Moving in Semantic Space in Prodromal and Very Early Alzheimer's Disease: A Characterisation of the Semantic Fluency Task

Aino Matleena Saranpää
Master's Thesis
Psychology
Faculty of Medicine
April 2020
Supervisors: Sasa Kivisaari, Aalto University and
Jari Lipsanen, University of Helsinki
Research Project: Ambizione

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty Faculty of Medicine		Koulutusohjelma – Utbildingsprogram – Degree Programme Master of Arts (Psychology)
Tekijä – Författare – Author Aino Matleena Saranpää		
Työn nimi – Arbetets titel – Title Moving in Semantic Space in Prodromal and Very Early Alzheimer's Disease: A Characterization of the Semantic Fluency Task		
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track Psychology		
Työn laji – Arbetets art – Level Master's Thesis	Aika – Datum – Month and year April, 2020	Sivumäärä – Sidoantal – Number of pages 50

Tiivistelmä – Referat – Abstract

Objectives. The semantic fluency task is a widely used clinical tool for diagnosing Alzheimer's disease (AD). Performance differences between semantic categories such as animals versus tools have been widely studied, but in the present study, we examine finer-grained performance within the animal-category. We compare amnestic Mild Cognitive Impairment (aMCI) and very early AD patients with healthy controls to investigate whether the groups move in semantic space differently and examine whether the patient groups exhibit decline in discrimination of semantically similar objects, which is putatively a very early sign of AD.

Methods. In the semantic fluency task, participants (42 healthy, 24 aMCI, 18 AD) named as many animals as they could within a minute. We condensed the semantic space using a dimensionality reduction algorithm (t-SNE) on the word2vec vector-model. Sub-categories were formed of the t-SNE result based on visual inspection. Moving in semantic space was estimated with the number of words, sub-categories, and switches and returns to sub-categories. Multinomial logistic regression models were used to predict the diagnostic group with these independent variables.

Results and conclusions. We discovered eight meaningful sub-categories inside the animal-category and found differences between groups in how they utilised the semantic space. Our results did not provide direct support for decline in processing semantically similar objects in prodromal AD. In classifying the groups, only returning to sub-category provided additional information on top of the number of words produced. Our findings provide new perspectives on navigation in sub-category level fine-grained semantic space in the context of prodromal AD.

Avainsanat – Nyckelord – Keywords
Alzheimer's disease, Semantic processing, Semantic fluency task, Verbal Fluency

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Sasa Kivisaari and Jari Lipsanen

Säilytyspaikka – Förvaringställe – Where deposited
Helsingin yliopisto – Helda / E-thesis (opinnäytteet), e-thesis.helsinki.fi

Muita tietoja – Övriga uppgifter – Additional information

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta – Fakultet – Faculty Lääketieteellinen tiedekunta	Koulutusohjelma – Utbildingsprogram – Degree Programme Psykologian maisteriohjelma	
Tekijä – Författare – Author Aino Matleena Saranpää		
Työn nimi – Arbetets titel – Title Semanttisessa avaruudessa liikkuminen erittäin varhaisessa Alzheimerin taudissa: kuvaus semanttisen fluenssin tehtävästä		
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track Psykologia		
Työn laji – Arbetets art – Level Pro Gradu -tutkielma	Aika – Datum – Month and year Huhtikuussa 2020	Sivumäärä – Sidoantal – Number of pages 50

Tiivistelmä – Referat - Abstract

Tavoitteet. Semanttisen fluenssin tehtävää käytetään laajasti kliinisenä työkaluna varhaisen Alzheimerin taudin (AT) tutkimiseen. Ylempiä semanttisia kategorioita kuten eläimiä ja työkaluja on tutkittu laajasti, mutta kategorioiden sisälle ei aikaisemmissa tutkimuksissa ole menty. Tämän tutkimuksen tarkoituksena oli syventyä semanttisen avaruuden alakategorioihin eläinkategorian sisällä sekä kuvailla varhaisten AT- ja aMCI-potilaiden (amnestic Mild Cognitive Impairment eli lievä kognitiivinen heikentyminen) suoriutumista semanttisen fluenssin eläintehtävässä verrattuna terveisiin verrokkeihin. Erityisesti tarkastelimme, heikentyykö varhaisessa AT:ssa semanttisen samankaltaisuuden erottelu, minkä on väitetty liittyvän varhaisen AT:n neuropatologiaan.

Menetelmät. Semanttisessa eläinfluenssitehtävässä koehenkilöitä (42 kontrollia, 24 aMCI- ja 18 AT-potilasta) pyydettiin nimeämään mahdollisimman monta eläintä minuutissa. Tehtävässä nimetyt eläimet muodostivat tutkimuksessa käytetyn semanttisen avaruuden, jota mallinnettiin word2vec-vektoriavaruusmallin avulla. Moniulotteinen vektoriavaruus tiivistettiin t-SNE–dimensionredusointimallilla, jonka avulla jaoimme semanttisen eläinavaruuden kahdeksaan alakategoriaan. Semanttisessa avaruudessa liikkumista arvioitiin kaikkien sanojen, tuotettujen alakategorioiden, alakategoriasta toiseen vaihdosten ja alakategorioihin paluiden määrällä. Vaihdosten ja paluiden itsenäinen vaikutus eristettiin jakamalla ne koko sanojen määrällä. Muuttujien erottelukykyä ryhmien välillä arvioitiin multinomiaalisilla logistisilla regressiomalleilla.

Tulokset ja johtopäätökset. Semanttinen avaruus jaettiin kahdeksaan alakategoriaan, ja löysimme eroja ryhmien välillä siinä miten semanttisessa avaruudessa liikuttiin. Tulokset eivät suoraan tukeneet sitä, että hyvin varhaisessa AT:ssa semanttisesti samankaltaisten asioiden prosessointi olisi heikentynyt. Emme myöskään löytäneet tukea aikaisemmalle tutkimustiedolle, jonka mukaan alakategorioiden määrä ja kontrolloidut vaihdokset tuovat lisäarvoa potilaiden luokittelemiseen sanamäärän lisäksi. Kontrolloitu alakategoriaan paluu taas oli merkitsevä AT-potilaiden luokittelussa, kun sanamäärä oli otettu huomioon. Tuloksemme tarjoavat lisätietoa siitä, miten varhaisessa vaiheessa olevat AT-potilaat liikkuvat semanttisessa avaruudessa semanttisessa eläin-fluenssitehtävässä, mikä voi tuoda lisähyötyä AT:n diagnosointiin. Lisäksi tämä tutkimus osoittaa, että semanttista avaruutta voidaan tarkastella myös alakategorioiden tasolla.

Avainsanat – Nyckelord – Keywords Alzheimerin tauti, muistisairaudet, semanttinen sujuvuus
Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors Sasa Kivisaari ja Jari Lipsanen
Säilytyspaikka – Förvaringställe – Where deposited Helsingin yliopisto – Helda / E-thesis (opinnäytteet), e-thesis.helsinki.fi
Muita tietoja – Övriga uppgifter – Additional information

Acknowledgements

First and foremost, I would like to express my gratitude to my first supervisor, Dr. Sasa Kivisaari (Aalto University) for providing me invaluable support throughout this project and giving me their expert opinions on the substance matter in this research. Further, I want to thank them for their understanding and compassion throughout this very intense journey. I would also like to thank my second supervisor Mr. Jari Lipsanen (University Instructor, Department of Psychology and Logopedic, University of Helsinki) who gave me valuable insights in choosing the methods for the present study and providing essential comments during the writing process.

The data for the present thesis was collected in the Memory Clinic FELIX PLATTER, University Department of Geriatric Medicine FELIX PLATTER, Basel, Switzerland. As I was not part of the original study, I would like to thank Dr. Sabine Krumm (Felix Platter Hospital), Dr. Kirsten I. Taylor (University of Basel & Roche Pharma Research and Early Development) and Prof. Andreas U. Monsch (University of Basel) for giving me the opportunity to utilise this exceedingly valuable data that has broadened our insights in the early Alzheimer's disease and its detection.

Finally, I want to thank my family and friends for their support through this research. I wish to express my deepest gratitude for my partner Niilo with whom I could freely converse my ideas and who was responsible for doing most of the household work during this academic year. This project would not have been possible without you.

Aino Saranpää

In Helsinki
28th of April, 2020

Contents

1 Introduction

The semantic fluency task, where the participants need to produce as many words as possible in a certain semantic sub-category in a specific time frame, is widely used in clinical settings to identify difficulties in speech production, executive functioning and semantic memory performance (Lezak, Howieson, Loring, & Fischer, 2004). In the task, participants are asked to name as many words as possible in a given semantic category (typically animals). Participants may use different strategies to map the semantic space and therefore the task provides rich information on cognitive functioning and semantic processes. This study has two aims. Firstly, I examine, whether words produced in a single semantic category (i.e., animals) can reliably be divided into sub-categories which would provide more detailed information on how individuals move in the semantic space. Secondly, I examine whether moving in and between these sub-categories can be used to provide more understanding of the semantic fluency task as a tool for early detection of Alzheimer's disease (AD).

Difficulty in the semantic fluency task is a viable predictor of early AD (Henry, Crawford, & Phillips, 2004). This is since the task requires many cognitive processes such as accessing and retrieving information from the semantic memory (Birn et al., 2010). It has been shown in behavioural studies that patients with early AD produce fewer words in this task compared to healthy controls, especially if they are asked to name living things (Krumm et al., 2019). Troyer, Moscovitch, and Winocur (1997) have suggested that the strategies individuals use in producing as many words as possible in the semantic fluency task involve producing clusters ("sub-categories") of objects (e.g., naming *dog, cat* and *rabbit*) and switching ("crossing") from one sub-category to another (e.g., moving to *cow* and *horse*). There is evidence from previous literature that in addition to AD patients naming less objects, they also produce smaller sub-categories and perform less crossings between the sub-categories (Troyer, Moscovitch, Winocur, Leach, & Freed-

man, 1998). However, previously these sub-categories have been estimated with a subjective evaluation protocol such as the one proposed by Troyer and colleagues (1997). To our knowledge, there are few studies that have used on a quantitative estimation method in attempting to extensively study the sub-categories that constitute the semantic space inside the specific task, such as naming animals. Furthermore, it is still unclear whether there are differences between healthy controls and AD patients in the way they move from one sub-category to another inside the semantic space in the naming task, when the number of words produced is taken into account.

The present study analysed the longitudinal, originally normative sample from the Ambizione study, collected in the Memory Clinic FELIX PLATTER in Basel, Switzerland. The data consists of neurologically screened healthy controls and highly functional AD and amnestic mild cognitive impairment (aMCI) patients, as indicated by the very high average scores of the patients in the Mini-Mental-State Examination (MMSE). Our aim was to examine whether there are differences in the way that very early and prodromal AD patients move in the semantic space compared to healthy participants in the semantic fluency task, more specifically in the animal fluency task. To do this, we divide the semantic space into smaller sub-categories with a dimensionality reduction algorithm. More specifically, the aim was to investigate whether the patient groups have difficulties in producing semantically similar words. We define this difficulty as not being able to stay within a sub-category but instead exhibiting more crossings from one sub-category to another compared to healthy controls. Furthermore, we want to describe whether the healthy participants exhibit different strategies in how they name objects in the semantic fluency task. Finally, our objective is to analyse whether using the category and crossing dimensions in addition to the number of the words brings additional information into the diagnostics of the early and prodromal AD.

