Department of Neuroscience and Biomedical Engineering

Processing of subword information in the human brain

Tero Hakala

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 7/2020

Processing of subword information in the human brain

Tero Hakala

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination at the Auditorium F239a, Otakaari 3A, on 24 January 2020 at 12 noon

Aalto University
School of Science
Department of Neuroscience and Biomedical Engineering
Imaging Language

Supervising professor
Professor Riitta Salmelin, Aalto University, Finland

Thesis advisors
Dr. Annika Hultén, Aalto University, Finland
Dr. Minna Lehtonen, University of Oslo, Norway

Preliminary examiners
Professor Piers Cornelissen, Northumbria University Newcastle, UK
Professor Teija Kujala, University of Helsinki, Finland

Opponent
Professor Alec Marantz, New York University, USA

Aalto University publication series
DOCTORAL DISSERTATIONS 7/2020

ISBN 978-952-60-8911-9 (printed)
ISBN 978-952-60-8912-6 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-8912-6

Images: Illustrations and cover art, Tero Hakala

Unigrafia Oy
Helsinki 2020

Finland

Printed matter
4041-0619

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Tero Hakala

Name of the doctoral dissertation
Processing of subword information in the human brain

Publisher School of Science

Unit Department of Neuroscience and Biomedical Engineering

Series Aalto University publication series DOCTORAL DISSERTATIONS 7/2020

Field of research Systems neuroscience

Manuscript submitted 8 December 2019 **Date of the defence** 24 January 2020

Permission for public defence granted (date) 17 December 2019 **Language** English

☐ **Monograph** ☒ **Article dissertation** ☐ **Essay dissertation**

Abstract

In agglutinative languages, such as Finnish, a single word can have a large number of possible inflected and derived forms. It is necessary for the human brain to recognize regularities in the subword structures. In study IV it was observed that the brain responses to linguistic stimuli are related to fine-grained predictions of the language input at least at the syllable level. Studies I-III tested quantitative models for describing the relationship between subword structure and the responses related to human word processing.

Statistical machine-learning models developed for automated applications in Natural Language Processing have proven useful for describing morphological regularities in languages. In this thesis, these models are applied to human word processing.

Visual word recognition evokes a distinct pattern of neural responses that can be functionally, temporally and spatially separated using magnetoencephalography (MEG). In study I these responses were linked to language models describing different levels of linguistic abstraction. The early occipital and occipito-temporal responses could be modeled using visual and orthographic features, whereas the responses in the bilateral temporal areas were best described by models that represented words as compositions of morphemic units or as whole words.

In the statistical model of morphology used in these studies, the subword structure emerges from optimization of information representation. The structure is determined by the cost of storing distinct morphemic units and the cost of combining them. Study III found that the bestperforming model for describing eye-movements used compositions of morphemic segments to represent many, but not all, complex words. Many words were also kept intact. The optimal morphemes were generally more coarse-grained than those implicated by linguistic analysis. In Study II, the morphemes from the optimal statistical model were compared to linguistic morphemes in a neural decoding task in which words were identified from the cortical responses. Both statistically and linguistically structured models were successful in the decoding task.

The results of this thesis suggest that the neural responses to words are related to word representation by compositions of morphemic units. The units may not be strictly linguistically determined; instead, the word structures can reflect the statistical regularities of language environment. This thesis demonstrates that quantitative modeling of cortical responses is useful for describing even relatively abstract linguistic phenomena such as morphology.

Keywords MEG, morphology, computational linguistics, Morfessor, word recognition

ISBN (printed) 978-952-60-8911-9 **ISBN (pdf)** 978-952-60-8912-6

ISSN (printed) 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2020

Pages 132 **urn** http://urn.fi/URN:ISBN:978-952-60-8912-6

Tekijä
Tero Hakala

Väitöskirjan nimi
Sanan osien käsittely aivoissa

Julkaisija Perustieteiden korkeakoulu

Yksikkö Neurotieteen ja lääketieteellisen tekniikan laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 7/2020

Tutkimusala Systeeminen neurotiede

Käsikirjoituksen pvm 08.12.2019 **Väitöspäivä** 24.01.2020

Väittelyluvan myöntämispäivä 17.12.2019 **Kieli** Englanti

☐ **Monografia** ☒ **Artikkeliväitöskirja** ☐ **Esseeväitöskirja**

Tiivistelmä

Suomen kielen kaltaisissa agglutinatiivisissa kielissä voi sanan kantaan liittää erilaisia päätteitä kuten taivutuksia ja johdoksia. Aivot hyödyntävät sanan rakenteen säännönmukaisuuksia. Esimerkiksi tämän väitöskirjan tutkimuksen IV tulokset viittaavat siihen, että aivot pyrkivät ennustamaan kielellistä ärsykettä hienojakoisesti ainakin tavujen tasolla. Tutkimuksissa I-III selvitettiin, miten yhteyttä sanan osien ja ihmisen sanankäsittelyä kuvaavien vasteiden välillä voidaan mallintaa kvantitatiivisesti.

Luonnollisen kielen automaattiseen prosessointiin liittyviin teknisiin sovelluksiin on kehitetty tilastollisia koneoppimiseen pohjautuvia malleja, joissa sanojen rakenteen säännönmukaisuus eli morfologia opitaan ilman kieliopillista informaatiota. Tässä väitöskirjassa näitä malleja sovelletaan kuvaamaan ihmisen sanankäsittelyä.

Kirjoitettujen sanojen tunnistukseen liittyy sarja aivotason vasteita, jotka voidaan erottaa toisistaan ajan, paikan ja toiminnallisuuden suhteen magnetoenkefalografialla. Tutkimuksessa I vertailimme näitä vasteita kuvauksiin, jotka heijastivat informaation eri abstraktiotasoja. Varhaiset takaraivolohkon vasteet pystyttiin ennustamaan ärsykkeen visuaalisten ja ortografisten piirteiden avulla, mutta myöhemmät ohimolohkoilla havaitut vasteet selittyivät malleilla, jotka esittivät sanat kokonaisina tai kokoelmana erillisiä osia eli morfeemeja.

