most typical VF task types are phonemic verbal fluency (PVF, phonemic, or letter cue, e.g. /f/) and semantic verbal fluency (SVF, semantic category cue, e.g. animals).

When selecting the categories for VF tasks, language and, culture-specific factors should be considered (Abwender et al. 2001; Gollan, Sandoval, and Salmon 2011; Olabarrieta-Landa et al. 2017; Roberts and Dorze 1997; Rosselli et al. 2002). In PVF, letter frequency in the target language can impact the number of words generated, and localising letter prompts to high-frequency word-initial letters in the target language has been shown to result in similar norms across languages (Mardani et al. 2020; Oberg and Ramirez 2006; Schmid 2011a; Tombaugh, Kozak, and Rees 1999). In English, a widely used letter compilation is FAS, selected initially due to the high yield of items for these two consonants and one vowel (Borkowski et al. 1967; Ross 2003; Strauss et al. 2006).

In SVF, various demographic factors and cultural settings can influence semantic memory organisation, semantic category size, and content (Abwender et al. 2001; Olabarrieta-Landa et al. 2017; Roberts and Le Dorze 1997; Rosselli et al. 2002; Strauss et al. 2006; Troyer 2000). Selecting a culturally and linguistically relatively neutral category, such as 'animals' (Pekkala et al. 2009), can stabilise these effects between groups.

All VF tasks require rapid lexical retrieval and inhibition of unsuitable candidates, thus engaging verbal knowledge and executive control skills. PVF tasks rely more on strategic cognitive organisation and maintenance of effort, requiring higher cognitive skills, while SVF tasks rely more on semantic categorisation and hierarchical mental lexicon (Luo, Luk, and Bialystok 2010; Patra, Bose, and Marinis 2020; Schmid 2011a; Strauss et al. 2006). VF task total scores can differentiate performance among various populations and reflect verbal and executive processes underlying task performance. In bilingual research, total scores have been used to infer vocabulary size and access, language dominance patterns, and executive functions during task performance (Luo, Luk, and Bialystock 2010; Marsh et al. 2019; Patra, Bose, and Marinis 2020; Roberts and Le Dorze 1997; Rosselli et al. 2002; Schmid and Köpke 2009).

In PVF tasks, mono- and bilingual groups have shown either similar total scores (Rosselli et al. 2000; Rosselli et al. 2002; Soltani et al. 2021) or bilinguals have outperformed monolinguals (Marsh et al. 2019). Higher scores in the bilingual groups have been associated with enhanced executive functions, potentially stemming from inhibiting language interference while switching between languages (Ljungberg et al. 2013; Luo, Luk, and Bialystock 2010; Marsh et al. 2019; Patra, Bose, and Marinis 2020; Sandoval et al. 2010). Studies that focus on language attrition populations rarely apply PVF tasks (Jarvis 2019; Schmid and Köpke 2009). Some studies suggest a trend of language attriters generating a smaller number of words compared to monolinguals in PVF tasks (Lazaridou-Chatzigoga and Karatsareas 2022; Opitz 2011) contradicting general bilingual findings.

In SVF, bilinguals, including language attriters, systematically generate lower total scores than monolinguals. This bilingual disadvantage has been linked to language interference, weaker lexical connections, or smaller vocabularies (e.g. Badstübner 2011; Dostert 2009; Gollan, Montoya, and Werner 2002; Opitz 2011; Rosselli et al. 2000; Sandoval et al. 2010; Schmid and Dusseldorp 2010; Schmid and Keijzer 2009; Schmid and Köpke 2009), but VF tasks have been shown to have little effectiveness in distinguishing individuals as attriters or monolinguals (Schmid and Jarvis 2014).

VF task total scores reflect, but do not allow insight into the processes that support or hinder optimal outcomes. Additional analysis methods focusing on temporal parameters, errors, and clustering and switching strategies, have been proposed to further investigate VF data (Lehtinen, Luotonen, and Kautto 2021; Thiele, Quinting, and Stenneken 2016). Following these suggestions, Lehtinen, Kautto and Renvall (2023) analysed VF performance for total scores, temporal parameters (15 second segments), and errors in three PVF tasks (letters: L1 /k/,/a/, /p/; L2 /f/, /a/, /s/) and one SVF category ('animals') in a group of Finnish-English language attriters and Finnish monolinguals. The current study further extends the analysis of this VF task performance via clustering and switching analysis. In the following, we provide a brief overview of the previous literature on clustering and switching strategies in the context of bilingual verbal fluency tasks before describing the current study in detail.

Clustering and switching

In VF task analysis clustering refers to the ability to form sub-clusters in a category (e.g. in the category 'animals': farm animals – pets – wild animals), and switching is the ability to transition between these clusters (e.g. shifting from farm animals to pets to wild animals). Productive performance in a VF task involves both clustering and switching strategies, typically measured respectively as the mean cluster size and the number of switches. Various methods can be employed in determining clusters and switches, and the technique used to compute clusters also influences the number of transitions between those clusters. Thus, direct comparisons across studies employing different methods should be made with careful consideration (Lehtinen, Kautto, and Luotonen 2021; Strauss et al. 2006; Thiele, Quinting, and Stenneken 2016; Troyer 2000; Troyer, Moscovitch, and Winocur 1997).

Task congruent clusters in PVF, namely phonemic clusters, are based on phonemic characters, such as the same onset-nucleus sequence as in 'simple, simile, sieve'. Generating phonemic clusters requires a cognitively effortful non-routine search based on phonemic attributes without the help of semantic categorisation (Luo, Luk, and Bialystock 2010; Strauss et al. 2006; Troyer 2000; Troyer, Moscovitch, and Winocur 1997). In SVF, generating task-congruent clusters (semantic clusters) involves verbal semantic memory and semantic categorisation, resembling a relatively spontaneous everyday systematic semantic search, such as generating a shopping list. The transition between the clusters (i.e. switching), involves higher executive functions, including cognitive flexibility (Thiele, Quintin, and Stenneken 2016; Troyer, Moscovitch, and Winocur 1997; Troyer 2000). Consequently, Troyer, Moscovitch, and Winocur (1997) showed that in a monolingual population, switching was more crucial for optimal fluency than task-congruent clustering in PVF, while in SVF, both clustering and switching strategies were more equally important in reaching a high total score.

In addition to task congruent clusters, task-discrepant clustering (semantic clusters in PVF and vice versa) represents an additional effortful and often intentional strategy. It involves higher cognitive processes than task-congruent clustering and can suggest the activation of semantic mechanisms in an effortful phonemic task (Abwender et al. 2001; Sung et al. 2013).