1.1 Semantic processing

Internal, conceptual knowledge of the meanings of words is called a semantic representation (Ellis & Young, 2013). These meanings can be modal such as visual or auditory characteristics (e.g., 'has a nose'), functional properties (e.g., 'barks') or encyclopaedic information (e.g., 'is a pet'). Same features can also describe other words, but an object can be recognised by an individual property, that is, individuals can produce an image of a dog in their mind just by hearings its bark. Next, we will present some mechanisms that try to explain how semantic information is represented in neural networks and how it is possible to unite these fragments of information into one object.

Most contemporary theories on semantic processing support the idea that semantic processes involve brain regions responsible for perception and action (Patterson, Nestor, & Rogers, 2007). Semantic information often consists of visual, auditory, tactile, gustatory and olfactory information combined with the object's motor affordances (i.e., how an object can be used) and the language used to describe the object (Patterson et al., 2007). Therefore, the basis of semantic processing lies in the co-activation of sensorimotor tracts in order to produce concepts such as a dog. This view abandons the previously held theoretic assumption that there are single neurons responsible for producing individual objects (such as the Jennifer Aniston neuron, see, e.g., Quiroga, 2012).

Contemporary theories on what brain regions are responsible for semantic processing can be divided by whether they assume that distributed networks are sufficient for semantic processes or whether the existence of a semantic hub is needed. Patterson and colleagues (2007) have suggested that networks themselves are insufficient to explain how objects with conceptual overlap can be distinguished. For instance, it is unclear how individuals are able to separate different insects from each other, when these objects share so many similar features. This suggests that

the differences between objects cannot be established purely by the co-activation of perceptual features. In their model, Patterson and colleagues have added a semantic hub, which represents a high-level process that is able to generalize over semantic sub-categories. They assume that in order to achieve the higher-order generalizations which our semantic system relies on, the existence of a hub is vital. Based on clinical and neuroimaging studies, they suggest that the hub is located in the anterior temporal lobe. This assumption is supported by research which has revealed that many primary sensory and motor areas are connected to the anterior temporal lobe (see, e.g., Bonner & Price, 2013).

Taylor and colleagues present an alternative view to Patterson's hub-based theory: the Conceptual Structure Account (CSA; Taylor, Devereux, & Tyler, 2011). The CSA assumes that semantic processing is structured according to the statistical properties of the object's features, and processing of semantic concepts corresponds to the co-activation of the concept's features. The properties that structure the semantic space are called feature distinctiveness and feature correlation (Taylor et al., 2011). The former refers to the extent which a feature is shared by other concepts, and the latter stands for feature co-occurrence. For instance, the features 'has eyes' and 'has a tail' are high in feature correlation, that is, they can simultaneously describe many objects. Features such as 'has a trunk' are distinctive and can uniquely distinguish between objects. Taylor and colleagues (2011) claim that "feature co-occurrence and distinctiveness interact to determine conceptual processing as a function [of] *task demands*, i.e., the information required to perform the task at hand".

According to the CSA theory, differences in feature co-occurrence and distinctiveness explain why some objects are easier to discern from each other (Taylor et al., 2011). If an object is low on feature distinctiveness, that is, has many ambiguous features, and those features are shared with other objects, this makes it hard to distinguish the object from others within its category. As living things, such as

animals, are often high in feature correlational strength and have fewer distinctive features less correlated with other features compared to non-living things, this makes them more prone to difficulties in distinction (Randall, Moss, Rodd, Greer, & Tyler, 2004). Therefore, the CSA suggests that objects belonging to the living or non-living categories differ in their internal structures. Living objects (such as animals, or fruits and vegetables) form larger categories that have highly shared features, such as 'has legs' and 'has eyes'. (Taylor et al., 2011). However, since living things are not high in distinctive features, distinguishing living things *from each other* tends to be impaired. Based on clinical studies, there is evidence that impaired naming of living objects is connected to lesions in bilateral antero-medial structures and inferior temporal lobes (Gainotti, Silveri, Daniel, & Giustolisi, 1995; Krumm et al., 2019).

Feature statistics are combined with a neurocognitive approach in the CSA (Taylor et al., 2011). From non-human primate studies, there is evidence that visual objects are processed based on their features in a hierarchical system extending from posterior occipital to anteromedial cortex (see, e.g., Mishkin, Ungerleider, & Macko, 1983). Similar functions are proposed for the object processing system in humans (Damasio, 1989). Since semantic processing involves the combination of multi-modal sensory inputs, the CSA suggest that conceptual processing relies on hierarchically organized sensory streams (Taylor et al., 2011). At the apex of all of these streams lies the perirhinal cortex (PRC) (Suzuki & Amaral, 1994). In the CSA framework, the PRC has a critical role in processing the most fine-grained feature conjunctions and combining unimodal information into multimodal object representation, though this function may not be limited to the PRC (Taylor et al., 2011).

In the anterior medial temporal lobe, the PRC is located inside the collateral sulcus in the fusiform gyrus, and its medial portion (mPRC) corresponds to the transentorhinal cortex (Kivisaari, Probst, & Taylor, 2013). MPRC has a key role in ob-

ject recognition memory (Hirni, Kivisaari, Monsch, & Taylor, 2013), and therefore for a long time, it was only studied in the context of amnesia and the wider medial temporal lobe "memory system" (see, e.g., Squire, Stark, & Clark, 2004). More recently, the mPRC has been suggested to have a key role in recognizing fine differences between visual objects (Buckley, Booth, Rolls, & Gaffan, 2001), but also differences of semantic nature (Connolly et al., 2012) and processing object-specific information (Clarke & Tyler, 2014). One explanation for the importance of the mPRC to semantic object processing has been suggested by Libby, Ekstrom, Ragland, and Ranganath (2012), who considered the connections from the PRC to multi-modal sensory areas important. Kivisaari, Tyler, Monsch, and Taylor (2012) demonstrated that damage in the PRC disproportionately hinders the processing of visually and semantically ambiguous objects as compared with objects which have more distinctive features. Therefore, atrophy in the PRC, and more specifically in the mPRC, could affect identifying and naming objects that are semantically similar to each other.

To conclude, semantic information is processed in wide networks of sensory-motor and executive areas, but the processes involving multimodal object representations and distinguishing between fine-grained differences of objects seem to involve the participation of the medial temporal lobe, and more specifically the mPRC. Based on the previously presented literature, we suggest that if individuals that have atrophy in the medial temporal lobes have difficulties in distinguishing similar objects, these difficulties may be reflected in how they move in the semantic space. In the semantic fluency task, we can assume that healthy participants utilize a strategy in which they are able to name similar words (such as *dog* and *cat*) and then move to another set of similar words (such as *cow* and *pig*). However, patients who have atrophy in the medial temporal lobes or more specifically, in the mPRC, might display a strategy in which they are less likely to name similar objects but rather more likely to move all around the semantic space (naming objects such as *dog, elephant* and *eagle*).

1.2 Alzheimer's disease

Alzheimer's disease (AD) is a progressive neurodegenerative disease which is the main cause of dementia (see, e.g., Terry & Katzman, 1983). If we extrapolate linearly from the results of Brookmeyer, Johnson, Ziegler-Graham, and Arrighi (2007), approximately 40 million people have AD worldwide at the present moment, and that the prevalence of AD grows to 106.8 million by the year 2050. Therefore, AD will become an even more pressing problem concerning public health and a great cost to healthcare and social welfare systems all over the world (Wimo, Jonsson, & Winblad, 2006). Further, there is evidence that the pathological process of AD begins years or even decades before the diagnosis of clinical AD (Amieva et al., 2008), which indicates a long preclinical stage of the disease (Sperling et al., 2011). Thus, there is a distinct need for tools that help clinicians distinguish AD at the earliest stage possible.

Mild cognitive impairment (MCI) is considered a transitional stage between healthy cognitive functions and clinically probable AD according to the Winbald's and colleagues' (2004) criteria and is sometimes referred to as prodromal AD. Especially when MCI is amnestic (aMCI), patients have a high probability of being diagnosed with AD later on, compared to patients with non-amnestic MCI (Fischer et al., 2007). It has been suggested that some risk factors, such as anxiety, predict the progression from aMCI to AD (Palmer et al., 2007). The criteria for aMCI include memory complaints usually by an informant, objective memory impairment relative to age, but preserved general cognitive function without the individual being demented, and normal daily living activities (Petersen, 2004). Therefore, aMCI patients may exhibit some decline in memory functions but overall are more high functioning compared to AD patients at even very early stages. Based on these findings, we will examine aMCI and very early AD patients as separate groups in the present study.

1.2.1 Neuropathology of AD

The neuropathology of AD is related to accumulation of amyloid plaques (Hardy & Higgins, 1992) and changes in intracellular neurofibrillary tangles and neuropil threads (Braak & Braak, 1997). Braak and Braak (1991) have divided the neuropathological changes in AD into six different stages, which are characterized by patterns in neurofibrillary tangles and neuropil threads. Stages I–II are characterized by alterations in the transentorhinal layer, which corresponds to the mPRC (Taylor & Probst, 2008). Stages III–IV involve the limbic areas, the entorhinal and transentorhinal layer. Finally, at stages IV–V the alterations progress to the isocortical areas (Braak & Braak, 1991).

There is evidence that the neuropathological stages correspond to the clinical progression of AD (Fewster, Griffin-Brooks, MacGregor, Ojalvo-Rose, & Ball, 1991). In their review, Almkvist (1996) presents a model which matches the neuropathological progression to preclinical and clinical stages of AD. In the transentorhinal stage, episodic memory is affected. In the limbic stage, impairment in verbal abilities, visuospatial functions, attention, and executive functions are possible. In the isocortical stage, where AD is fully developed, there is severe impairment in primary memory functions and episodic and verbal skills as well as visuospatial and executive functions. However, later evidence has shown that prior to the changes in the episodic memory as proposed by Almkvist (1996), decline in semantic memory and conceptual formation may appear up to 14 years prior to clinically defined dementia (Amieva et al., 2008). Based on the previously presented literature on the mPRC, we consider this decline in semantic processing to reflect the transentorhinal stage of AD. Thus, tasks measuring semantic memory, such as the semantic fluency task, could be utilized to expose AD at its very early stages. Since the mPRC is affected in the early stages of AD, we suggest that in the semantic fluency task especially separating similar things may be diminished in the early stages of AD.

1.3 Semantic fluency task performance in AD

The semantic fluency task, which has been widely used to measure semantic memory in AD, is considered to measure not only basic language capacity but also the executive functions related to naming objects, for example self-monitoring, flexible thinking and working memory (Lezak et al., 2004). The semantic animal fluency task is instructed to the participant by asking them to name as many animals as possible within a minute. The category provides structure for the task and the participant must find a suitable strategy to efficiently produce words in the task (Lezak et al., 2004).