Tässä tutkimuksessa käytetyssä morfologian mallissa sanojen pilkkominen erillisiin morfeemeihin perustuu siihen, kuinka morfeemien muistamiseen ja toisaalta niiden yhdistämiseen liittyvät kustannukset optimoidaan ja kuinka niitä painotetaan. Tutkimuksessa III ihmisten silmänliikkeitä kuvasi parhaiten malli, joissa osa monimutkaisista sanoista jaettiin osiin mutta joissa monet esitettiin myös jakamattomina. Optimaalisessa mallissa morfeemien pituus oli keskimäärin suurempi kuin kieliopillisesti määriteltyjen morfeemien. Tutkimuksessa II vertailtiin tilastollisiin morfeemiyksiköihin ja toisaalta kieliopillisiin morfeemeihin pohjautuvia malleja tehtävässä, jossa aivovasteen perusteella pyrittiin ennustamaan niihin liittyvä sana. Sekä kieliopilliset että tilastolliset morfeemit mahdollistivat sanan määrittämisen aivovasteen perusteella.

Tutkimusten tulokset viittaavat siihen, että aivot hyödyntävät sanojen käsittelyssä morfeemien kaltaisia yksiköitä, mutta nämä yksiköt eivät määräydy tai perustu yksinomaan kieliopillisiin sääntöihin, vaan ne opitaan tilastollisesti kieliympäristöstä.

Aivojen toiminnan kartoittamisessa pyritään käyttämään jatkuvasti enemmän eksplisiittisiä, määrällisiä malleja. Tämä väitöskirja toimii esimerkkinä siitä, miten myös verrattain korkean abstraktiotason kielellinen ilmiö, morfologia, voidaan yhdistää matemaattisesti hermostollisiin vasteisiin.

Avainsanat MEG, morfologia, laskennallinen lingvistiikka, Morfessor, sanantunnistus

ISBN (painettu) 978-952-60-8911-9 **ISBN (pdf)** 978-952-60-8912-6

ISSN (painettu) 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Helsinki **Painopaikka** Helsinki **Vuosi** 2020

Sivumäärä 132 **urn** http://urn.fi/URN:ISBN:978-952-60-8912-6

Acknowledgements

Science is a collaborative effort - without the support of many amazing people sharing the passion this thesis would not have been possible. First, I would like to thank my supervising professor Riitta Salmelin for her brilliant guidance, patience and tireless effort. Your comments were always encouraging, always constructive and it was a privilege working in your group. Thank you for not giving up on me.

I could not have hoped for better advisors, Annika Hultén and Minna Lehtonen. Thank you, Annika, for everything. You always believed in me and our work, even in some moments when I had lost the plot, you found a way to point me to the correct direction: Onwards and Upwards. Your positivity, energy and enthusiasm for life was indispensable. And thank you Minna for all those hour-long skype conferences brainstorming ideas and possibilities among other things.

This work was carried out in the department of neuroscience and biomedical engineering with resources from Aalto Neuroimaging infrastructure and support from the doctoral program Brain and Mind. I thank everyone involved for providing such a professional, yet relaxed working environment. I thank Mia Ilman and Marita Kattelus for the help in conducting the experiments, and all my co-authors for their efforts. Special thanks to Tiina Lindh-Knuutila and Marijn van Vliet for their help with the language models. I'm also grateful to all the other members of the Imaging language group: Anna-Mari, Heidi, Sasu, Hanna, Mia, Jan, Timo, Sasa, Lotta, Silvia, Ali, and others. Thank you for making our time in the offices so enjoyable.

From the University of Helsinki, I would like to thank Sari Ylinen for getting me involved with neuroscience of language. I also thank the wonderful people of the Cognitive Science unit, Alina Leminen, Otto Lappi and others who making it possible for me to enter this fascinating field of study.

I would also like to thank my parents, Riitta and Oiva, for always supporting me and encouraging my pursuit of higher education. Thanks to Birgit and Jari and Jarmo and Tuula for all kinds of help whenever needed and for getting me involved with the coolest hobbies. And of course, big thanks to my dear friends, Elle, Jani, Janne, Juha & Seidi, Marko, Mikko, and all the other weird and wonderful hippies.

Finally, I'm very happy with the Finnish system of education that did not force me into a mold and allowed my not-so-straightforward journey across widely different fields of study to feed my curiosity and to learn how the world works.

Helsinki, December 2019
Tero Hakala

Contents

List of Abbreviations and Symbols

ECD	Equivalent current dipole
EEG	Electroencephalograpy
EOG	Electro-oculography
ERF	Event related field
ERP	Event related potential
ET	Eye tracking
IR	Infrared
MEG	Magnetoencephalography
(f)MRI	(Functional) Magnetic resonance imaging
NDR	Naïve discriminative reader
NLP	Natural language processing
PSP	Postsynaptic potential
SQUID	Superconducting quantum interference device
tSSS	Spatiotemporal signal space separation
STG	Superior temporal gyrus
VWF	Visual word form area

List of Publications

This doctoral dissertation consists of a summary and of the following publications which are referred to in the text by their numerals

1. Hakala, T., Hultén, A., Lehtonen, M., Lagus, K., Salmelin, R. (2018) Information properties of morphologically complex words modulate brain activity during word reading, Human Brain Mapping, 39, 2583-2595

2. Hakala, T., Hulten, A., Lehtonen, M., Lindh-Knuutila, T., Salmelin, R. Learned morphemic representations successfully decode brain activity to written words. Under revision.