Within the field of bilingualism, clustering and switching strategies have been analysed to study executive processes (Mardani et al. 2020; Marsh et al. 2019; Patra, Bose, and Marinis 2020) and cross-linguistic fluency strategies (Roberts and Le Dorze 1997; Rosselli et al. 2002). It has also been suggested that bilinguals, who regularly switch between languages in everyday life, may demonstrate superior cognitive flexibility through more efficient switching strategies than monolinguals (Gollan, Montoya, and Werner 2002).

However, many studies include only one of the participants' languages, and the various methods used to calculate clusters and switches can complicate the interpretation of the findings. For example, Patra, Bose, and Marinis (2020) studied the L2 performance of Bengali-English bilingual immigrants (time spent in L2 environment $M = 7.48$ years) and showed that bilinguals generated larger clusters than monolinguals in PVF, resulting in a higher total score, and groups performed similarly in SVF. They suggested that the ability to maintain a more demanding strategy of PVF clustering better than monolinguals could signal superior executive performance in the bilingual group. In contrast, Mardani et al. (2020) reported a higher total score but no significant difference in L1 PVF cluster size between monolinguals and self-reported Farsi-Balochi bilinguals with no sudden change in their language environment. Thus, bilinguals relied more on switching to reach a higher total score than monolinguals. In SVF, bilinguals generated smaller clusters than monolinguals in L1 with no significant difference in total scores, interpreted as a disadvantage in verbal-semantic memory performance in the bilingual group.

As for studies investigating language attrition populations per se, we are not aware of a robust VF task clustering and switching analysis. Ammerlaan (1996) notes that attriters rely heavily on effortful strategies in VF but does not include an analysis of these strategies in his study.

When determining clusters and switches in bilingual populations, it is important to consider varied language backgrounds. Roberts and Le Dorze (1997) studied L1 and L2 in a French-English bilingual group and reported larger semantic clusters in L1 than in L2 in the category 'animals' but not 'foods'. They suggested that as the names of animals were learned during childhood when the participants were exposed primarily to L1, there may be a stronger semantic base in L1 for 'animals' than 'foods', since food-related words are typically learned later in life when the participants had more exposure to L2. Rosselli et al. (2002) also observed that differences in the semantic clustering of two monolingual groups were less evident in English-Spanish bilinguals (specifically that English monolinguals named more wild animals and Spanish monolinguals more birds and insects). The authors interpreted this as indicating different semantic category structures between bilingual and monolingual groups.

In clustering analysis, language-specific factors can be taken into consideration by analysing naturally occurring clusters rather than using predetermined categories, especially in SVF (Abwender et al. 2001; Gollan, Sandoval, and Salmon 2011; Lehtinen, Luotonen, and Kautto 2021; Olabarrieta-Landa et al. 2017; Roberts and Dorze 1997; Rosselli et al. 2002). Naturally occurring semantic subcategories can include taxonomic subcategories (e.g. wolf, dog), environmental (e.g. gazelle, lion), geographical (e.g. wombat, kangaroo), or visual semantics (e.g. snake, eel) in the semantic category. In the phonemic category, these can include, e.g. two initial phonemes (e.g. fossil, foster) or rhyming words (e.g. fight, flight). For a more detailed discussion on naturally occurring clusters, we refer the reader to the VF task administration and analysis guide in Lehtinen, Luotonen, and Kautto (2021).

Extralinguistic variables

Sociolinguistic parameters, such as language exposure and use, can result in shifts in individual language dominance patterns over time affecting language performance (Treffers-Daller 2019; Yilmaz 2019). As language attrition populations share a history of prolonged immersion in an L2 environment with minimal exposure to their native L1 over an extended time (Schmid 2011b; Schmid 2019; Schmid and Jarvis 2014; Schmid and Köpke 2017), in this study we focus on two variables that reflect the participants' shared language history and may have a direct effect on language attrition: length of residence in the L2 environment (LoR) and frequency of L1 use during that time.

Research shows that in general, LoR typically affects L1 language task performance during the initial years of residence, while L2 is simultaneously rapidly acquired, with the effect typically levelling out after the first decade (Linck and Kroll 2019; Opitz 2011; Schmid 2011b; Schmid 2019). The impact of LoR may also become significant when combined with very little L1 use (de Bot and Clyne 1994).

The frequency of L1 use as a single measure has not been found to systematically impact L1 performance (for an overview, see Schmid 2019). However, L1 use in professional settings has been shown to positively impact lexical retrieval and lexical diversity in L1 (Schmid and Dusseldorp

2010; Yilmaz and Schmid 2012). In addition, frequent L1 use in informal settings with peers has been shown to contribute to increased variability in the phonemic domain, potentially suggesting contact-induced language change in the community rather than at the individual level (De Leeuw, Schmid, and Mennen 2010).

In previous studies on language attrition, VF total scores have been largely unrelated to the time spent in an L2 environment (LoR) or frequency of L1 use (Schmid 2011a; Schmid and Jarvis 2014; Schmid and Köpke 2009). In this study, we investigate the impact of LoR, frequency of L1 use, and their interactions on the efficacy of clustering and switching in PVF and SVF among a group of language attriters to better understand the impact extralinguistic variables have on the processes underlying VF task performance.

Research objectives, questions, and hypotheses

The present study examines the effectiveness of clustering and switching strategies in phonemic (PVF) and semantic (SVF) verbal fluency tasks among Finnish-English bilinguals who have experienced prolonged immersion in an English-speaking environment, thus considered attriters. First, we examine VF task performance in their first (L1, Finnish) and second language (L2, English) and assess the influence of the length of residence (LoR) and frequency of L1 use on the use of clustering and switching. Second, we compare L1 task performance between attriters and L1 monolinguals to investigate differences in the efficacy of clustering and switching between the groups. Our research questions and hypotheses are as follows:

Research question 1: L1 and L2 within the attriter group

(a) Do task-congruent mean cluster size and the number of switches predict total score similarly for the attriters in L1 and in L2 in PVF and SVF tasks?
(b) Does LoR, the frequency of L1 use, and their interactions impact the efficacy of clustering and switching strategies within the group?

Hypothesis: We expected attriters to demonstrate a robust semantic network in L1 by relying more efficiently on clustering than switching in L1 compared to L2. We anticipated frequent L1 use to support switching in both languages and task types, potentially reflecting cognitive flexibility stemming from more frequent L1 use in an L2-dominant environment.