Troyer and colleagues (1997) have suggested that the strategies in the semantic fluency task can be divided to two: 1) producing objects inside a sub-category and 2) crossing, that is, moving to a new sub-category. Both of these strategies are needed to efficiently produce objects in the task. Healthy participants are able to navigate in the semantic space systematically and utilize different sub-categories of objects in the task. For instance, a commonly used strategy for healthy individuals is to start producing sub-categories (e.g., farm animals) and cross to another (e.g., birds) when one sub-category is exhausted.

There is ample evidence that AD patients perform differently from healthy control participants in the semantic fluency task, especially regarding the category production and crossing aspects of the task. Not only do AD patients name fewer words compared to healthy controls (Fagundo et al., 2008; Price et al., 2012; Raoux et al., 2008; Troyer et al., 1998), but they also create smaller sub-categories (Fagundo et al., 2008; Troyer et al., 1998), cross less between sub-categories (Fagundo et al., 2008; Raoux et al., 2008) and create fewer sub-categories (Pekkala, 2004). However, there have also been studies that have not found differences between sub-category sizes (Epker, Lacritz, & Munro Cullum, 1999; Pekkala, 2004; Raoux et al., 2008) and crossing behaviour (Price et al., 2012). These inconsistencies have

been explained by the severity of the dementia, demographic variables, sampling differences and study design differences (Raoux et al., 2008). Further, there are studies that have found that already aMCI patients differ from healthy controls in the number of words (Lonie et al., 2009) and sub-category sizes (Price et al., 2012), which supports the idea that semantic processing becomes gradually more limited in the succession of the disease. However, there are only a few studies that have researched whether the sub-category production and crossing performance can be used to differentiate healthy controls from amnestic patients. Epker et al. (1999) performed hierarchical clustering analyses to test whether there were differences in how well the number of words, sub-category size and crossings would differentiate between healthy controls, both amnestic and non-amnestic Parkinson's disease patients and AD patients in semantic and phonemic fluency task. They found that in the semantic fluency task, the number of words classified correctly 60% of the AD patients and crossing 62%, while sub-category size only classified 8% of AD-patients correctly. However, they also did not control for the total number of words in assessing differences in crossing or sub-category size, so it might be that a major part of the classification effect of the crossing sub-category could be explained simply with the number of words the participants produce in the task.

Fagundo et al. (2008) have examined the overall effects of producing sub-categories and crossings. In their study, they combined the number of words, sub-category size and the number of crossings into one model to see which of these variables predicted the development of AD. They found that when comparing individuals with memory complaints, some of which developed an AD diagnosis, only the mean sub-category size was significant in predicting the development of AD. They suggested that in the progression of AD, the participants produce smaller sub-categories. However, in the model, the coefficients for the number of words and number of crossings were non-significant but *positive*, which indicates that the number of words and crossings should *increase* rather than decrease in the progression of AD, which is in conflict with the previous literature. Intuitively, the per-

formance in producing sub-categories can have a very strong positive relationship with the number of words produced in the task and the number of crossings. The more words are produced within a sub-category and the more crossings are made between sub-categories, the more words are produced overall. Therefore, we suggest that these results can be interpreted as possible multicollinearity problems in the model, which in turn limits confidence in the results presented by the authors on the differentiating power of the sub-category size.

To our knowledge there are no studies that would have investigated the independent effects of the production of sub-categories and crossings in the semantic fluency task in addition to the well-established effect in the decline of the number of words in the progression of AD. We aim to examine the individual effects of producing sub-categories and crossing from a sub-category to another, and analyse whether they have clinical significance in diagnosing early AD.

1.4 Sub-categories in the semantic space

Previously presented behavioural studies on the semantic fluency task have all used the Troyer et al. (1997) method, where the categories have been divided into sub-categories based on subjective evaluation. Even though this method entails clear instructions, some studies have questioned the method's validity (Epker et al., 1999). Other subjective measures have also been presented, but they too suffer from issues such as low inter-rater reliability or test-retest reliability (Abwender, Swan, Bowerman, & Connolly, 2001). Therefore, the results from studies that use these methods may be difficult to replicate.

In the neurocognitive framework, research has mostly focused on comparing general higher level categories, such as living versus non-living (Kivisaari et al., 2012; Krumm et al., 2019; Tyler et al., 2013) and animals, fruits, tools and vehicles (Clarke & Tyler, 2014; Kivisaari et al., 2019). However, few studies have investigated finer-

grained performance within single semantic categories. Especially the animal category has been considered to be more dense compared to for instance tools and vehicles, meaning that the animals are more similar among one another than tools or vehicles which in turn contain more distinctive features (Randall et al., 2004). Examining performance within such a dense category seems a fruitful approach in examining the cognitive decline in AD as difficulties in fine-grained semantic and perceptual discrimination are putatively the very early signs of the disease (Barense, Henson, Lee, & Graham, 2010; Kivisaari et al., 2012; Taylor, Moss, Stamatakis, & Tyler, 2006). In the present study, we aim to investigate whether the difficulties in semantic discrimination can be found already in prodromal and very early AD patients using the dense semantic space constructed of animals that participants name in the semantic fluency task. To achieve this objective, we use a modern corpus-based method for producing feature-vectors and a dimensionality reduction model that allow us to accurately and efficiently model the semantic space and the sub-categories within.

1.5 Research questions and hypotheses

Neurocognitive and behavioural frameworks provide two alternative assumptions for the performance of the participants in the semantic fluency task. In the present study, we wanted to examine whether the diagnostic status, driven by putative medial temporal lobe pathology, is related to naming similar things in early onset and prodromal AD. Based on the evidence from neurocognitive studies, we assumed that aMCI and AD patients have difficulty in naming semantically similar objects inside a sub-category (i.e., birds or forest animals within the animal category) and therefore exhibit more crossing behaviour from one sub-category to another, when the number of words is controlled for. Furthermore, we wanted to examine whether using the category and crossing dimensions in addition to the number of the words brings additional information into the diagnostics of very early and prodromal AD.

Finally, we aimed to describe how the participants move in the semantic space, and whether the healthy participants exhibit different strategies in how they move in the semantic space in the semantic fluency task.

2 Methods

2.1 Participants

In total, 181 native Swiss-German or German speaking adults were recruited in the original Ambizione study. All participants with the available data to match the aims and research questions of the present study (42 patients and 42 matching healthy controls) were chosen from the original Ambizione study to form the final sample of 84 participants. 42 participants (21 identified as male; mean age 74.4 yrs; SD 7.3 years) belonged to the control group and confirmed cognitively healthy through medical and neuropsychological screening. In the patient group, there were 42 participants (20 identified as male; mean age 74.3 yrs; SD 6.8 yrs) that were matched to the control group according to age, gender and education. In the patient group, 24 participants were diagnosed with aMCI due to AD (Albert et al., 2011) according to DSM-IV (American Psychiatric Association, 1994) and Winblad et al. (2004) criteria. Eighteen participants were diagnosed with very early Alzheimer's dementia according to DSM-IV (American Psychiatric Association, 1994) and NINCDS-ADRDA (McKhann et al., 2011) criteria. The diagnoses were based on a consensus between an interdisciplinary team of experienced clinicians. Demographic information of the different groups can be found in Table 1. Since the groups differed in age and there is evidence that age affects the performance in the semantic fluency task (Troyer et al., 1997), we used age as a covariate in later analyses. As expected, the participants differed in the MMSE scores but both the aMCI and AD groups scored very high points in the test, which indi-

cates a very early stage of AD. The data for the original study was collected in the Memory Clinic FELIX PLATTER, University Department of Geriatric Medicine FELIX PLATTER, Basel, Switzerland. All participants provided an informed consent prior to participating in the study. The study was approved by the local ethical committee of Both Basels.

Table 1. Demographic information.

	Healthy (n=42)	aMCI (n=24)	AD (n=18)	χ^2	p
Sex	21 males	10 males	10 males	0.84	.657
Variable	Mean (SD)			F	
Education	12.86 (3.14)	13.08 (3.16)	12.22 (3.14)	0.41	.667
Age (yrs.)	74.38 (7.32)	71.34 (6.59)	78.32 (4.76)	5.68	.005
MMSE	29.31 (1.00)	28.67 (1.46)	26.61 (1.79)	25.83	$< .001$

aMCI = amnestic Mild Cognitive Impairment, AD = Alzheimer's disease, MMSE = Mini-Mental State Examination

2.2 Task

In the semantic fluency task, participants were asked to name as many objects inside a certain semantic category as they could within a minute. The object categories were animals, fruits, tools and vehicles. In this thesis, we focused on the animal category, due to the research on the difficulties AD patients have distinguishing living things in particular (Krumm et al., 2019). As instructed by Troyer et al. (1997), we utilised all words produced in the task in the further analyses, and did not differentiate between correct and non-correct (indicated by repetitions or perseveration) words.

2.3 Data analyses

In the present study, we described semantic space as a set of distance-based clusters rather than continuous distance metrics. This is since the information related to the semantic distance of the objects is meaningful only when the objects are relatively close in semantic space. For example, *dog* and *cat* are intuitively semantically quite similar and close, but is there a difference in the distance between object-pairs such as *dog-whale* and *dog-worm*? Therefore, we describe semantic proximity as belonging to the same cluster or sub-category. We first estimated semantic distance, that is similarity, with cosine distance to best model it, and then created distinct sub-categories based on this distance metric.

We estimated the semantic similarity of the named objects using a text corpus, that is, the 3B-token Google News dataset (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). The words were first translated from Swiss-German to English. We used a pre-trained word2vec skip-gram model to find vector representations that predict surrounding words of the given object in a sentence. With the method, dense vector representations of words were comprised from unstructured text data, that is, corpus (Mikolov et al., 2013). The code can be found online at `https://code.google.com/archive/p/word2vec`. The text corpus was used to estimate semantic similarity via measuring the cosine distance between all concept-feature vectors. Each row in the acquired matrix described how semantically similar the object is to all other objects within that category, estimated from zero to one, where values close to zero indicate very similar representations and values close to one very distant representations.

2.3.1 Dimensionality reduction

For the classification of the semantic distance, we used an unsupervised, non-linear dimensionality reduction technique known as t-Distributed Stochastic Neighbor

Embedding (t-SNE) developed by Maaten and Hinton (2008) which can be used to visualise the structure of high-dimensional (HD) data with low-dimensional (LD) maps such as two-dimensional scatter plots (Maaten & Hinton, 2008). T-SNE aims to retain local structures of the data by preserving the distances between points and their nearest neighbours from the original HD data to the LD map. This is done by plotting Gaussian distributions for each point in the HD data and measuring the density of the other points under the Gaussian. The acquired probability functions are compared to similarly acquired t-distributed similarity functions in the LD data and are measured by the Kullback-Leibler divergence, which t-SNE tries to minimize. Student's t-distribution is used because it allows for better modeling of far apart distances, since it does not give as much emphasis on values at the extreme ends of the distribution (Maaten & Hinton, 2008).