3. Lehtonen, M., Varjokallio, M., Kivikari, H., Hultén, A., Virpioja, S., Hakala, T., Kurimo, M., Lagus, K., Salmelin, R. (2019) Statistical models of morphology predict eye-tracking measures during visual word recognition. Memory & Cognition, 1-25

4. Ylinen, S., Nora, A., Leminen, A., Hakala, T., Huotilainen, M. Shtyrov, Y. Mäkelä, J.P., Service, E. (2014) Two distinct auditory-motor circuits for monitoring speech production as revealed by content-specific suppression of auditory cortex. Cerebral Cortex, 25, 1576-1586

Author's Contribution

Publication 1: Information properties of morphologically complex words modulate brain activity during word reading

I was the principal author and had the main responsibility for designing and conducting experiments, analysing the data and writing the manuscript

Publication 2: Learned morphemic representations successfully encode brain responses to written words

Same as Study 1

Publication 3: Statistical models of morphology predict eye-tracking measures during visual word recognition

I participated in the data analysis and in writing the manuscript

Publication 4: Two distinct auditory-motor circuits for monitoring speech production as revealed by content-specific suppression of auditory cortex

I implemented the experiment and participated in writing the manuscript

1. Introduction

Visual representation of language by written symbols was a remarkable innovation that allowed humans to circumvent unreliable memory and construct stories that transcend space and time. It literally makes this thesis possible. Even when speech and video can be recorded and transmitted with relative ease, the written word has properties that cannot be easily replicated. In recent times, books may have somewhat fallen out of favour, and the physical substrate for text may have metamorphosed from paper into screens, but writing in the form of messaging, news media, advertisements, etc., saturates the visual landscape. Cellphones may have made it possible to speak with people at distance wherever and whenever, but people at distance do not always appreciate this. Therefore, the preferred method of communication has largely shifted from voice to some form of text messaging due to its time-independent nature (Pinchot, Douglas, Paullet, & Rota, 2012).

In the timescale of civilization, written language is a relatively recent invention, and widespread literacy is an even more modern phenomenon. Unlike speech, reading and writing abilities have not had time to apply evolutionary pressure to develop distinct biological capabilities (Dehaene, 2009). Children of illiterate parents can learn to read as well as anybody, suggesting that epigenetics does not play a major role in skill acquisition either. However, the increased dependence on text messaging in the modern dating culture might result in future generations with extraordinarily flexible thumbs. Without an evolutionarily moulded reading organ, it is especially remarkable that an accomplished reader can convert text into thought with efficiency, to the point that it is nearly impossible to look at a common word and not immediately see its meaning. Sometimes we can even look at a word form that we have never encountered but still understand it. What kind of mechanisms allow such a rapid word recognition? How is it implemented in the brain? How should we go about describing it?

The study of language processing is a vast interdisciplinary project. The field of cognitive science draws elements from linguistics, psychology, neurobiology, computer science, physics, philosophy and others. In this thesis, I investigate brain processes associated with recognition of words and word parts, by combining time-sensitive measures of cortical activation with computational models developed originally for applications in Natural Language Processing (NLP).

1.1 The science of language processing

Language ability can be studied from several angles. Psycholinguistic experiments are used to measure human reaction times and accuracies in linguistic tasks or more automatic responses such as eye movements. Neurolinguistics refers to the use of functional brain monitoring methods to study language processing as it unfolds on the neural level.

Models of language processing attempt to link linguistic concepts to properties of cognition and neural systems. These models could, for example, resolve whether and what type of constraints of grammar are hardcoded in the architecture of the human brain, as proposed by the proponents of innate language ability (Chomsky, 2014; Pinker, 2003). A linguistic theory can provide hypotheses for empirical studies, as well as conceptual structure and interpretations for neuroscientific results.

However, it is not entirely obvious how the different fields of study conform. Generative grammar and the symbol manipulation approach have been a cornerstone of the linguistic theory (Chomsky, 2014), but it is not guaranteed that the linguistic rules and constructs should correspond to distinct functional structures that can be observed in neurological architectures.

Linguistics and neuroscience deal with fundamentally different types of concepts. Linguistic constructs describe the language itself and the detailed rules operating on objects such as words, sentences or morphemes. Neuroscience describes the physical behaviour of cell assemblies using concepts like activation, potentiation and oscillation. One way to state the discrepancy is by Poeppel & Embick (2017) who named two endemic difficulties:

> Granularity Mismatch Problem (GMP): Linguistic and neuroscientific studies of language operate with objects of different granularity. In particular, linguistic computation involves a number of fine-grained distinctions and explicit computational operations. Neuroscientific approaches to language operate in terms of broader conceptual distinctions.
>
> Ontological Incommensurability Problem (OIP): The units of linguistic computation and the units of neurological computation are incommensurable

In other words, linguistic theory provides such a detailed description that it is unlikely that we can observe corresponding detail with our current, rather crude imaging methods. Second, even if we could measure the brain with a very high precision, it would be challenging to identify a correspondence between a given linguistic rule and communication between neurons. The problem is akin to trying to debug a computer program by measuring the physical properties of a processor, which is not feasible even with full knowledge of microscopic details (Jonas & Kording, 2017). Indeed, connecting physical implementation to algorithmic level of analysis is a fundamental difficulty in all neuroscience, and the

problem is especially pronounced when the algorithms in questions are very detailed.

However, regardless of these challenges, linguistic structure provides at least a starting point. In order to study the reading process and word recognition, one can start from the overall framework of what a word is and what it is made of. Further along the line, we can ask how these abstract properties might be linked to physical observables. In linguistics, a word is the smallest element of language that has a practical meaning in isolation, however, a word itself may be composed of smaller building blocks called morphemes.