Research question 2: L1 in the attriter and monolingual groups

Do task-congruent mean cluster size and the number of switches predict total score similarly in the attriter and monolingual groups for L1 PVF and SVF tasks?

Hypothesis: We expected attriters to demonstrate superior executive flexibility by relying more on switching than monolinguals in PVF. In SVF, we anticipated attriters to display a weaker semantic foundation and greater cognitive flexibility than monolinguals by relying less on clustering and more on switching strategies.

Materials and methods

Participants

Two groups of healthy, neurotypical adults participated in this study: L1 attriters ($n = 38$, L1 Finnish, L2 English) and Finnish monolinguals ($N = 50$). Attriters were first-generation immigrants living in Northern California who had immigrated from Finland between 1948 and

1998 [age $M = 60.89$, $SD = 8.42$, range 45–79; education: no academic degree ($n = 16$) and academic degree ($n = 22$), gender: female ($n = 29$) and male ($n = 9$)]. Attrition-specific data is referenced in Table 1.

The monolingual participants identified themselves as monolinguals who had always lived in Finland and used only Finnish in their everyday life [age $M = 62.58$, $SD = 7.59$, range 49–79, education: no academic degree ($n = 27$) and academic degree ($n = 23$), gender: female ($n = 35$) and male ($n = 15$)]. The groups were matched for age, education (no academic degree/academic degree), and gender with no significant differences detected for age (attriters $M = 60.90$, monolinguals $M = 62.60$) $z = -0.97$, $p = .332$, education ($\chi(1) = 0.49$, $p = .27$), or gender ($\chi(1) = 0.43$, $p = .51$) between the groups.

Administration and analyses of data

Task administration, scoring, and analyses were conducted following guidelines in Lehtinen, Luotonen, and Kautto (2021). For clarity, we briefly describe the methods below.

During administration, participants were asked to generate as many words as possible in a 60-second time frame following the given criteria. The only exclusion was proper names and instructions included the phrase 'individual words' to discourage participants from generating inflections of the same word in Finnish [e.g. *kirja* (book), *kirjassani* (in my book)].

L1 language-appropriate phonemic prompts /k/, /a/, /p/ were selected following a high-frequency dictionary approach for two consonants and one vowel to correspond to L2 letter selection (Lehtinen, Luotonen, and Kautto 2021; Mardani et al. 2020; Oberg and Ramirez 2006; Schmid 2011a). For L2 PVF, letters /f/, /a/, /s/ were selected based on their frequent use in the literature (Borkowski et al. 1967; Ross 2003; Strauss et al. 2006). For the SVF, we selected the culturally and linguistically relatively neutral category 'animals' (Pekkala et al. 2009).

We calculated the mean cluster size for naturally occurring semantic clusters in SVF and the mean cluster size for naturally occurring phonemic clusters in PVF. The minimum cluster size was two words. The number of switches was based on the task- congruent clusters. Our preliminary analysis showed that task-discrepant clusters (semantic clusters in PVF and vice versa) occurred only rarely in the data. Due to the small sample size, analysis for task-discrepant clusters was omitted, but the data is included in the data repository of this study.

Extralinguistic factors were extracted from a sociolinguistic questionnaire and included self-reported length of residence in an L2 environment in years (LoR) and overall frequency of L1 use during that time [How often do you speak Finnish? 5-point Likert scale: rarely ($n = 1$), few times a year ($n = 1$), monthly ($n = 7$), weekly ($n = 22$), daily ($n = 7$)].

R software (R Core Team 2019) with packages dplyr (Wickham et al. 2019), tidyr (Wickham 2020), lme4 (Bates et al. 2015), and lmerTest (Kuznetsova, Brockhoff, and Christensen 2017) were used in data clean-up and analyses. Packages sjPlot (Lüdecke 2018), jtools (Long 2020), ggeffects (Lüdecke 2018), and ggplot2 (Wickham 2016) were used in tables and figures. Scripts and data used for the analyses are available at https://osf.io/95q3j/.

Table 1. Age at emigration, length of residence in L2 environment, frequency of L1 use, and self-reported language proficiency in L1 and L2 in the attriter group ($N = 38$).

	M	Range	*SD*
Age at emigration	26.68	9–48	7.38
Length of residence in L2 environment	34.24	20–50	10.83
	Very good	*Good*	*Good enough to get by*
Self-report on L1 before emigration	23	15	0
Self-report on L1 at the time of interview	6	26	6
Self-report on L2 at the time of interview	21	16	1

Statistical analyses

For comparison between the efficacy of clustering and switching strategies in L1 and L2 within the attriter group, we used linear mixed-effects models to predict total scores as a function of task-congruent cluster size, the number of switches and language (L1/L2) separately for PVF and SVF with LoR and L1 use as predictors. Cluster size and the number of switches were centred to the sample mean. Participant ID was used as a random factor to account for individual variation in performance. Comparisons of Bayesian Information Criterion (BIC) values in models with and without LoR as a predictor suggested the models without LoR to be the most parsimonious fits to the data.

To compare the L1 performance in the attriter and monolingual groups, we modelled total scores as a function of task-congruent cluster size, the number of switches, and the participant group separately for PVF and SVF. As our PVF task data included three observations for each participant (/k/, /a/, and /p/), participant ID was used as a random factor in that model.

Results

Descriptive statistics for total scores in L1 and L2 in the attriter group and L1 in the monolingual group are shown in Table 2.

Research question 1: L1 and L2 within the attriter group

The model summaries, including estimates, *t*-values, and *p*-values for PVF are presented in Table 3, and for SVF in Table 4. For both tasks, we found significant main effects for mean cluster size and the number of switches with higher total scores linked to efficient use of clustering and switching strategies. We found no significant interactions for task language or frequency of L1 use. The relationship between clustering and switching was consistent in L1 and L2 for both task types, and the frequency of L1 use did not modify the use of clustering and switching within the attriter group.

Research question 2: L1 in the attriter and monolingual groups

The model summary for switching and clustering in the PVF task is presented in Table 5. A significant interaction between group and cluster size was observed, but no significant interaction was detected for the number of switches and group. The stronger relationship between cluster size and the total score in the attriter group compared to the monolingual group is shown in Figure 1.