It has been suggested by Maaten and Hinton (2008) to use some other dimensionality reduction technique for the data prior to using t-SNE to improve t-SNE performance in data sets with a high number of features. In the present study, we used Multidimensional scaling (MDS) prior the t-SNE. MDS is a technique for visual representation of distances or dissimilarities between sets of objects and can be used as a dimensionality reduction technique (Buja et al., 2008). We used MDS to reduce the number of the dimensions from 224 (the number of the unique words in the data) to 50 (as suggested by Maaten & Hinton, 2008) which was the number of dimensions that explained 96 percent of the variance in the data. Finally, t-SNE was implemented on the 50 acquired dimensions from the MDS. Based on the visual inspection of the two-dimensional plot that was acquired as the t-SNE result, we divided the 224 unique objects into sub-categories, so that each object belonged to one sub-category.

The dimensionality reduction model was executed with Python 3.7 using the package sklearn.manifold (Pedregosa et al., 2011). Multiple model solutions with different perplexity parameter values were executed. The perplexity parameter defines

the number of points falling under the probability distribution, thus perplexity can be considered to set the number of effective nearest neighbors estimated for each point, and is suggested to be somewhere between 5 to 50 in a t-SNE model (Maaten & Hinton, 2008). As we were interested in the local clusters, we used a perplexity value small enough which was 20 in the final model. However, different perplexity values did not greatly affect the overall output of the model (see Appendix). A number of 1500 iterations was found to establish a stable model. T-SNE was run multiple times, as suggested by Maaten and Hinton (2008) to achieve the lowest Kullback-Leibler divergence, which was 0.88 in the final model.

2.3.2 Statistical analyses

Statistical analyses were executed with IBM SPSS Statistics 25. To examine the participants' performance in the semantic fluency task, we calculated the sum of objects in each sub-category per participant (i.e., how many pets or birds the participant named) and summed all these objects to get the total number of words produced in the task ('Number of words'). We also examined the number of words in each sub-category and divided it by the number of words produced in the task to get proportional information of each sub-category. In addition, we recorded the number of sub-categories visited in the task ('Sub-categories named'). We defined crossings ('Crossings') as moving from one sub-category to another, calculated the sum of crossings for each participant and divided that by the number of words each participant produced ('Adjusted crossings'). As we tried to examine movement in the semantic space as thoroughly as possible, we also decided to inspect a novel variable that would capture not only unidirectional movement from cluster to cluster but also describe revisiting previously utilised areas in the semantic space. For this purpose, we examined the number of times a participant returned to a sub-category which they had previously visited ('Returns') and adjusted that number with the total number of words produced ('Adjusted returns').

Normality of the data was evaluated with Q-Q plots. The effect of belonging to a diagnostic group on the number of words produced in the task was examined with one-way analysis of variance (ANOVA). To examine the effects of belonging to a diagnostic group (healthy, aMCI and AD) on moving in the semantic space, the number of sub-category, crossings, adjusted crossings, and adjusted returns variables were used as dependent variables in one-way multivariate analysis of variance (MANOVA), which was conducted to minimize the likelihood of type 1 error. Different sub-categories were also examined with one-way MANOVA with diagnostic group as an independent variable. Since some of the variables were not normally distributed, we used bootstrapped parameter estimates with bias-corrected and accelerated bootstrap interval in all pairwise comparisons. Separate linear multinomial logistic regression models were used to predict diagnostic group with the number of categories, adjusted crossings and adjusted returns as independent variables. Since there is evidence that age affects the performance in the semantic fluency task (Troyer et al., 1997), we used it as a control variable in addition to the number of words in each model. We did not combine the independent variables into one model, due to multicollinearity issues between some variables.

3 Results

3.1 Corpus data

The results from the t-SNE analysis based on the 50 components produced by MDS can be seen in Figure 1. Based on visual inspection, we formed eight subcategories based on the t-SNE result: pets, birds, forest animals, jungle animals, aversive animals (consisted of reptiles and insects), farm animals and sea life. T-SNE was able to produce some very tight categories (such as sealife in the left bottom corner) and some sub-categories that are more loose (such as birds in the bot-

tom centre). However, the visualization of the semantic space appears to produce sensible categories that were meaningful for further analyses.

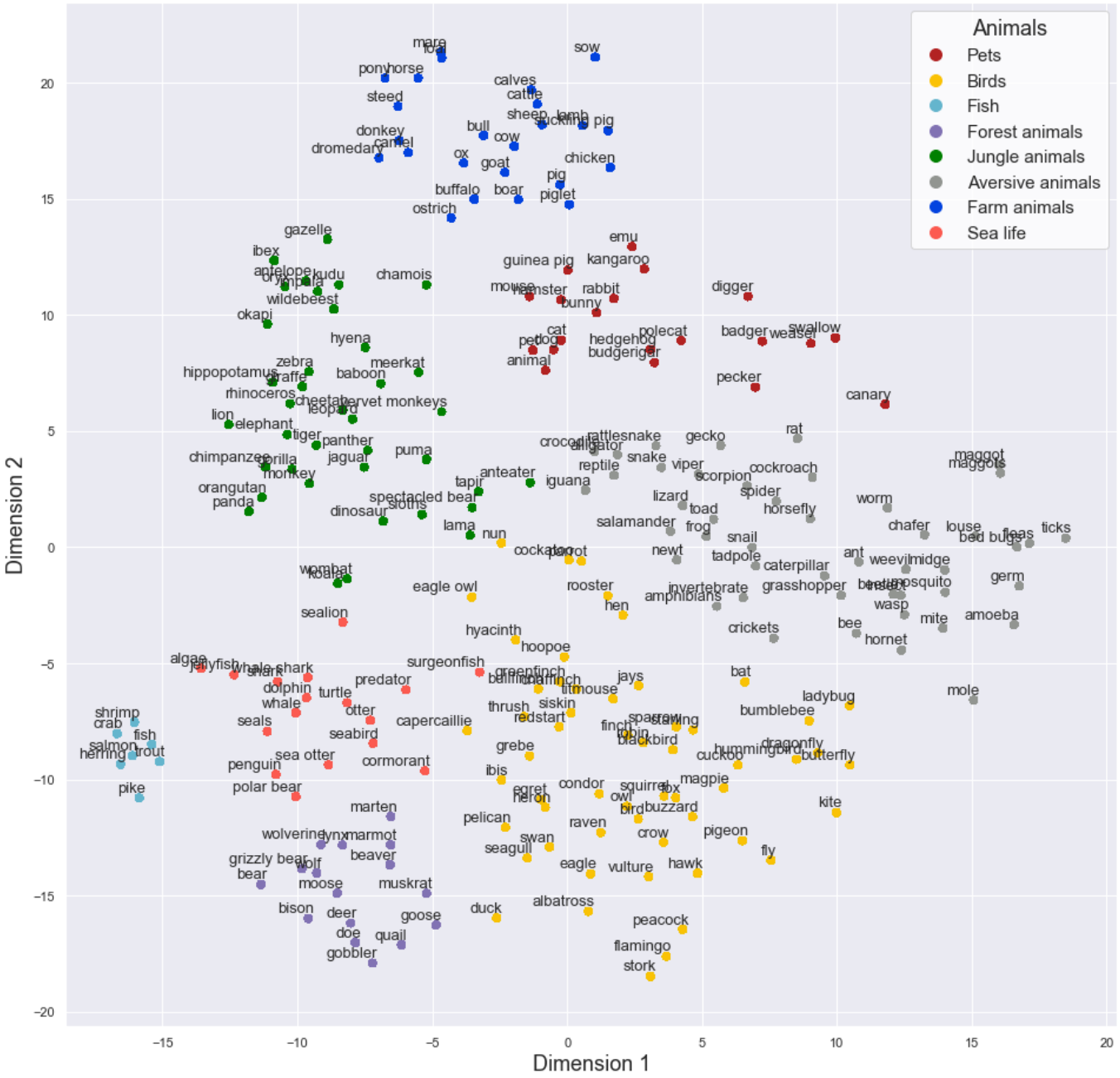

Figure 1. Two-dimensional visualisation of the animals produced in the semantic fluency task by the t-SNE model on the 50 dimensions of multidimensional scaling. In the figure, different sub-category labels are presented based on visual inspection as different colours.

3.2 Behavioural data

We found that there was a significant difference between groups in the number of words produced in one-way ANOVA [F(2,81) = 23.49, $p < .001$, partial $\eta^2 = 0.37$]. As expected, healthy participants named more animals compared to both aMCI and AD patients, and aMCI patients named more compared with AD patients (Figure 2). The sub-categories produced by each participant are presented in Figure 3. Overall, farm animals, jungle animals, pets and birds were the most often named sub-categories. Interestingly, there were many participants in all groups that named multiple birds in the task. Fish and sea life were sub-categories that were used less often, as many participants named zero to one objects from these sub-categories. There were few outliers in the data.

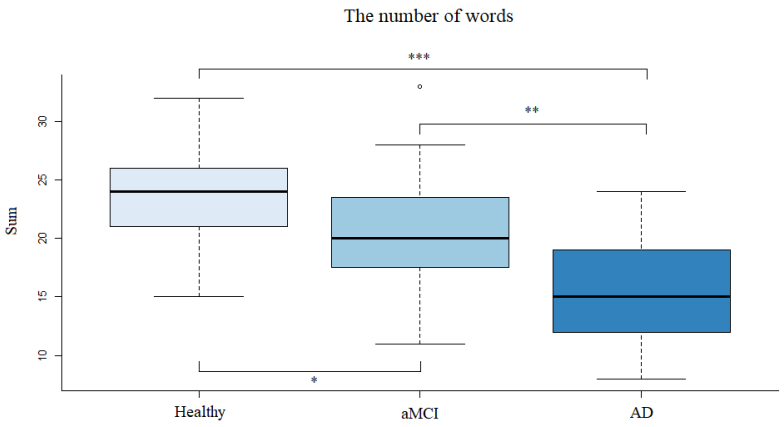

Figure 2. The number of words produced in the task by groups. Healthy participants name more words compared to both patient groups and aMCI patients name more compared to AD patients. In the boxplot, minimum, first quartile, median, third quartile and maximum are shown. The circle represent an outlier.
*$p < .05$, ** $p < .01$, *** $p < .001$.