1.2 Morphology

The linguistic domain of word structures is called morphology. Words can be categorized into monomorphemic words (simplex) and multimorphemic words (complex). A complex word such as 'un-think-able-s' consists of multiple distinct parts called morphemes. Morphemes are considered the minimal meaningful units in a language, thus building blocks that are more basic than words. Some morphemes are bound, i.e., they must always be combined with another morpheme (e.g. 'un- '/'-s'/). A morpheme that can stand alone as a word (e.g., 'think') is called a free morpheme.

Morphology describes regularity in the otherwise arbitrary relationship between word forms and meaning. For example, English plural suffix '-s' placed at the end of a noun marks the plural. When we learn that a novel item is called a 'wug', we can deduce that the word 'wugs' refers to many such items (Pinker, 2003).

The study of cognitive and neurological foundations of morphology is a very active area of neurolinguistics. A key question is whether identification and production of complex words is a generative process in which rules or subword units are used to generate the complex word form. Alternatively, every word form could be stored as a distinct memory engram. These alternatives pose different demands in terms of computation and storage, and it is conceivable that the neural system could also seek to balance these two requirements (Bertram, Schreuder, & Baayen, 2000). Furthermore, if complex word forms are generated 'online' by some type of computational process, it may be possible to track these processes in the neural machinery of language and perhaps build models that describe the neural computations.

The 'wugs' example shows that we make use of rules when faced with a novel word, but how about those complex words that are frequently used? Is there a qualitative difference on how bound and free morphemes are represented on the neural level? Are the linguistic categories of inflections, derivations and compounds reflected in the neural implementation? These questions exist at an interesting intersection where relatively high-level linguistic concepts can provide insights for empirical neuroscience and synergy might be found between the two fields.

There has been a significant amount of research on the topic, and experimental data and theories abound, but we are still lacking a widely accepted consensus on many of these questions. The apparent lack of clarity has been attributed, e.g., to somewhat conflicting pieces of evidence and to the confusing zoo of models that keeps expanding, whereas models are rarely falsified (Amenta & Crepaldi, 2012). The lack of progress may also stem from the fundamental difficulties of bridging the conceptual gap between linguistics and neuroscience that was noted in the previous chapter.

1.2.1 Psycholinguistic models of morphology

Models of morphological processing attempt to describe a general system or subsystem of the human language faculty that coincides with the linguistic notion of morphology. The problem has attracted much interest, and numerous models have been proposed. Since there are so many of them, it has become customary to divide them into categories. A division by the model type distinguishes between classical computational models and connectionist models. This division reflects a more general discussion in the cognitive sciences about ways to describe cognition and the nature of representations and computations (Fodor & Pylyshyn, 1988; Sloman, 1996).

The classical approach describes language processing on an abstract level without recourse to the details of implementation. This paradigm proposes a set of symbols and formal rules that operate on the symbols. A connectionist approach, in turn, assumes that linguistic constructs emerge from the properties of neural networks and cannot be abstracted. Models of this type comprise a network of nodes that represent inputs, outputs, and hidden features, as well as weighted associations between the nodes. The processing is described by associations between the input and output.

The difference between computationalism and connectionism is the approach to description, not a fundamental ontological difference. Any symbol posited by a classical approach would still be encoded in the brain by some type of pattern of neural organization. It is possible to construct neural networks that perform classical computation (Siegelmann & Sontag, 1991). On the other hand, all connectionist neural networks can be simulated symbolically with classical computers and therefore abstracted as Turing machines, although the description might be rather obscure. There is also nothing that forbids hybrid models with features from both approaches (A. Graves, Wayne, & Danihelka, 2014).

Historically, the first cognitive models of morphology were classical computational models envisioned at the time that was marked by the advent of computers, and thus were undoubtedly influenced by them. These models are information processing models that, on the abstract level, resemble how one would try to implement a computer program which is sensitive to morphological features. These models posit distinct processing stages, much like subroutines in a programming language. These models can be related to human word processing by associating the stages with processing time. An influential model of word recognition by Taft and Foster (Taft & Forster, 1975) is illustrated in Fig. 1. This model proposes that processing of an affixed word involves distinct steps. The

difference in the processing steps between affixed and non-affixed words would explain why the observed recognition times are slower for affixed words. This type of models could also describe selective damages to language ability if the damaged functions can be associated with distinct stages of the model.

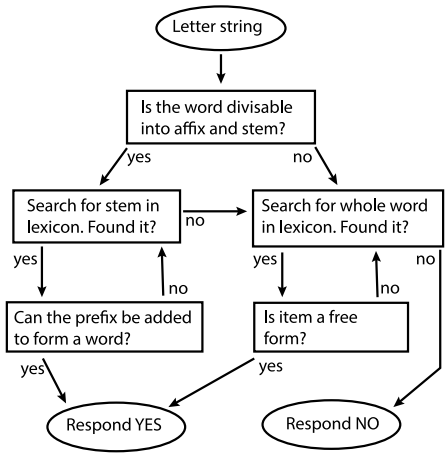

Figure 1. A Box-and-arrow model of visual word recognition by Taft & Foster, 1975.

There is a whole family of these box-and-arrow models that differ slightly on the stages and on the linguistic features that affect the processing of those stages. However, comparison between the models can be difficult if they do not provide quantitative predictions. While in Tafts's model all affixed words are decomposed, many recent models propose some type of hybrid or a dual route architecture that allows some affixed words to be processed as full-form. The idea in a dual architecture is that full word processing is supposedly computationally simpler and therefore faster. However, due to additional cost of storage, it is only reasonable for those affixed word forms that are relatively frequent (Baayen & Schreuder, 1999; Caramazza, Laudanna, & Romani, 1988; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Frauenfelder & Schreuder, 1992).