The Model summary for switching and clustering in the SVF is presented in Table 6. We discovered a significant interaction between group, cluster size, and the number of switches. Figure 2 shows the relationship between cluster size and the total score becoming weaker as the number of switches increases in the monolingual group, while in the attriter group, the connection between cluster size and the total score remains consistent regardless of the number of switches.

Table 2. Descriptive data for total scores in L1 and L2 in the attriter Group and L1 in the monolingual group.

	L1				L2			
	k	a	p	animals	f	a	s	animals
Attriters								
Mean	18.92	14.90	16.45	21.90	14.05	10.60	15.32	20.60
SD	4.94	5.03	4.54	4.31	4.23	5.13	5.21	4.85
Range	11–33	6–24	7–27	13–33	7–27	1–20	6–28	9–32
Monolinguals								
Mean	19.38	14.68	17.44	25.96				
SD	6.12	5.65	5.50	5.90				
Range	5–35	5–31	4–30	12–43				

Note. L1 attriters $n = 38$; L2 attriters $n = 37$; L1 monolinguals $n = 50$. Overall VF performance and performance changes during the tasks were previously reported in Lehtinen, Kautto, and Renvall (forthcoming) for attriters and in Lehtinen, Luotonen, and Kautto (2021: 1–13) for the monolingual group.

Table 3. Phonemic Verbal Fluency task total scores as a function of task-congruent cluster size, the number of switches, and language (L1/L2) in the attriter group.

	Total score			
Predictors	Estimates	95% CI	*t*	*p*
(Intercept)	14.66	13.95–15.38	40.29	<0.001
Number of switches	3.92	3.28–4.56	12.04	<0.001
Mean cluster size	2.17	1.40–2.94	5.56	<0.001
Task language	0.72	−0.08–1.51	1.78	0.08
Frequency of L1 use	0.57	−0.23–1.36	1.41	0.16
Number of switches* Mean cluster size	0.02	−0.62–0.67	0.06	0.95
Number of switches* Task language	−0.56	−1.39–0.28	−1.31	0.19
Mean cluster size* Task language	0.51	−0.49–1.50	1.01	0.32
Number of switches* Frequency of L1 use	0.16	−0.59–0.91	0.43	0.67
Mean cluster size* Frequency of L1 use	0.43	−0.25–1.11	1.24	0.22
Task language* Frequency of L1 use	0.33	−0.54–1.2	0.75	0.46
Number of switches* Mean cluster size* Task language	0.09	−0.80–0.99	0.21	0.84
Number of switches* Mean cluster size* Frequency of L1 use	0.01	−0.65–0.68	0.03	0.97
Number of switches* Task language* Frequency of L1 use	0.03	−0.89–0.96	0.07	0.94
Mean cluster size* Task language* Frequency of L1 use	0.49	−0.43–1.40	1.05	0.30
Number of switches* Mean cluster size* Task language* Frequency of L1 use	0.29	−0.63–1.22	0.63	0.53
Random effects				
σ^2	6.77			
τ_{00} ID	1.81			
ICC	0.21			
N_{ID}	38			
Observations	*225*			
Marginal R^2 / Conditional R^2	*0.697 / 0.761*			

Discussion

Verbal fluency tasks are one of the most utilised tasks in language attrition studies (Schmid 2019). However, their analyses are typically limited to total scores, offering little insight into the processes underlying task performance. In the current study, we extended the analyses of VF task performance

Table 4. Semantic Verbal Fluency task total scores as a function of task-congruent cluster size, the number of switches, and language (L1/L2) in the attriter group.

	Total score			
Predictors	Estimates	95% CI	*t*	*p*
(Intercept)	21.08	20.09–22.07	42.64	<0.001
Number of switches	3.56	2.49–4.64	6.64	<0.001
Mean cluster size	2.43	1.44–3.43	4.90	<0.001
Task language	0.87	−0.42–2.16	1.36	0.18
Frequency of L1 use	1.09	−0.06–2.23	1.90	0.06
Number of switches* Mean cluster size	0.33	−0.88–1.53	0.55	0.59
Number of switches* Task language	−0.64	−2.10–0.82	−0.87	0.39
Mean cluster size* Task language	−0.89	−2.43–1.66	−1.15	0.26
Number of switches* Frequency of L1 use	−0.13	−1.34–1.08	−0.22	0.83
Mean cluster size* Frequency of L1 use	−0.57	−1.85–0.72	−1.88	0.38
Task language* Frequency of L1 use	0.55	−1.08–2.18	0.68	0.50
Number of switches* Mean cluster size* Task language	0.65	−0.92–2.21	0.83	0.41
Number of switches* Mean cluster size* Frequency of L1 use	−0.37	−1.84–1.10	−0.51	0.62
Number of switches* Task language* Frequency of L1 use	0.27	−1.41–1.96	0.33	0.75
Mean cluster size* Task language* Frequency of L1 use	1.18	−0.90–3.27	1.14	0.26
Number of switches* Mean cluster size* Task language* Frequency of L1 use	1.88	−0.44–4.21	1.62	0.11
Random effects				
σ^2	5.71			
τ_{00} ID	1.25			
ICC	0.18			
N_{ID}	38			
Observations	75			
Marginal R^2 / Conditional R^2	0.689 / 0.745			

Table 5. Phonemic Verbal Fluency task total scores as a function of task-congruent cluster size, the number of switches, and participant group in the attriter and monolingual groups.

	Total score			
Predictors	Estimates	95% CI	*t*	*p*
(Intercept)	16.93	16.06–17.80	38.39	<0.001
Group [monolingual]	−0.01	−1.16–1.13	−0.02	0.98
Number of switches	3.93	3.18–4.67	10.38	<0.001
Mean cluster size	3.51	2.42–4.60	6.34	<0.001
Group [monolingual]* Number of switches	−0.26	−1.16 - 0.63	−0.58	0.56
Group [monolingual]* Mean cluster size	−1.53	−2.74 – −0.33	−2.51	0.013
Number of Switches* Mean cluster size	0.55	−0.26–1.37	1.33	0.18
Group [monolingual]* Number of switches* Mean cluster size	−0.72	−1.61–0.17	−1.59	0.11
Random effects				
σ^2	5.88			
τ_{00} ID	5.17			
ICC	0.47			
N_{ID}	88			
Observations	264			
Marginal R^2 / Conditional R^2	0.603 / 0.789			

Phonemic Verbal Fluency Task Total Scores as a Function of Task-Congruent Cluster Size, the Number of Switches, and Participant Group in the Attriter and Monolingual Group.