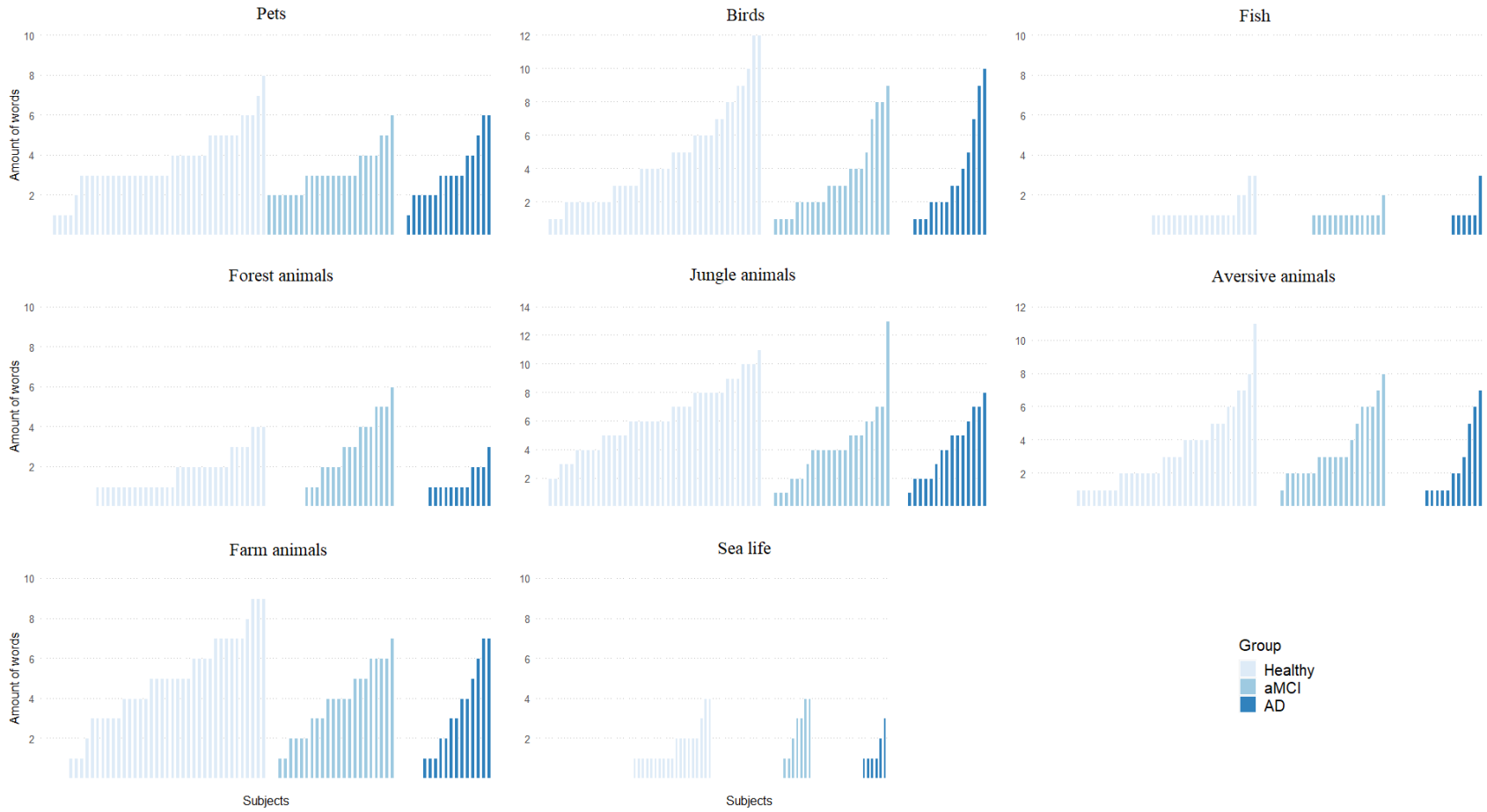

Figure 3. Total numbers of words produced in each animal sub-category for each individual participant by group. Scores are sorted by magnitude in each sub-category.

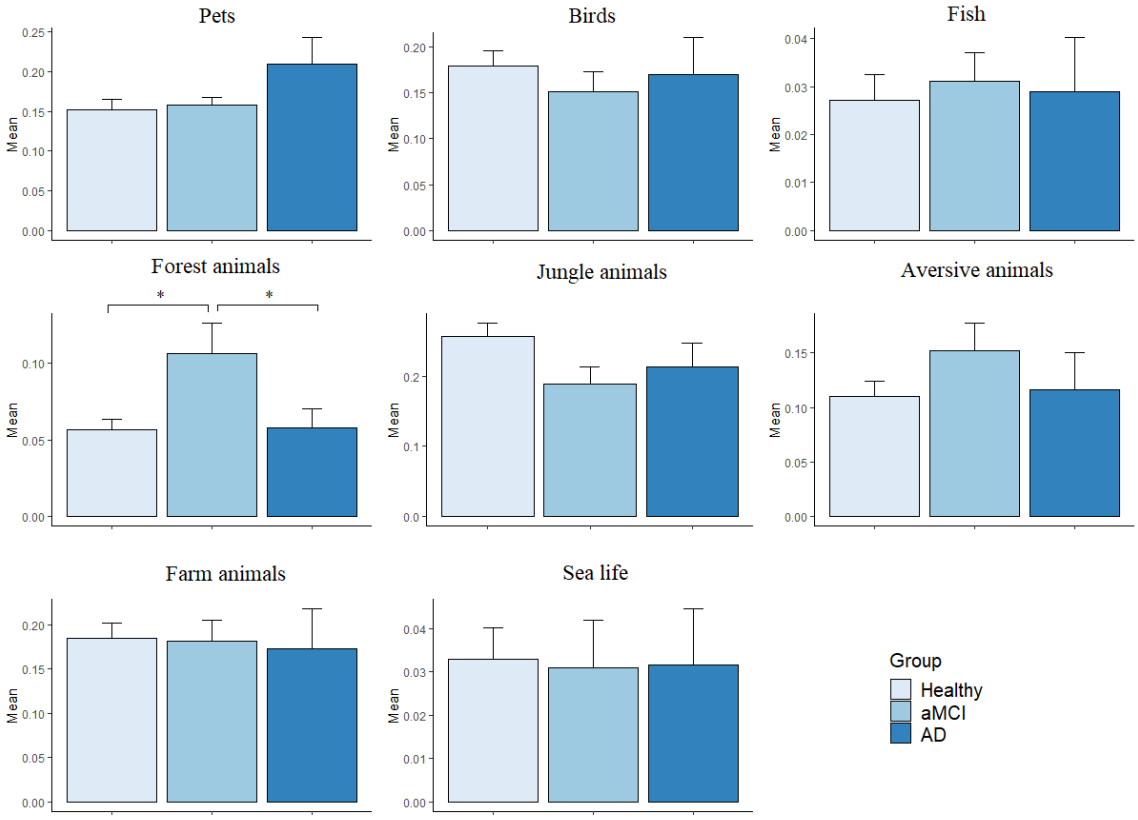

Figure 4. The means and standard errors of the proportional share of each sub-category (i.e., number of objects in sub-category divided by the total number of words) by groups. *$p < .05$

There was a statistically significant difference in the number of words in each sub-category based on the group in one-way MANOVA [F (16,148) = 4.17, $p < .001$, Wilk's $\Lambda = 0.48$, partial $\eta^2 = .31$]. We found differences between groups in forest animals [F (2,81) = 4.61, $p = .013$, partial $\eta^2 = .10$], jungle animals [F (2,81) = 7.26, $p = .001$, partial $\eta^2 = .15$] and farm animals [F (2,81) = 3.71, $p = .029$, partial $\eta^2 = .08$]. Healthy controls named more jungle animals compared to aMCI ($p = .004$, bootstrapped 95% CI: -3.26 – -0.62) and AD group ($p = .002$, bootstrapped 95% CI: -3.84 – -1.07), and produced more farm animals compared to AD patients ($p = .011$, bootstrapped 95% CI: 0.37 – 3.18). Further, aMCI patients named more forest animals compared to AD patients ($p = .008$, bootstrapped 95%

CI: 0.42 – 2.01). To examine whether these effects were only related to the number of words the participants produced, we ran a one way MANOVA on forest, jungle and farm animals when the number of words was controlled for (Figure 4). There was a statistically significant effect in the adjusted number of words in the sub-categories based on participant group [F (6,158) = 2.17, p = .048, Wilk's Λ = 0.85, partial η^2 = .08]. We found statistically significant differences between groups in forest animals [F (2,81) = 4.85, p = .010, partial η^2 = .11]. AMCI group named more forest animals compared to both healthy (p = .020, bootstrapped 95% CI: -0.09 – -0.01) and AD patients (p = .022, bootstrapped 95% CI: -0.092 – -0.004).

Table 2. Spearman correlations between semantic fluency variables and age.

Variables	1	2	3	4	5	6
Number of words						
Sub-categories named	.46***					
Crossings	.65***	.68***				
Adjusted crossings	-.01	.44***	.61***			
Returns	.59***	.39**	.93***	.59**		
Adjusted returns	.24*	.25*	.78***	.80***	.90***	
Age	-.16	-.21	-.33**	-.25*	-.28**	-.22*

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

The number of words had a strong positive correlation with the number of crossings and sub-categories the participants produced (Table 2). When the total number of words was controlled for, the adjusted crossings did not have a significant relationship with the number of words. Since the crossings were highly positively correlated with the number of words produced in the task (Figure 5), we used adjusted crossings with the number of words in the logistic model. Further, returns had a very strong correlation with crossings, so we used only adjusted returns in later analyses. When adjusted, this variable had a small positive correlation with the number of words. Age did not correlate with the number of words or the num-

ber of sub-categories, but had significant negative correlations with all crossing and return variables. Thus, older participants tended to cross sub-categories less often.

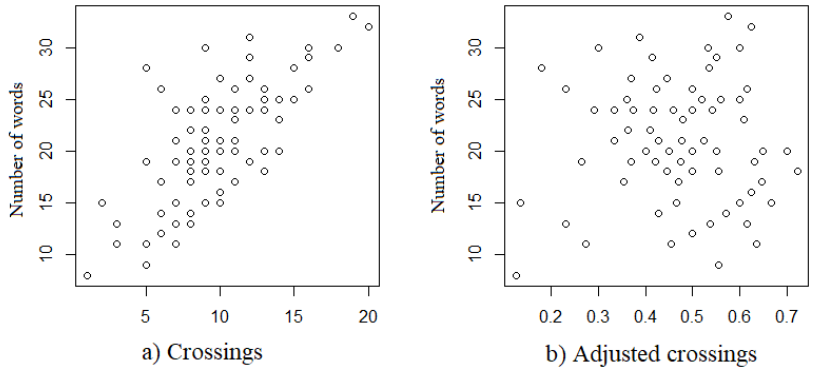

Figure 5. The correlations between the number of words and both a) non-adjusted and b) adjusted crossings as a scatterplot.

When we inspected the associations between the fluency variables within each group separately, we found that the relationship between the number of words and crossings was especially strong in aMCI patients (*Spearman* $r = 0.60$, $p = .002$; healthy: $r = 0.47$, $p = .002$; AD: $r = 0.51$, $p = .030$). Therefore, especially in aMCI patients, when the number of crossings increased, the number of produced words increased. Moreover, the linear relationship between named sub-categories and crossings was the strongest in AD patients ($r = 0.84$, $p < .001$) when compared with aMCI ($r = 0.61$, $p = .002$) and healthy controls ($r = 0.62$, $p < .001$). In AD patients, the participants that visited only a few sub-categories also crossed them less. Interestingly, in healthy participants, there was a negative correlation between the number of sub-categories and returns that were adjusted with the number of crossings ($r = -0.33$, $p = .034$). The relationship was non-significant and positive in the patient groups. Therefore, if healthy participants named more sub-categories, they were less likely to return to them, and vice versa.