An example of a connectionist model is the Naïve Discriminative Reader (NDR) (Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011; Baayen & Smolka, 2019). This model proposes a layered network of nodes and weighted associations between the nodes, illustrated in Fig. 2. The input layer describes the word as a set of letter bigrams that act as visual cues. The cues are associated with the second layer of the network that consists of nodes representing abstract word-like entities called lexemes (in practice, these can be, e.g., word lemma forms). The lexemes are associated with each other by weighted connections that represent linguistic and semantic relationships between them.

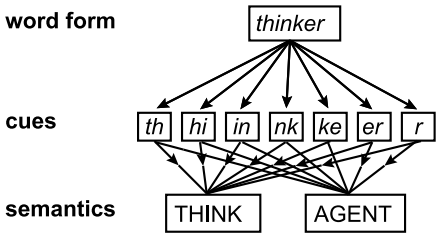

Figure 2. A connectionist model of visual word recognition, Naïve Discriminatory Reader by Baayen et al., 2011. Orthographic cues consisting of letters and letter-bigrams are extracted from the word form and associated with abstract word units.

A network model can learn associations from data and can work without a priori linguistical structures. The connection weights can be turned into quantitative predictions of e.g., processing times. The models may also have some correspondence with a neural implementation. For example, on a neural level, the cues on the first layer of NDR could correspond to features extracted by early stages of the visual processing stream. However, the weights are rather hard to interpret in a linguistic framework.

A further, third approach makes use of the concepts of information and computational complexity that dictate fundamental limitations for computational and representational systems operating on abstract symbols (Papadimitriou, 1994). These concepts can be used to make quantitative predictions about the operation of such a system, in this case, the brain. This approach falls in the domain of classical models, as it does not deal with networks. Yet, models based on this principle are quite distinct from examples such as that of Taft (1975). These models are interesting because they provide predictions for quantities that are conceptually compatible with properties of physical systems such as energy usage, while retaining some form of symbolic features that are compatible with linguistics. This idea is explored in more detail in the following chapter.

1.3 Words and information

Good words are worth much, cost little
(George Herbert, 1651)

Words contain information. This is true in many ways, but the mathematical definition that originates from the field of information theory proves particularly useful. In the following, we define information of a word in that sense.

Information resolves uncertainty. This statement was formalized mathematically by Shannon, who associated information with the effort needed to communicate the occurrence of a probabilistic event (Shannon, 1948). For an event with probability p, the amount of information is,

$$I = -log_2(p) \quad (1)$$

This quantity is also called surprisal (it quantifies how surprising an event is) and its unit is bit. This is also the minimum message length required to represent the event. Surprisal associated with a rare event is greater than that of a common event. For example, tossing a coin yields tails with the probability of 1/2, while getting a 6 with a roll of dice has a probability of 1/6. The result of a coin toss can be communicated with one bit, $I = -log_2\left(\frac{1}{2}\right) = 1$, while the result of a roll of dice requires $I = -log_2\left(\frac{1}{6}\right) = 2.58$ bits. On average, the more possible outcomes there are, the more effort it takes to communicate which outcome occurred. The expected amount of information, the entropy,

$$S = -\sum_i p_i \, log_2(p) \quad (2)$$

depends also on the shape of the distribution, with more equal distribution yielding higher entropy (Pathria & Beale, 2011).

To calculate the surprisal of a word, one needs to treat the occurrence of a given word from a set of possible words as a probabilistic event. A statistical language model is a model that defines some reasonable way to calculate this probability, i.e., it defines a probability distribution over words. We could, for example, define a crude language model which assumes that a word has 10 characters, and each character occurs at an equal probability. Then there are 27^10 possible words, each with a surprisal of $log_2(1/27\text{^}10) = 48$ bits. Thus, it requires at least 48 bits to represent a word in the context of this model. However, this is not a very good model for a natural language.

We may devise a language model with different assumptions. Another way would be to consider a large dictionary, say >1 million words, and assume the word is picked at random. Any given word can now be specified with just $-log_2(1/10\text{^}6) = 20$ bits. Of course, in this model, one can represent only those words that occur in the dictionary.

Not all words occur at equal probability. In most (if not all) natural languages, the frequency of words is distributed according to Zipf's law (Zipf, 1935). Zipf noticed that there are a small number of words with relatively high frequencies and a long tail of low-frequency words. It follows that some words are less surprising, on average, and they can be represented with a smaller number of bits in an optimized encoding. Observed word lengths in natural languages seem to respect this kind of optimization of representation: the encoding length (i.e., the word length) is proportional to the probability of the word, thus common words tend to be short, and those long words that are used often tend to get truncated in common use. This idea is generalized in the principle of least effort that states that people use the least possible effort to communicate a concept (Zipf, 2016).

This optimization of representation is independent of the implementation. Words can be represented with written letters, acoustic utterances, magnetic moments of molecules, neural spikes etc. Independent of the medium, less surprising words can be represented with less effort than the surprising ones. The assumption that the system prefers to optimize the overall long-term cost, i.e., minimize the energy, has been suggested as a means for understanding many aspects of brain organization (Friston & Stephan, 2007).

In the study of human language processing we can use this idea by considering the neural pathways as an information channel. To communicate a word, the neurons must represent the word somehow. If we assume that the neural system is optimized to represent words, the average representation effort over all possible words is minimized, and the encoding cost can be described by a language model that approximates the real language environment. The neural system is therefore associated with a model of words. The better this model corresponds to the actual language use the less effort is needed. This general optimization principle might also explain why the observed speech information rate across different languages is similar (Pellegrino, Coupé, & Marsico, 2011).

How might we estimate whether a neural system is associated with a model that is relevant to coding of words? First, we guess a model. Second, we use this model to calculate the surprisal value for a number of words. Third, we compare the results of our calculation with a measure of neural activity. If the model-derived and neural values do not correlate, then the initial guess is wrong. This may happen if 1) the model is a poor description for that neural system, 2) the assumption of optimization is wrong, or 3) the measure of activity, such as the amplitude of the MEG response, does not relate to effort and computational properties of a neural population.