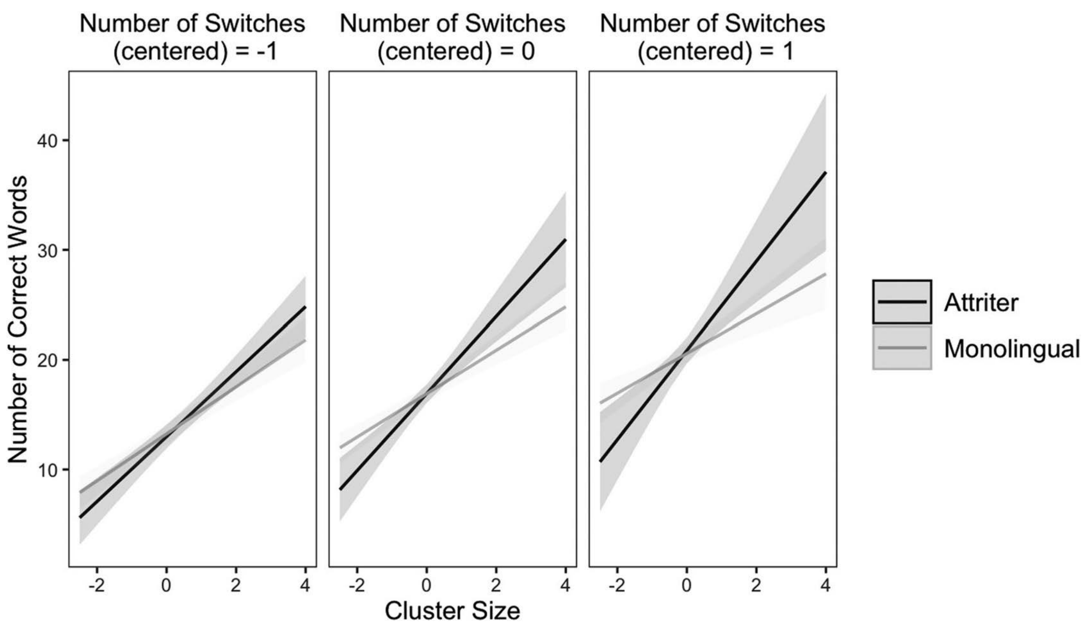


Figure 1. Total scores as a function of task-congruent cluster size (centred), the number of switches (centred), and participant group in the Phonological Verbal Fluency task.

to the efficacy of clustering and switching strategies in three phonemic (PVF) and one semantic (SVF) verbal fluency tasks. Our first objective was to explore processes that support optimal performance in L1 and L2 in a group of Finnish-English language attriters and to investigate the influence LoR and frequent L1 use has on these processes. We then set out to determine if the processes of optimal VF task performance differ between attriters and L1 monolinguals.

We expected attriters to rely more efficiently on clustering in L1 than L2 as a marker of a more robust semantic network in L1. Contrary to our expectations, attriters utilised clustering and switching strategies equally efficiently to reach a high score across languages and tasks. Thus, the performance in L1 and L2 SVF and PVF tasks mirrors typical, monolingual SVF performance where both

Table 6. Semantic Verbal Fluency task total scores as a function of task-congruent cluster size, the number of switches, and participant group in the attriter and monolingual groups.

	Total score			
Predictors	Estimates	95% CI	*t*	*p*
(Intercept)	23.44	22.21–24.66	38.17	<0.001
Group [monolingual]	1.39	−0.18–2.95	1.76	0.08
Number of switches	3.59	2.31–4.88	5.58	<0.001
Mean cluster size	3.08	1.64–4.52	4.27	<0.001
Group [monolingual]* Number of switches	0.79	−0.80 - 2.39	0.99	0.33
Group [monolingual]* Mean cluster size	−1.36	−3.05–0.33	−1.61	0.11
Number of switches* Mean cluster size	0.72	−0.69–2.13	1.02	0.31
Group [monolingual]* Number of switches* Mean cluster size	−2.12	−3.85 – −0.40	−2.45	0.017
Observations	264			
Marginal R^2 / Conditional R^2	0.603 / 0.789			

Note: Descriptive data for total scores in both languages and groups is shown in Table 2.

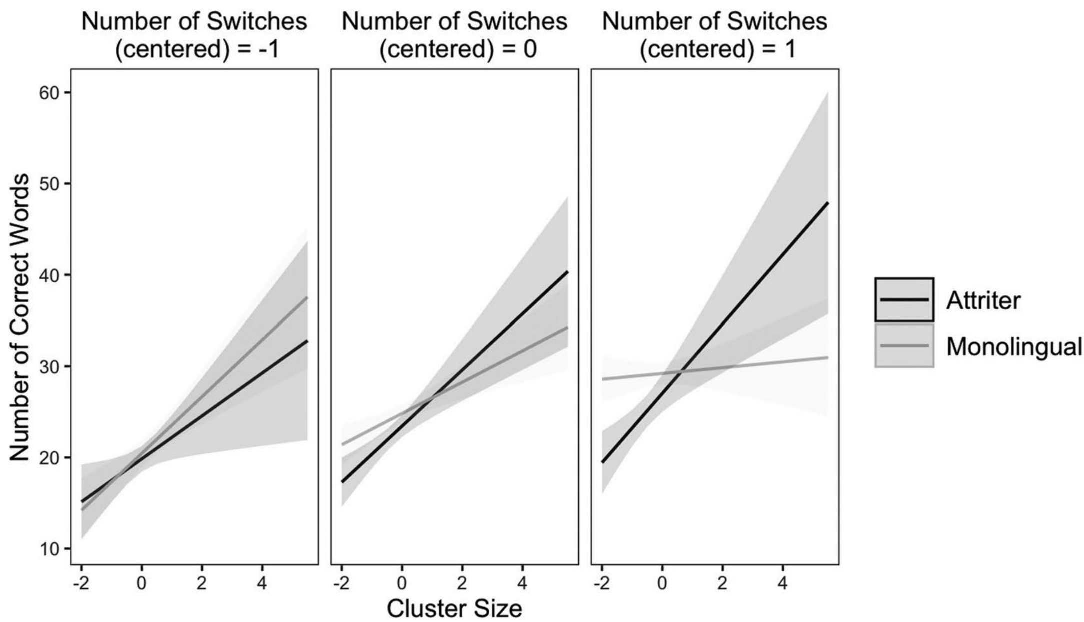


Figure 2. Total scores as a function of task-congruent cluster size (centred), the number of switches (centred), and participant group in the Semantic Verbal Fluency task.

strategies contribute equally to a high score as reported by Troyer, Moscovitch, and Winocur (1997). This finding is discussed below in relation to the performance of the monolingual group. Our findings of similar performance in L1 and L2 point to the direction of a general bilingual effect in VF task performance across languages rather than L1 language attrition among attriters. Consequently, our results highlight the importance of investigating performance in both languages in language attrition studies to account for the broad impact bidirectional cognitive and neural processes of bilingualism can have on the linguistic performance of a bilingual individual as a whole (Abutalebi and Green 2007; Gurunandan, Carreiras, and Paz-Alonso 2022; Laine and Lehtonen 2018; Schmid and Köpke 2017).