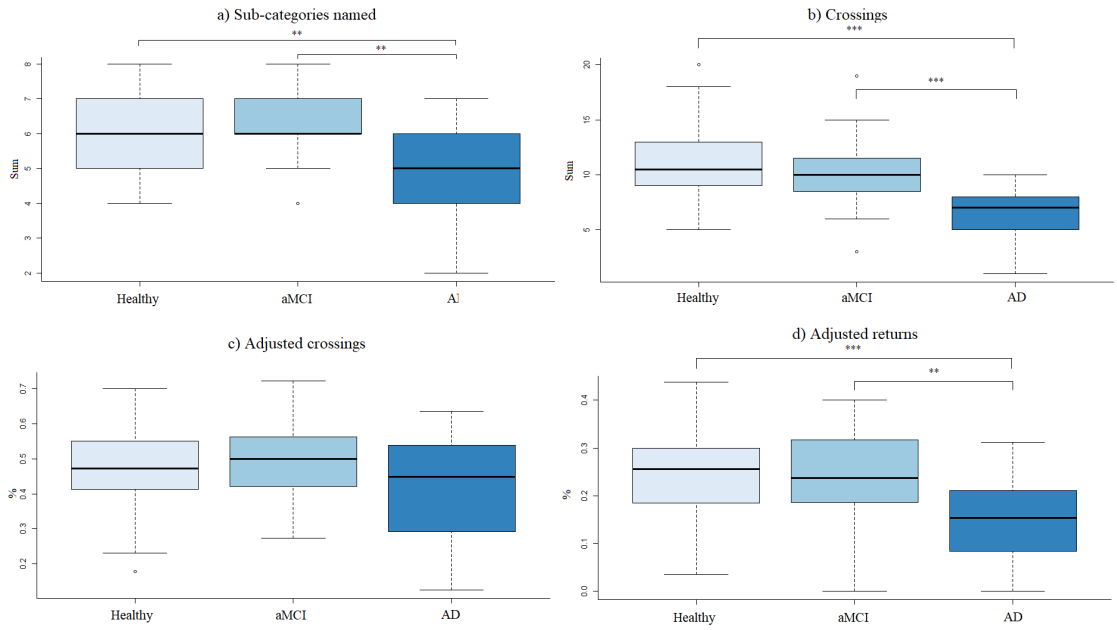

Figure 6. a) Number of sub-categories named, b) crossings, c) adjusted crossings and d) adjusted returns as means and standard errors by groups. In the boxplots, minimum, first quartile, median, third quartile and maximum are shown. The circles represent outliers. *$p < .05$, ** $p < .01$, *** $p < .001$.

In line with our hypotheses, we found that there were differences between groups in how they moved in the semantic space in one way MANOVA [F(8,156) = 6.31, $p < .001$, Wilk's $\Lambda = 0.57$, partial $\eta^2 = .25$]. We found differences in how many sub-categories each group visited during the task in one way ANOVA [F(2,81) = 5.45, $p = .006$, partial $\eta^2 = .12$]. AD patients named fewer sub-categories compared to both aMCI patients and healthy controls (Figure 6). In our data, no AD patient was able to visit all of the eight sub-categories. Consistent with previous literature, the groups also differed from each other in crossing from a sub-category to another [F(2,81) = 16.80, $p = < .001$, partial $\eta^2 = .29$], where the AD group crossed sub-categories less often than aMCI patients and healthy controls (Figure 6). Contrary to our hypothesis, we did not find a statistically significant difference between groups in the adjusted crossings [F(2,81) = 2.24, $p = .113$, partial $\eta^2 = .05$]. However, when we examined the adjusted returns variable, we found

a highly significant difference between the groups [F(2,81) = 8.69, $p < .001$, partial $\eta^2 = .18$]. In the pairwise analyses, we found that both healthy participants and aMCI patients returned to the sub-categories they had previously visited more often compared to AD patients even when the number of words produced was controlled (Figure 6).

Table 3. Multinomial logistic regression models.

Model	$\tilde{\chi}^2$	Pseudo R^{2+}	BIC	*P*-value
Model 1	46.74	0.49	153.65	<.001
Number of words	35.94			<.001
Age	9.24			.010
Model 2	51.74	0.53	157.52	<.001
Number of words	38.77			<.001
Age	8.37			.015
Adjusted crossings	5.00			.082
Model 3	47.23	0.49	162.03	<.001
Number of words	29.34			<.001
Age	8.33			.016
Sub-categories	0.49			.783
Model 4	54.78	0.55	154.48	<.001
Number of words	30.49			<.001
Age	10.27			.006
Adjusted returns	8.03			.018

$^{+}$ Nagelkerke's Pseudo R^2 is used.

The results from the multinomial logistic models are presented in Table 3. Model 1, which consisted of only the number of words and age, was statistically significant and explained half of the variation in the data. To Models 2 and 3 we added the adjusted crossings and the number of sub-categories, respectively. Even though both models themselves were statistically significant, neither the adjusted crossings or the number of sub-categories were statistically significant in the models. Therefore, these variables did not improve the models' fit to the data. Neither adjusted crossings or the number of sub-categories had a significant, independent effect in

the models when the number of words and age were controlled. However, in Model 4, adjusted returns was a statistically significant predictor even when the number of words and age were controlled. Adding adjusted returns into Model 1 improved the explanatory power by six percent.

Pairwise comparisons between groups from Model 4 are seen in Table 4. As expected, the number of words differentiated AD patients from both aMCI patients and healthy controls. When the number of words increased with one, the odds ratio for Healthy and aMCI grew 1.54 and 1.27, respectively, compared to the AD group. Age differentiated aMCI patients from AD, so that when age increased with one year, the odds ratio for being in the AD group grew 1.27. Finally, the adjusted returns differentiated healthy and aMCI from AD patients. When adjusted returns increased by one percent, the odds ratio for healthy controls and aMCI patients grew 1.13 and 1.10, respectively, compared to the AD group.

Table 4. Model 4. Multinomial logistic regression analysis on belonging to a group with adjusted returns, the number of words, and age as independent variables.

	Healthy vs. AD						aMCI vs AD					
	B	S.E	Wald	Exp(B)	95% C.I		B	S.E	Wald	Exp(B)	95% C.I	
					L	U					L	U
Intercept	-4.38	5.01	0.76				6.98	4.70	2.21			
Adjusted returns	0.12	0.05	6.72**	1.13	1.03	1.23	0.09	0.05	4.24*	1.10	1.00	1.20
Number of words	0.43	0.11	16.78***	1.54	1.25	1.90	0.24	0.19	5.86*	1.27	1.54	1.91
Age	-0.07	0.06	1.47	0.93	0.82	1.05	-0.17	0.06	7.25**	1.10	1.00	1.20

Reference category is the AD group. Adjusted returns was multiplied by hundred for easier interpretation.

$^{*}p<.05, ^{**}p<.01, ^{***}p<.001$

Finally, we inspected the classification rates of the Models 1 and 4 (Table 5). Adding the adjusted returns to Model 1 did not drastically improve the overall classification power of the model (one percent), but it did greatly improve the classification of AD patients by 16 percent. However, Model 4 was not as accurate in classifying

healthy controls and it classified more healthy controls to the aMCI patient sub-category. Yet, Model 1 mistook more AD patients as healthy controls compared to Model 4 and the AD patients were overall better classified in Model 4. Neither of the models were able to classify aMCI patients above the 0.5 random cut-off point. Therefore, we conclude that based on the present data, neither the number of words or returns were useful in categorising aMCI patients in the semantic fluency task. However, accounting for the number of adjusted returns in the task provided additional information for the distinction of the groups, and especially in distinguishing AD patients.

Table 5. Classification amounts and rates (%) for Models 1 and 4 presented as a confusion matrix.

Model 1

	Predicted			
Observed	Healthy	aMCI	AD	Correct (%)
Healthy	33	7	2	78.6
aMCI	9	10	5	41.7
AD	7	0	11	61.1
Overall (%)	58.3	20.2	21.4	64.3

Model 4

	Predicted			
Observed	Healthy	aMCI	AD	Correct (%)
Healthy	31	10	1	73.8
aMCI	11	10	3	41.7
AD	4	0	14	77.8
Overall (%)	54.8	23.8	21.4	65.5

4 Discussion

The present study used a novel dimensionality reduction method t-SNE to classify words produced inside the animal category of the semantic fluency task. This enabled us to identify eight distinct sub-categories of animals. Thus, we were able to extensively describe how both healthy controls and very early and prodromal AD patients utilised the semantic space in the animal fluency task. We found that in addition to healthy controls naming more words compared to aMCI and AD patients, the aMCI patients also produced more forest animals when the number of words was controlled. After controlling for age and number of words, the number of categories and adjusted crossings did not differentiate the patient groups from

healthy participants. However, we discovered that returning to a sub-category provided additional information besides total words named in the classification of the AD patients. Our results provide more insight in how both healthy and amnestic individuals move in the semantic space and what strategies they use in producing as many words as possible in the semantic fluency task. Further, these findings may have clinical implications in diagnosing very early AD.

4.1 Findings from the behavioural data

In line with previous behavioural studies, we found that healthy participants named more words in the semantic fluency task than aMCI and AD patients. Furthermore, healthy participants named more sub-categories and performed more crossings compared to AD patients. AMCI patients also performed better than AD patients in producing more words, sub-categories and crossings. This supports the notion that the overall performance in the semantic fluency task seems to deteriorate in the progression of the AD. However, these differences may have been due to the deterioration in the capability to produce as many words as possible.

We found that in all diagnostic groups most sub-categories were well represented in the answers, with the exception of fish and sea life. Of these sub-categories, participants often named no words or a single word. Interestingly, there were many different bird species in our data. This can be explained by the fact that in the Swiss-German area, there is a well-known children's song that lists birds, which may have affected our results. In tasks that rely on the production of words, it is important to take into account that the words produced in the task are connected to the language of the area where the study is conducted. For instance, previous studies that have found musical memory such as remembering melodies and lyrics can remain relatively intact in the progression of AD (Cuddy & Duffin, 2005), which may have affected the number of birds in our data.

When we compared the groups based on what sub-categories they named, we found that healthy participants named more jungle animals compared to aMCI and AD patients and more farm animals compared to AD patients. Further, aMCI patients named more forest animals compared to AD patients. However, after controlling for the number of words, the only significant effect between groups was that aMCI patients named more forest animals compared to both healthy and AD groups. This is a novel finding, but needs to be replicated by future studies, since it could also be a demonstration of this particular data set's features.

There was a strong positive connection between the number of words and crossings. Therefore, the more words the participant names, the more crossings they make. This supports our claim that the number of crossings mostly reflects the overall number of the produced words. Thus, studying a non-adjusted crossing variable may conceal valuable information on the crossing behaviour that is not dependent on how many words the participants produce. Similar connections were inspected between the named sub-categories and the number of words: the more words the participant produced, the more likely it was for them to name more sub-categories. Returns to a sub-category also had a strong positive connection to the number of words produced and this relationship remained statistically significant even after controlling for the number of words. However, the results of multinomial logistic model suggest that returning to a sub-category provides an individual effect over and above to the effect of number of words produced.