For 3) there is half a century of experiments with MEG that have produced applicable results in a wide range of topics (Supek & Aine, 2016). As regards 2), the hypothesis of optimal coding is a strong candidate for the overall theory of organization of brain functions (Friston, 2010, 2012). Measures of surprisal have been successful in predicting neural responses in language tasks (Brennan, 2016; Frank, Otten, Galli, & Vigliocco, 2015; Henderson, Choi, Lowder, & Ferreira, 2016).

1.3.1 Information in morphological processing

Information theory has been used in the study of morphological processing (del Prado Martin, Kostić, & Baayen, 2004; Kostic, 2013). There are various ways to incorporate morphological properties in the probability distribution over words. One approach is to consider the word as a member of a set of morphologically related words and take into account the information associated also with the other members of that set.

For example, (del Prado Martin et al., 2004) calculate the entropy of a set of words that includes all morphological relatives. Derivations and inflections are considered hierarchically (Fig. 3) so that each derived word form (including the base form) is associated with a set of inflections, termed inflectional paradigm.

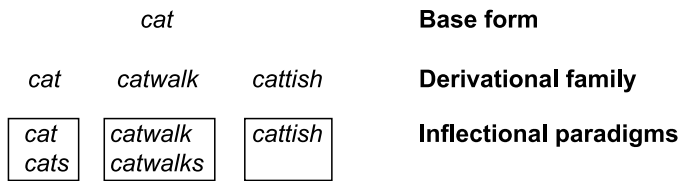

Figure 3. Hierarchical model of morphological structure

Given a word form, the inflectional entropy is calculated over a set of inflected word forms of that derivation,

$$H(P) = -\sum_{x \in P} p(x|P)\, log_2\, \mathrm{p(x|P)} \qquad (3)$$

where P is the inflectional paradigm (e.g., P:={cat, cats}), and p(x|P) is the conditional probability of encountering the word x given that the word is one of those in P. The total paradigmatic entropy is then obtained by summing the inflectional entropies associated with all derivations $H_{tot}(\mathrm{w}) = H(P_1) + H(P_2) +$ …. Finally, the word surprisal, $I(w) = -log_2\, p(w)$, and its paradigmatic entropy are summed together to obtain a quantity which the authors call the information residual, $I_R(w) = I(w) - H_{tot}(\mathrm{w})$. del Prado Martin and collegues then model recognition times for Dutch words in three separate datasets. By applying linear regressions, they show that a model using the information residual as the single predictor generally outperforms a model that uses a combination of three traditional variables associated with morphology (surface frequency, morphological family size, and cumulative root frequency).

The information residual thus considers the surprisal from word frequency and the information over all morphological relatives. In simplified terms, the word should be easier to process if it has a lot of relatives (inflections, derivations, and inflected derivations) that occur frequently. Essentially, morphological regularity in language tends to reduce the surprisal of the word compared to what is suggested by its surface frequency alone. This type of approach to incorporate morphological properties in the surprisal measures uses linguistic theory

to define which words belong to a given morphological paradigm. Thus, the details and idiosyncrasies of linguistic analysis are intertwined in the model. There are practical limitations in implementing this method for languages for which there are no comprehensive databases on morphological relations.

Information theory can also be used to define morphological properties themselves, instead of just using the morphological structure to calculate the surprisal of words: we can directly search for morphological structure that would minimize surprisal. Remarkably, such an approach can produce morphological features that resemble linguistically defined structures so well that the principle is used in NLP to automate morphological analysis (Creutz & Lagus, 2002). In NLP, a morphological segmentation is essential to reduce the lexicon size for example in speech recognition tasks in highly inflective languages.

The particular model employed in this thesis, that applies this idea of minimization of surprisal or, more precisely, the minimum description length principle (Rissanen, 1978), is the baseline variant of the Morfessor model (Creutz & Lagus, 2007). In this model, words are assumed to be composed by concatenation of morphemic units, e.g., 'think' + 'er'. The model learns a lexicon of morphemes, and words are represented by references to morphemic units (Fig. 4). Words sharing one or more units implicates a morphological relationship between those words.

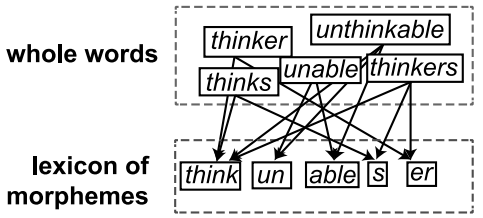

Figure 4. Morfessor seeks an optimal lexicon of morphemes. Words are represented by reference to units in the morpheme lexicon.

Here, the morphemes need not be defined a priori, instead, they are learned from data during the model training. Morfessor seeks to determine a set of morphemic units that minimize the average surprisal of all words in the corpus, while trying to keep the set of morphemes small.

Suppose M is a lexicon of morphemes m_i, with frequency-based probabilities in p_i. A word is composed of a sequence of morphemes, say $m_1\ m_2\ m_3$. In the simplest case, the units are independent of each other. The probability of the word is then $p_1 p_2\ p_3$, and the surprisal is $I = -\sum_{i=1..3} log_2\ p_i$. If there are multiple possible segmentations for the word, using some other morphemes, the segmentation with lowest surprisal is chosen.

The cost associated with a corpus is the sum of surprisals of all words in the corpus, whereas the cost of the lexicon is the total number of characters in the morphemic lexicon. Given a learning corpus, the cost function that must be minimized is the sum of these two costs, Cost(corpus) + Cost(lexicon).

The problem of finding the best possible segmentation and composition of the morphemic lexicon is a type of optimization problem that is studied widely in computational linguistics, and there are reasonably effective algorithms available. For example, one can use an iterative process: Start with a lexicon of all whole word forms. Take one word and try all possible ways to segment it into two parts. If the cost is reduced, keep that segmentation and apply the procedure recursively to the individual segments. The problem can be solved easily at least for a corpus with a few million words.