Our first research question also addressed the impact of LoR and frequency of L1 use on strategies that underlie task performance. These extralinguistic variables did not appear to contribute to the efficacy of either strategy. In future research, analysing task-discrepant clusters and extralinguistic factors' influence on their occurrence in L1 and L2 can offer insight into the impact of extralinguistic factors on processing strategies between languages.

To answer our second research question, we investigated differences in the efficacy of clustering and switching strategies in L1 task performance between attriters and L1 monolinguals. Based on the literature, we anticipated that attriters would demonstrate a cognitive advantage by relying more on switching than monolinguals in PVF. However, we detected no significant differences in the effectiveness of switching strategies between the groups. On the contrary, in PVF attriters seemed to be able to maintain the effortful cognitive strategy of generating task-congruent phonemic clusters more efficiently than monolinguals. Our results align with Patra, Bose, and Marinis (2020), who interpreted their finding of larger task-congruent clusters in the bilingual group as superior executive functions compared to a monolingual group. In the current study, we similarly infer that rather than relying on switching (i.e. returning to a previous subcategory or initiating a new subcategory when a subcategory has been depleted), attriters generated larger clusters to reach a similar total score as monolinguals in PVF. In line with this finding, a comparison between L1 and L2 performance among the attriters showed that attriters rely more equally on clustering and switching in PVF than anticipated in both languages.

The phenomenon of attriters relying on clustering for optimal performance was also apparent in SVF, where the connection between cluster size and the total score remained consistent in the attriter group regardless of the number of switches. In contrast, monolinguals used clustering and switching strategies more equally to achieve higher scores, in line with Troyer, Moscovitch, and Winocur (1997). Thus, our findings underscore that while monolinguals were able to support their performance by switching between categories and clustering, support from switching between the categories was less efficient for optimal performance in the attriters group and these tendencies were reflected in L2 performance.

Our results for PVF align partly with Patra, Bose, and Marinis (2020), where bilinguals generated higher total scores in their L2 by relying on larger clusters than monolinguals. While their study did not address language attrition per se, their participants were immigrants who had spent a significant time in an immersive L2 environment ($M = 7.48$ years). In contrast, Mardani et al. (2020) found that bilinguals with no sudden change in their language environment relied more on switching than clustering, generating a higher total score than monolinguals in L1 PVF. They also report that bilinguals generated smaller clusters in L1 SVF, resulting in similar total scores with monolinguals. When interpreting these findings, it is important to note that there are various methods for determining clusters and, consequently, the number of switches between these clusters. Both Patra, Bose, and Marinis (2020) and Mardani et al. (2020) used predetermined cluster categories in their analysis, but scarce details were given of their specific analysis. We calculated the size of naturally occurring task- congruent clusters to include all unique clustering strategies and natural switches generated by participants without the constraints of predetermined semantic categories (Lehtinen, Luotonen, and Kautto 2021; Roberts and Dorze 1997; Rosselli et al. 2002).

Thus, comparison across these studies requires careful consideration. In addition, the relatively small study samples and different language pairs call for further consideration when generalising our results. However, we remain confident that our analysis based on well-defined guidelines on analysing naturally occurring clusters and switches offers a solid foundation and direction for future studies involving larger data sets, diverse language pairs, and various language contact situations. Future comparative studies between attriters and bilinguals with no change in their language environment could also determine the role of extralinguistic circumstances in the use of clustering and switching in VF tasks across bilingual populations.

We set out to investigate the processes that support or hinder optimal outcomes within a Finnish-English language attrition population L1 and L2 VF task performance and between attritters and L1 monolinguals. Our results demonstrate that an extended VF task analysis has the potential to yield insight into the subtle language processes underlying VF task performance and provide a solid starting point for future studies to further investigate the similarities in language attrition populations in contrast to monolinguals and bilinguals with varied backgrounds.

Acknowledgments

We thank Ida Luotonen for her contributions to data analysis and overall support, Professor Marja-Liisa Helasvuo for guidance and insightful discussions on Finnish language characteristics, and Professor Monika Schmid for encouragement during the formulation of this study. We also want to thank Lotta Weckström for her help in recruiting participants for the study in the US. Finally, we express our gratitude to the participants for the generous donation of their time and enthusiasm in participating in this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Finlandia Foundation National; The Alfred Kordelin Foundation; The Migration Institute of Finland; Turun Yliopistosäätiö; Turun yliopiston tutkijakoulu.

Data availability statement

The data supporting this study's findings are available in [Verbal Fluency Switching and Clustering in Language Attrition] at [https://osf.io/95q3j].ote.

ORCID

Nana Lehtinen http://orcid.org/0000-0002-3606-1302
Anna Kautto http://orcid.org/0000-0003-4885-9167
Kati Renvall http://orcid.org/0000-0002-3329-4602

References

Abutalebi, Jubin, and David Green. 2007. Bilingual language production: the neurocognition of language representation and control. *Journal of Neurolinguistics* 20, no. 3: 242–75. https://doi.org/10.1016/j.jneuroling.2006.10.003.

Abwender, David A., Jeffrey G. Swan, John T. Bowerman, and Sean W. Connolly. 2001. Qualitative analysis of verbal fluency output: review and comparison of several scoring methods. *Assessment* 8, no. 3: 323–38. https://doi.org/10.1177/107319110100800308.

Ammerlaan, Ton. 1996. 'You get a bit wobbly …' : exploring bilingual lexical retrieval processes in the context of first language attrition. PhD diss., Radboud University, Nijmegen.

Badstübner, Tina. 2011. L1 attrition: German immigrants in the U.S. PhD diss., The University of Arizona.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Ting linear mixed-effects models using Lme4. *Journal of Statistical Software* 67, no. 1. https://doi.org/10.18637/jss.v067.i01.