There were differences between groups in how the fluency variables were related to each other. The connection between the number of words and crossings was especially strong in aMCI patients, even though with both the healthy and the AD group the relationship was also positive. These findings might indicate that aMCI patients might rely more on crossing sub-categories in trying to produce as many words as possible. This could imply that according to our hypothesis, aMCI patients might be more prone to crossing sub-categories instead of naming similar

objects.

Interestingly, the number of sub-categories and the returns divided by all crossings were negatively correlated in healthy controls, but not other groups. These results suggest that healthy participants may utilize two parallel strategies in naming as many words as possible: 1) naming as many objects inside one sub-category and then moving to the next one, and 2) naming few sub-categories but crossing between these sub-categories. Since this differentiation of strategies was not apparent in patient groups, we suggest that it could be that these groups are not able to choose between strategies, but the participants that name more objects also cross sub-categories more often.

Contrary to our hypothesis, when we inspected the adjusted crossings variable, there were no statistically significant differences between groups in crossings. However, when inspecting Figure 6, the aMCI group seemed to make more crossings when the number of words was controlled for compared to the healthy participants even though this effect was not statistically significant. On the contrary, AD patients seem to name less objects compared to healthy participants. It could be that the differences between patient groups reflect differences in the stage of cortical damage in aMCI and AD. Despite the fact that the AD patients in the present study are at a very early stage of the disease, they may already present symptoms of more limited cognitive processes such as perseveration (i.e., becoming stuck on repeating words), which may outweigh the effect of having difficulties in naming semantically similar words. Therefore, we suggest that the aMCI patients might provide a more useful subgroup for studying the difficulties in naming semantically similar words.

As a novel finding, we found that AD patients returned to already visited sub-categories less compared to both aMCI patients and healthy participants. In addition, we found that only the adjusted returns provided additional information in discriminating between groups in addition to the number of words produced

and age (Table 3). Adding returns to the model enhanced the classification of the whole model, particularly in discriminating AD patients from healthy participants. Adjusted crossings and sub-categories were not significant in the models. These findings suggest that in addition to the number of words produced in the semantic fluency task, the number of sub-categories and the adjusted number of crossings do not provide additional information in discriminating AD and aMCI patients from healthy controls. However, the use of adjusted returns might aid clinicians in diagnosing early AD.

Contrary to our hypothesis, we found significant effects only in returning to a category but not crossings. These variables were highly correlated and the returns were formed as being the percentage of returns from all crossings. This suggests that crossing between sub-categories might consist of two differing aspects: 1) crossing to a new sub-category and 2) returning to a previously visited sub-category. It may be that in early AD, only the ability to return to a sub-category is declined, and not moving to new sub-categories, which is kept intact. However, when using algorithms like word2vec and t-SNE in estimating sub-categories in the semantic space, one semantic structure is imposed on all participants. Our results could also be explained by the fact that healthy participants may have more rich connections between different objects, which is why they exhibit more of the returning behaviour. Healthy participants may be able of utilising different strategies in naming objects besides just naming objects inside one biological category. For instance, moving from *duck* (bird) to *pig* (farm animal) to *pigeon* (bird) may be logical (all something you might eat) even though the individual seems to return to a sub-category. Our results may imply that the capability of using various types of connections between objects to move in the semantic space is damaged in already very early AD.

Contrary to the findings presented by Troyer et al., we did not find differences between groups in the number of sub-categories and the adjusted crossings, when

the number of words was taken into account. However, we did find differences in the adjusted returns, which is in line with overall semantic difficulties in the progression of AD previously studied. In addition, we did not find statistically significant effects indicating decline in the processing of semantically similar objects in very early AD. Yet, there was a non-significant visually inspected effect that the aMCI patients would present more adjusted crossings compared to healthy participants, while the AD patients seemed to present less adjusted crossings compared to the control group. Further, the relationship between crossings and the number of words was the strongest in aMCI patients. These factors may indicate that among aMCI patients, the amount of words is especially dependant on how many crossings the individual makes. Therefore, aMCI patients might rely more on the strategy of moving from one sub-category to another in producing as many words as possible, which may be an indicator of decline in processing similar objects, as suggested by previous neurocognitive literature (Krumm et al., 2019). For future studies, we suggest that especially the aMCI patients should be of interest in researching the processing of similar objects.

4.2 Findings from the corpus data

From the corpus data, we were able to visualise feature-based vectors of animals in a two-dimensional space with the t-SNE dimensionality reduction algorithm and form sub-categories of the animal category. The results of the t-SNE model were stable and did not change drastically with different perplexity parameters (see Appendix). Since the sub-categories created were based on the condensed feature vectors of words, they were not purely biological sub-categories. For instance, words fox and squirrel are here categorized as birds, which can be explained by their closeness to words owl and crow as these animals also often occur together in fairy tales, for example. There were some other outliers that did not clearly belong to one specific sub-category but were always situated near them. For instance,

swallow, pecker, weasel and badger seemed to cluster together and close to pets in all of the models, so they were put into the Pet sub-category in the final model. Overall, the sub-categories were remarkably logical and t-SNE seemed to work very effectively with corpus data as previously has been suggested (Maaten & Hinton, 2008).

Based on the results in the current study, we consider the t-SNE method based on the concept-feature vectors very effective on classifying objects in sub-categories in the semantic fluency task. Thus, t-SNE should be considered an alternative method to the subjective evaluation method proposed by Troyer et al. (1997). T-SNE can be used for more consistent categorization of named objects in the fluency task. Further, t-SNE provides a uniform category solution for the data of the specific task, and therefore allows us to compare crossings and the number of objects in the sub-categories. Finally, since semantic space is formed of categories based on multi-dimensional feature vectors, these categories do not follow only one logic (e.g., a fox belongs to birds). Dimensionality reduction algorithms such as t-SNE are able to capture the multidimensional nature of the semantic space and visualise it into two-dimensional space, which is not possible with a traditional subjective evaluation measure.

To our knowledge, there have not been studies that have delved into the sub-categories of the semantic fluency task in a similar scope as the present study. The results of this Master's thesis demonstrate that in addition to classic higher-level categories, such as living versus non-living, or animals, fruits, vehicles and tools, it is also possible to examine the sub-categories within the category. In the present study, this was made possible with the use of an internet-based corpus, feature vectors, and a dimensionality reduction algorithm. These results give new insights in zooming inside the higher-level categories and thus we encourage these methods to be used to gain more knowledge on how individuals utilise the semantic space in tasks such as the semantic fluency task.

4.3 Limitations

For the classification of the data, we used the t-SNE dimensionality reduction method to be able to compare crossings and the number of objects in the sub-categories. However, the method has some limitations that need to be accounted for. Since t-SNE is data-driven, its results might be difficult to replicate in other data sets. However, with large enough data sets, we can assume that most of the sub-categories we have presented in the present study are likely to replicate, if the category size is kept relatively big and not divided into small subgroups. In choosing to use bigger categories, we may lose information on individual categorisation strategies. For instance, in our data, mouse and rat belong to different sub-categories (pets and aversive animals, respectively), which could also be categorised to the same sub-category using another logic. However, our approach provides a possibility to examine how individuals move in the semantic space abiding by overall semantic structure. Another issue is that the categories produced by t-SNE require manual labeling based on visual inspection, since it is not advisable to use an actual clustering algorithm on the t-SNE results because t-SNE does not preserve distances between sub-categories or alternatively regards them meaningless (Maaten & Hinton, 2008). However, we consider that in the present study, the visual inspection was faithful to the t-SNE solution since the sub-categories were relatively well-defined.

4.4 Conclusions

In the present study, we aimed to extensively describe how patients with very early and prodromal AD perform in the semantic fluency task compared with healthy controls. Our results did not directly support the idea that there is decline in the processing of semantically similar objects in early and prodromal AD. Based on our findings, we propose that the number of sub-categories and crossings produced

by participants in the task are not meaningful in attempting to differentiate between patients and healthy controls in a clinical context, since these variables do not provide additional information over and above that information provided by the number of words the participants produce in the task. However, inspecting returns to a sub-category might provide to be useful for clinicians in diagnosing early AD. Furthermore, t-SNE provides a valuable tool for visualising the semantic space and its sub-categories which individuals seem to utilise efficiently in the semantic fluency task. We hope that these results provide insight for clinicians for the behaviour of very early and prodromal AD patients in the semantic fluency task to promote discovering these diseases at as initial a stage as possible.

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.

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., ... Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 270–279.

Almkvist, O. (1996). Neuropsychological features of early Alzheimer's disease: preclinical and clinical stages. *Acta Neurologica Scandinavica*, *94*(S165), 63–71.

American Psychiatric Association. (1994). Diagnostic and Statistical Manual of Mental Disorders. *American Psychiatric Association Press, Washington, DC*.

Amieva, H., Le Goff, M., Millet, X., Orgogozo, J. M., Pérès, K., Barberger-Gateau, P., ... Dartigues, J. F. (2008). Prodromal Alzheimer's disease: Successive emergence of the clinical symptoms. *Annals of Neurology*, *64*(5), 492–498.

Barense, M. D., Henson, R. N. A., Lee, A. C. H., & Graham, K. S. (2010). Medial temporal lobe activity during complex discrimination of faces, objects, and scenes: Effects of viewpoint. *Hippocampus*, *20*(3), 389–401.

Birn, R. M., Kenworthy, L., Case, L., Caravella, R., Jones, T. B., Bandettini, P. A., & Martin, A. (2010). Neural systems supporting lexical search guided by letter and semantic category cues: A self-paced overt response fMRI study of verbal fluency. *NeuroImage*, *49*(1), 1099–1107.

Bonner, M. F., & Price, A. R. (2013). Where Is the Anterior Temporal Lobe and What Does It Do? *Journal of Neuroscience*, *33*(10), 4213–4215.

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, *82*(4), 239–259.

Braak, H., & Braak, E. (1997). Diagnostic criteria for neuropathologic assessment of Alzheimer's disease. *Neurobiology of Aging*, *18*(4), S85–S88.

Brookmeyer, R., Johnson, E., Ziegler-Graham, K., & Arrighi, H. M. (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimer's & Dementia*, *3*(3), 186–191.

Buckley, M. J., Booth, M. C. A., Rolls, E. T., & Gaffan, D. (2001). Selective Perceptual Impairments After Perirhinal Cortex Ablation. *The Journal of Neuroscience*, *21*(24), 9824–9836.

Buja, A., Swayne, D. F., Littman, M. L., Dean, N., Hofmann, H., & Chen, L. (2008). Data Visualization With Multidimensional Scaling. *Journal of Computational and Graphical Statistics*, *17*(2), 444–472.

Clarke, A., & Tyler, L. K. (2014). Object-Specific Semantic Coding in Human Perirhinal Cortex. *The Journal of Neuroscience*, *34*(14), 4766–4775.

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., ... Haxby, J. V. (2012). The Representation of Biological Classes in the Human Brain. *Journal of Neuroscience*, *32*(8), 2608–2618.