1.4 Neural correlates of language processing

It takes only a few hundred milliseconds for a human to recognize a written or spoken word. During this time, information arriving at sensory cortices is decoded for linguistic content and made available for various cognitive functions. How exactly this happens is, unsurprisingly, a challenging problem for science (Geschwind, 1970). Last century and especially the last decades have provided a wide range of information on the brain basis for reading (Dehaene, 2009) and speech (Hickok & Poeppel, 2007), with converging evidence at least on the major principles. Some of the cortical areas associated with language functions are illustrated in Fig. 5, however, modern research points to parallel and intertwined processes that are not easily captured by any simplistic picture. Nevertheless, language processing in the brain seems to be at least partially compartmentalized with separate networks serving different aspect of language faculty (Ojemann, 1991; Vigneau et al., 2006). Most readily this is exemplified by various examples of brain lesions causing highly specific language disorders, such as selective problems in one modality, or difficulties with words in one grammatical class, while preserving other functions intact (Caramazza et al., 1988; Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996).

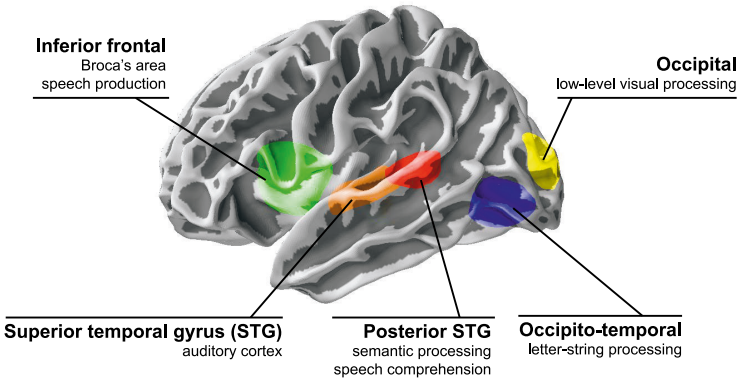

Figure 5. Some cortical areas associated with language processing.

In studies I and IV, the areas of specific importance are the bilateral temporal cortices. In particular, the left superior temporal gyrus (STG) has a central role in speech perception, word representation and processing of word meaning (Moerel, De Martino, & Formisano, 2014; Salmelin, 2007; Salmelin, Kujala, & Liljeström, 2019; Wilson, Bautista, & McCarron, 2018). Intracranial recordings show that invariant representations of speech sounds are organized in the posterior parts of STG (Chang et al., 2010; Mesgarani, Cheung, Johnson, & Chang, 2014), and the activation of this area is enhanced in tasks requiring access to phonological and lexical word forms (W. W. Graves, Grabowski, Mehta, & Gupta, 2008). In addition to language functions, STG is involved in a variety of cognitive tasks, and is especially critical for social cognition (Bigler et al., 2007).

Neural responses associated with reading and hearing necessarily differ during the early processing stages related to low-level sensory perception. However, both modalities seem to invoke similar language-related brain responses at later stages, and there is significant interaction between the modalities even in the cortical responses associated with the early processing (Klein et al., 2015; Vartiainen, Parviainen, & Salmelin, 2009). The high overlap suggests that language recognition in both modalities relies on a common neural mechanism evolved for integrating audio-visual speech (van Atteveldt, Formisano, Goebel, & Blomert, 2004).

Cortical activation during word recognition can be studied with modern brain monitoring methods. The high time-sensitivity of EEG and MEG makes these methods particularly well suited for tracking neural processing (Pylkkänen & Marantz, 2003; Salmelin et al., 2019). Time-sensitive measures that can be extracted from the signal and which have proven useful in analysis of language functions include event-related potentials/fields (ERP/ERF) and induced oscillations, often referred to as event-related synchronization and desynchronization. An evoked response (ERP/ERF) is detected when the waveform of the EEG/MEG signal is phase-locked to the timing of a stimulus (or other consistent trigger point in an epoch) whereas the induced responses refer to enchancement or attenuation of certain frequency components with respect to a trigger point but the signal is not exactly phase-locked. In this thesis, the analysis is mostly restricted to evoked responses.

The time-course of the activation to spoken words, as measured by EEG and MEG, shows a characteristic pattern. A transient response occurs at 80-120 ms after a sound onset, localized in the vicinity of the auditory cortex. The response is usually termed N1 or N100 in EEG studies or, for its magnetic counterpart, N100m (Hari, Kaila, Katila, Tuomisto, & Varpula, 1982). This response is sensitive to speech-specific acoustic features (Heinks-Maldonado, Nagarajan, & Houde, 2006; Parviainen, Helenius, & Salmelin, 2005; Tiitinen, Sivonen, Alku, Virtanen, & Näätänen, 1999) . The response is also modulated by top-down processes such as expectations and interaction with imagined speech (Sams, Möttönen, & Sihvonen, 2005), which is explored further in Study IV.

In the visual modality, presentation of a written word evokes the first distinct response in the occipital cortex at around 100 ms after the stimulus onset. This

response seems to reflect low-level visual features, such as overall visual complexity, but is insensitive to whether the stimulus contains letters (Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999; Wydell, Vuorinen, Helenius, & Salmelin, 2003). The occipital response is followed by an activation localized in the left occipito-temporal cortex peaking at around 150 ms. This is the first response that shows selective sensitivity to letters. The response is stronger when a stimulus contains alphabets rather than geometric symbols (Gwilliams, Lewis, & Marantz, 2016; Tarkiainen et al., 1999). Activation in this time window has shown some sensitivity to statistical properties of the letter-stimulus, such as the frequency of letter bigrams, but it does not, at least not consistently, reflect lexicality of the stimulus (Solomyak & Marantz, 2009a). The response is interesting as it seems to be a logical place to look for any morphemic representation that may activate before full word representations. This is explored further in Study I. The occipito-temporal activation seen in MEG may be related to activation in the fusiform area that is observed using fMRI in reading tasks. The fusiform area has been characterized as a visual word form area (VWFA), or the letter-box area, and considered to have key significance for the reading ability (L. Cohen & Dehaene, 2004; Dehaene, 2009; Dehaene & Cohen, 2011).