Borkowski, John G., Arthur L. Benton, and Otfried Spreen. 1967. Word fluency and brain damage. *Neuropsychologia* 5, no. 2: 135–40. https://doi.org/10.1016/0028-3932(67)90015-2.

Bot, Kees de, and Michael Clyne. 1994. A 16-year longitudinal study of language attrition in Dutch immigrants in Australia. *Journal of Multilingual and Multicultural Development* 15, no. 1: 17–28. https://doi.org/10.1080/01434632.1994.9994554.

De Leeuw, Esther, Monika S. Schmid, and Ineke Mennen. 2010. The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition* 13, no. 1: 33–40. https://doi.org/10.1017/S1366728909990289.

Dostert, Susan Claire. 2009. Multilingualism, L1 attrition and the concept of 'native speaker'. PhD Diss., die Philosophische Fakultät: Heinrich Heine Universitat Düsseldorf.

Gollan, Tamar H., Rosa I. Montoya, and Grace A. Werner. 2002. Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology* 16, no. 4: 562–76. https://doi.org/10.1037/0894-4105.16.4.562.

Gollan, Tamar H., Tiffany Sandoval, and David P. Salmon. 2011. Cross- language intrusion errors in aging bilinguals reveal the link between executive control and language selection. *Psychological Science* 22, no. 9: 1155–64. https://doi.org/10.1177/0956797611417002.

Grosjean, François. 2013. Bilingualism: a short introduction. In *The Psycholinguistics of Bilingualism*, eds. François Grosjean, and Ping Li, 5:5–25. 2. Hoboken: John Wiley & Sons.

Gurunandan, Kshipra, Manuel Carreiras, and Pedro M Paz-Alonso. 2022. Verbal production dynamics and plasticity: functional contributions of language and executive control systems. *Cerebral Cortex* 33, no. 3: 740–53. https://doi.org/10.1093/cercor/bhac097.

Jarvis, Scott. 2019. Lexical attrition. In *The Oxford Handbook of Language Attrition*, eds. Monika S. Schmid, and Barbara Köpke, 240–50. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.20.

Kasparian, Kristina, and Karsten Steinhauer. 2017. On missed opportunities and convenient 'truths'. *Linguistic Approaches to Bilingualism* 7, no. 6: 709–14. https://doi.org/10.1075/lab.00010.kas.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. Lmertest package: tests in linear mixed effects models. *Journal of Statistical Software* 82, no. 13: 1–26. https://doi.org/10.18637/jss.v082.i13.

Laine, Matti, and Minna Lehtonen. 2018. Cognitive consequences of bilingualism: where to go from here? *Language, Cognition and Neuroscience* 33, no. 9: 1205–12. https://doi.org/10.1080/23273798.2018.1462498.

Lazaridou-Chatzigoga, Dimitra, and Petros Karatsareas. 2022. Language attrition and lived experiences of attrition among Greek speakers in London. *Languages* 7, no. 4: 307. https://doi.org/10.3390/languages7040307.

Lehtinen, Nana, Anna Kautto, and Kati Renvall. 2023. Frequent native language use supports phonemic and semantic verbal fluency in L1 and L2: an extended analysis of verbal fluency task performance in an L1 language attrition population. *International Journal of Bilingualism* 0, no. 0. https://doi.org/10.1177/13670069231193727.

Lehtinen, Nana, Ida Luotonen, and Anna Kautto. 2021. Systematic administration and analysis of verbal fluency tasks: preliminary evidence for reliable exploration of processes underlying task performance. *Applied Neuropsychology: Adult* September: 1–13. https://doi.org/10.1080/23279095.2021.1973471.

Linck, Jared A., and Judith F. Kroll. 2019. Memory retrieval and language attrition: language loss or manifestations of a dynamic system? In *The Oxford Handbook of Language Attrition*, eds. Monika S. Schmid, and Barbara Köpke, 87–97. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.8.

Ljungberg, Jessica K., Patrik Hansson, Pilar Andrés, Maria Josefsson, and Lars- Göran Nilsson. 2013. A longitudinal study of memory advantages in bilinguals. *PLoS One* 8, no. 9: e73029. https://doi.org/10.1371/journal.pone.0073029.

Long, Jacob A. 2020. Jtools: analysis and presentation of social scientific data. https://cran.r-project.org/package=jtools.

Lüdecke, Daniel. 2018. SjPlot - data visualization for statistics in social science. *Zenodo*, https://doi.org/10.5281/ZENODO.1308157.

Luo, Lin, Gigi Luk, and Ellen Bialystok. 2010. Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition* 114, no. 1: 29–41. https://doi.org/10.1016/j.cognition.2009.08.014.

Mardani, Najme, Mohammad Pourjafari, Mohammad Amin Irandegani, Nassim Ahmadi, and Kowsar Baghban. 2020. The effect of bilingualism on the processing of clustering and switching in verbal fluency tasks. *Journal of Rehabilitation Sciences & Research* 7, no. 3: 114–7. https://doi.org/10.30476/jrsr.2020.86911.1096.

Marsh, John E., Patrik Hansson, Daniel Eriksson Sörman, and Jessica Körning Ljungberg. 2019. Executive processes underpin the bilingual advantage on phonemic fluency: evidence from analyses of switching and clustering. *Frontiers in Psychology* 10, no. June: 1355. https://doi.org/10.3389/fpsyg.2019.01355.

Oberg, Gunilla, and Maura Ramírez. 2006. Cross-linguistic meta-analysis of phonological fluency: normal performance across cultures. *International Journal of Psychology* 41, no. 5: 342–7. https://doi.org/10.1080/00207590500345872.

Olabarrieta-Landa, Laiene, Esther Landa Torre, Juan Carlos López-Mugartza, Ellen Bialystok, and Juan Carlos Arango-Lasprilla. 2017. Verbal fluency tests: developing a new model of administration and scoring for Spanish language. *NeuroRehabilitation* 41, no. 2: 539–65. https://doi.org/10.3233/NRE-162102.

Opitz, Cornelia. 2011. First language attrition and second language acquisition in a second language environment. PhD diss., Centre for Language and Communication Studies: University of Dublin, Trinity College. http://hdl.handle.net/2262/78154.

Patra, Abhijeet, Arpita Bose, and Theodoros Marinis. 2020. Performance difference in verbal fluency in bilingual and monolingual speakers. *Bilingualism: Language and Cognition* 23, no. 1: 204–18. https://doi.org/10.1017/S1366728918001098.