Cuddy, L. L., & Duffin, J. (2005). Music, memory, and Alzheimer's disease: is music recognition spared in dementia, and how can it be assessed? *Medical Hypotheses*, *64*(2), 229–235.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*(1-2), 25–62.

Ellis, A. W., & Young, A. W. (2013). *Human Cognitive Neuropsychology: A Textbook with Readings*. New York: Psychology Press.

Epker, M. O., Lacritz, L. H., & Munro Cullum, C. (1999). Comparative Analysis of Qualitative Verbal Fluency Performance in Normal Elderly and Demented Populations. *Journal of Clinical and Experimental Neuropsychology*, *21*(4), 425–434.

Fagundo, A. B., López, S., Romero, M., Guarch, J., Marcos, T., & Salamero, M.

(2008). Clustering and switching in semantic fluency: predictors of the development of Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *23*(10), 1007–1013.

Fewster, P. H., Griffin-Brooks, S., MacGregor, J., Ojalvo-Rose, E., & Ball, M. J. (1991). A topographical pathway by which histopathological lesions disseminate through the brain of patients with Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *2*(3), 121–132.

Fischer, P., Jungwirth, S., Zehetmayer, S., Weissgram, S., Hoenigschnabl, S., Gelpi, E., ... Tragl, K. H. (2007). Conversion from subtypes of mild cognitive impairment to Alzheimer dementia. *Neurology*, *68*(4), 288–291.

Gainotti, G., Silveri, M. C., Daniel, A., & Giustolisi, L. (1995). Neuroanatomical correlates of category-specific semantic disorders: A critical survey. *Memory*, *3*(3-4), 247–263.

Hardy, J. A., & Higgins, G. A. (1992). Alzheimer's Disease: The Amyloid Cascade Hypothesis. *Science; Washington*, *256*(5054), 184.

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. *Neuropsychologia*, *42*(9), 1212–1222.

Hirni, D. I., Kivisaari, S. L., Monsch, A. U., & Taylor, K. I. (2013). Distinct neuroanatomical bases of episodic and semantic memory performance in Alzheimer's disease. *Neuropsychologia*, *51*(5), 930–937.

Kivisaari, S. L., Probst, A., & Taylor, K. I. (2013). The Perirhinal, Entorhinal, and Parahippocampal Cortices and Hippocampus: An Overview of Functional Anatomy and Protocol for Their Segmentation in MR Images. In S. Ulmer & O. Jansen (Eds.), *fMRI* (pp. 239–267). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kivisaari, S. L., Tyler, L. K., Monsch, A. U., & Taylor, K. I. (2012). Medial perirhinal cortex disambiguates confusable objects. *Brain*, *135*(12), 3757–3769.

Kivisaari, S. L., van Vliet, M., Hultén, A., Lindh-Knuutila, T., Faisal, A., & Salmelin, R. (2019). Reconstructing meaning from bits of information. *Nature communications*, *10*(1), 1–11.

Krumm, S., Berres, M., Kivisaari, S. L., Monsch, A. U., Reinhardt, J., Blatow, M., ... Taylor, K. I. (2019). *Cats and Apples: Performance in naming living things predicts left medial perirhinal thickness.* (Unpublished manuscript)

Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). *Neuropsychological assessment*. Oxford University Press, USA.

Libby, L. A., Ekstrom, A. D., Ragland, J. D., & Ranganath, C. (2012). Differential Connectivity of Perirhinal and Parahippocampal Cortices within Human Hippocampal Subregions Revealed by High-Resolution Functional Imaging. *Journal of Neuroscience*, *32*(19), 6550–6560.

Lonie, J. A., Herrmann, L. L., Tierney, K. M., Donaghey, C., O'Carroll, R., Lee, A., & Ebmeier, K. P. (2009). Lexical and semantic fluency discrepancy scores in aMCI and early Alzheimer's disease. *Journal of Neuropsychology*, *3*(1), 79–92.

Maaten, L. v. d., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*(Nov), 2579–2605.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., ... Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 263–269.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems 26*, 3111–3119.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in neurosciences*, *6*, 414–417.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what

you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Pekkala, S. (2004). Semantic Fluency in Mild and Moderate Alzheimer's Disease. (Doctoral thesis, University of Helsinki, Dept. of Phonetics, Helsinki, Finland)

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*(3), 183–194.

Price, S. E., Kinsella, G. J., Ong, B., Storey, E., Mullaly, E., Phillips, M., ... Perre, D. (2012). Semantic verbal fluency strategies in amnestic mild cognitive impairment. *Neuropsychology*, *26*(4), 490–497.

Quiroga, R. Q. (2012). Concept cells: the building blocks of declarative memory functions. *Nature Reviews Neuroscience*, *13*(8), 587–597.

Randall, B., Moss, H. E., Rodd, J. M., Greer, M., & Tyler, L. K. (2004). Distinctiveness and Correlation in Conceptual Structure: Behavioral and Computational Studies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 393–406.

Raoux, N., Amieva, H., Le Goff, M., Auriacombe, S., Carcaillon, L., Letenneur, L., & Dartigues, J.-F. (2008). Clustering and switching processes in semantic verbal fluency in the course of Alzheimer's disease subjects: Results from the PAQUID longitudinal study. *Cortex*, *44*(9), 1188–1196.

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., ... Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association*, *7*(3), 280–292.

Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The Medial Temporal Lobe. *Annual Review of Neuroscience*, *27*(1), 279–306.

Suzuki, W. L., & Amaral, D. G. (1994, December). Perirhinal and parahippocampal cortices of the macaque monkey: Cortical afferents. *The Journal of Comparative Neurology*, *350*(4), 497–533.

Taylor, K. I., Devereux, B. J., & Tyler, L. K. (2011). Conceptual structure: Towards an integrated neurocognitive account. *Language and Cognitive Processes*, *26*(9), 1368–1401.

Taylor, K. I., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proceedings of the National Academy of Sciences*, *103*(21), 8239–8244.

Taylor, K. I., & Probst, A. (2008). Anatomic localization of the transentorhinal region of the perirhinal cortex. *Neurobiology of Aging*, *29*(10), 1591–1596.

Terry, R. D., & Katzman, R. K. (1983). Senile dementia of the Alzheimer type. *Annals of Neurology*, *14*(5), 497–506.

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.

Troyer, A. K., Moscovitch, M., Winocur, G., Leach, L., & Freedman, M. (1998). Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease. *Journal of the International Neuropsychological Society*, *4*(2), 137–143.

Tyler, L. K., Chiu, S., Zhuang, J., Randall, B., Devereux, B. J., Wright, P., ... Taylor, K. I. (2013). Objects and categories: Feature statistics and object processing in the ventral stream. *Journal of cognitive neuroscience*, *25*(10), 1723–1735.

Wimo, A., Jonsson, L., & Winblad, B. (2006). An Estimate of the Worldwide Prevalence and Direct Costs of Dementia in 2003. *Dementia and Geriatric Cognitive Disorders*, *21*(3), 175–181.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.-O., ... Petersen, R. C. (2004). Mild cognitive impairment – beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*, *256*(3), 240–246.

Appendix: Alternative t-SNE Models

Figure A1. Perplexity = 10

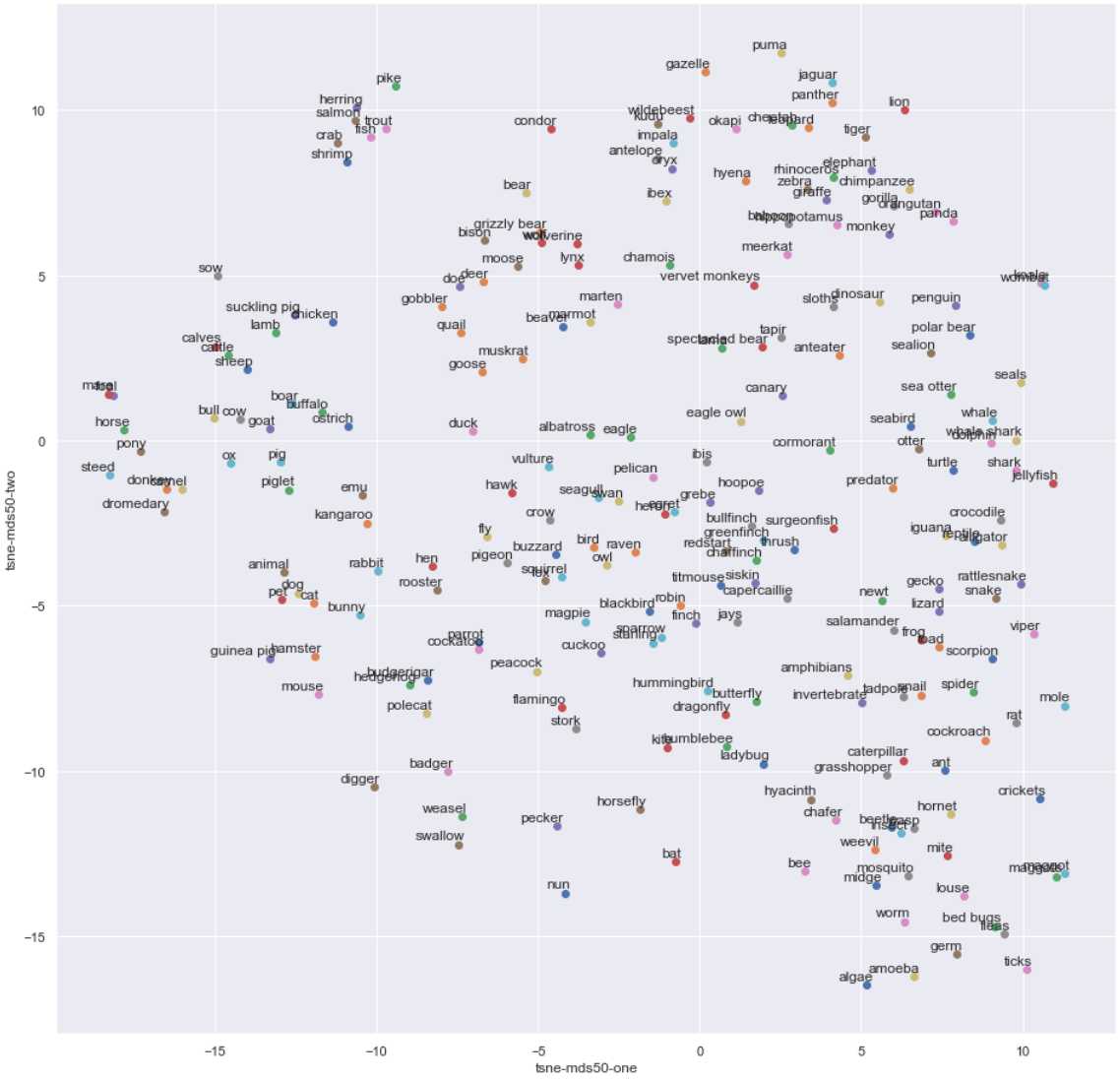

Figure A2. Perplexity = 25

Figure A3. Perplexity = 40