Pekkala, Seija, Mira Goral, JungMoon Hyun, Loraine K. Obler, Timo Erkinjuntti, and Martin L. Albert. 2009. Semantic verbal fluency in two contrasting languages. *Clinical Linguistics & Phonetics* 23, no. 6: 431–45. https://doi.org/10.1080/02699200902839800.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R- project.org/.

Roberts, Patricia M., and Guylaine Le Dorze. 1997. Semantic organization, strategy use, and productivity in bilingual semantic verbal fluency. *Brain and Language* 59, no. 3: 412–49. https://doi.org/10.1006/brln.1997.1753.

Ross, Thomas P. 2003. The reliability of cluster and switch scores for the controlled oral word association test. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists* 18, no. 2: 153–64.

Rosselli, Mónica, Alfredo Ardila, Katia Araujo, Viviana A. Weekes, Virginia Caracciolo, Mabel Padilla, and Feggy Ostrosky-Solí. 2000. Verbal fluency and repetition skills in healthy older Spanish-English bilinguals. *Applied Neuropsychology* 7, no. 1: 17–24. https://doi.org/10.1207/S15324826AN0701_3.

Rosselli, Mónica, Alfredo Ardila, Judy Salvatierra, Martha Marquez, Matos Luis, and Viviana A. Weekes. 2002. A cross-linguistic comparison of verbal fluency tests. *International Journal of Neuroscience* 112, no. 6: 759–76. https://doi.org/10.1080/00207450290025752.

Sandoval, Tiffany C., Tamar H. Gollan, Victor S. Ferreira, and David P. Salmon. 2010. What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition* 13, no. 2: 231–52. https://doi.org/10.1017/S1366728909990514.

Schmid, Monika S. 2011a. Contact x time: external factors and variability in L1 attrition. In *Studies in Bilingualism*, eds. Monika S. Schmid, and Wander Lowie, 43:155–76. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/sibil.43.11sch.

Schmid, Monika S. 2011b. *Language Attrition*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511852046.

Schmid, Monika S. 2019. The impact of frequency of use and length of residence on L1 attrition. In *The Oxford Handbook of Language Attrition*, eds. Monika S. Schmid, and Barbara Köpke, 287–303. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.24.

Schmid, Monika S., and Elise Dusseldorp. 2010. Quantitative analyses in a multivariate study of language attrition: the impact of extralinguistic factors. *Second Language Research* 26, no. 1: 125–60. https://doi.org/10.1177/0267658309337641.

Schmid, Monika S., and Scott Jarvis. 2014. Lexical access and lexical diversity in first language attrition. *Bilingualism: Language and Cognition* 17, no. 4: 729–48. https://doi.org/10.1017/S1366728913000771.

Schmid, Monika S., and Merel Keijzer. 2009. First language attrition and reversion among older migrants. *International Journal of the Sociology of Language* 2009, no. 200. https://doi.org/10.1515/IJSL.2009.046.

Schmid, Monika S., and Barbara Köpke. 2009. L1 attrition and the mental lexicon. In *The Bilingual Mental Lexicon: Interdisciplinary Approaches*, ed. Aneta Pavlenko, 209–38. Bilingual Education & Bilingualism. Bristol, UK; Buffalo, NY: Multilingual Matters.

Schmid, Monika S., and Barbara Köpke. 2017. The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism* 7, no. 6: 637–67. https://doi.org/10.1075/lab.17058.sch.

Soltani, Majid, Negin Moradi, Hosein Rezaei, Masoume Hosseini, and Ehsan Jasemi. 2021. Comparison of verbal fluency in monolingual and bilingual elderly in Iran. *Applied Neuropsychology: Adult* 28, no. 1: 80–7. https://doi.org/10.1080/23279095.2019.1594234.

Strauss, Esther, Elisabeth M. S. Sherman, Otfried Spreen, and Otfried Spreen. 2006. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary, 3rd ed*. Oxford; New York: Oxford University Press.

Sung, Kyongje, Barry Gordon, Sujeong Yang, and David J. Schretlen. 2013. Evidence of semantic clustering in letter-cued word retrieval. *Journal of Clinical and Experimental Neuropsychology* 35, no. 10: 1015–23. https://doi.org/10.1080/13803395.2013.845141.

Thiele, Kristina, Jana Marie Quinting, and Prisca Stenneken. 2016. New ways to analyze word generation performance in brain injury: a systematic review and meta-analysis of additional performance measures. *Journal of Clinical and Experimental Neuropsychology* 38, no. 7: 764–81. https://doi.org/10.1080/13803395.2016.1163327.

Tombaugh, T. N., J. Kozak, and L. Rees. 1999. Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists* 14, no. 2: 167–77.

Treffers-Daller, Jeanine. 2019. What defines language dominance in bilinguals? *Annual Review of Linguistics* 5, no. 1: 375–93. https://doi.org/10.1146/annurev-linguistics-011817-045554.

Troyer, Angela K. 2000. Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology* 22, no. 3: 370–8. https://doi.org/10.1076/1380-3395(200006)22:3;1-V;FT370.

Troyer, Angela K., Morris Moscovitch, and Gordon Winocur. 1997. Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology* 11, no. 1: 138–46. https://doi.org/10.1037/0894-4105.11.1.138.

Wickham, Hadley. 2016. Ggplot2: elegant graphics for data analysis. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley. 2020. Tidyr: Tidy Messy Data. https://CRAN.R-project.org/package=tidyr.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2019. Dplyr: a grammar of data manipulation. https://CRAN.R-project.org/package=dplyr.

Yilmaz, Gülsen, and Monika S. Schmid. 2012. L1 accessibility among Turkish-Dutch bilinguals. *The Mental Lexicon* 7, no. 3: 249–74. https://doi.org/10.1075/ml.7.3.01yil.

Yılmaz, Gülsen. 2019. L1 attrition, L2 development, and integration. In *The Oxford Handbook of Language Attrition*, eds. Monika S. Schmid, and Barbara Köpke, 303–13. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198793595.013.25.

Yılmaz, Gülsen, and Monika S. Schmid. 2018. Chapter 11. first language attrition and bilingualism: adult speakers. In *Studies in Bilingualism*, eds. David Miller, Fatih Bayram, Jason Rothman, and Ludovica Serratrice, 54: 225–50. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/sibil.54.11yil.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2024

ISBN 978-951-29-9983-5 (PRINT)
ISBN 978-951-29-9984-2 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)