The N1 activation in hearing, or the occipito-temporal activation in the case of reading, is followed by sustained activity starting from 250 ms, with a maximum around 400 ms, and hence termed the N400 in EEG studies, or N400m or M350 in MEG literature. This response is long lasting, and it is possible that it consists of several components with differing functional sensitivity (Pylkkänen & Marantz, 2003). The response is usually localized near the STG but has contributions from the wider temporal region and other nearby areas (Kutas & Federmeier, 2011; Van Petten & Luka, 2006). Importantly, both listening and reading tasks generate a very similar response, hence, it can reflect modality independent language functions (Vartiainen, Parviainen, et al., 2009). This makes N400(m) very relevant for studies seeking to link brain responses with abstract linguistic features, and it will be explored in our studies I, II and IV.

The N400(m) activation is sensitive to a wide array of semantic and syntactic manipulations of linguistic input (Kutas & Federmeier, 2011). For example, word frequency influences the response strength, with more frequent words eliciting weaker response when controlled for other factors (Halgren et al., 2002; Simon, Lewis, & Marantz, 2012). More generally, it seems that constraints imposed by the context are an important factor for the response strength. That is, if the word can be expected due to sentence context, or if it is primed with the same or related word, the response strength is reduced. Conversely, if the word has an illegal grammatical form or is otherwise unexpected, the response amplitude is increased (Halgren et al., 2002; Helenius, Salmelin, Service, & Connolly, 1998; Service, Helenius, Maury, & Salmelin, 2007). The response has been considered as index of difficulty for integrating the word in the current context (Hagoort, 2008), or alternatively, index of facilitated access of lexical information (Lau, Almeida, Hines, & Poeppel, 2009).

Most studies of lexical processing have used designs that contrast brain responses between word categories, such as low-frequency versus high-frequency

words, which are then linked to either increase or decrease of the response amplitude. However, recently it has been shown that the N400(m) response can also be quantitatively predicted by surprisal measures in sentence context (Frank et al., 2015). Again, this makes N400(m) an especially promising candidate for testing quantitative models.

Morphological processing in word recognition is thought to be indexed, e.g., by morpheme frequency effects. That is, frequencies of word stem and affix separately modulate the reaction times and brain responses, and the whole word frequency by itself does not explain all the data (Baayen, Wurm, & Aycock, 2007). Other linguistic measures that seek to capture the morphological aspects can also be used to identify traces of morphological processing. For example, morphological family frequency (lemma frequency including derivations) has been observed to modulate left hemispheric M350 amplitude, with high family frequency linked to higher amplitude (Pylkkänen, Feintuch, Hopkins, & Marantz, 2004). In general, morphological properties are most robustly associated with the late temporal responses in both reading and listening tasks (Cavalli et al., 2016; Leminen, Lehtonen, Bozic, & Clahsen, 2016; Leminen, Smolka, Duñabeitia, & Pliatsikas, 2019; Vartiainen, Aggujaro, et al., 2009).

2. Aims

The aim of this thesis was to study brain functions related to processing of words and their internal structure. The methods of choice were MEG and eye-tracking, both providing high time resolution. Specifically, we sought to interpret the neural responses using quantitative, well-defined statistical models emerging from the field of computational linguistics.

i. Are information measures associated with different levels of linguistic abstraction, i.e., letters, morphemes and words, related to the cortical responses during reading?

ii. Are the semantic features of morphemic units reflected in the cortical responses? Is there a difference between statistical and linguistically defined morphemic units?

iii. How is subword structure of words, as determined by a statistical model of morphology, related to eye-tracking measures during word recognition. What does this tell about the optimal morphological structure?

iv. Is explicit memory maintenance of language information reflected in the cortical responses when the memory matches sensory input? Is it possible to distinguish between effects related to whole-word vs. subword level matching?

3. Materials and Methods

3.1 Participants

Altogether 78 healthy adults volunteered in the experiments as detailed in Table 1. None participated in more than one experiment. All were native Finnish speakers with normal or corrected-to-normal vision and no reported neurological or language-related abnormalities. The participants in all MEG experiments were right-handed by Edinburgh inventory (Oldfield, 1971). Everyone gave informed consent and were reimbursed for their time with small monetary compensation (studies I and II) or movie tickets (studies III and IV).

Studies I, II and IV were approved by the Research Ethics Committee of the Helsinki University Central Hospital. Study III was approved by the ethics committee of Aalto University.

Table 1. Participants in Studies I-IV

	Number of participants	Females	Age (mean)
Study I & II	20	11	20-37 (24.4)
Study III a	24	22	19-44 (26.3)
Study III b	26	22	19-29 (22.6)
Study IV	24	13	19-33 (23.0)

3.2 Experimental procedures

Studies I, and II evaluated the same dataset. All participants took part in a non-primed visual word recognition task, while cortical signals were recorded using MEG. Study III involved two components: a single word decision task and a multiword reading task. In both experiemnts, the eye-movements were recorded using an eye-tracking camera.

In Study IV there were two tasks with identical stimuli. The stimuli consisted of both visual symbols and auditory non-words. The main task was an auditory memory task, while in the control task, only the visual symbols were to be attended.

3.3 Magnetoencephalography

Magnetoencephalography (MEG) is a non-invasive functional brain imaging method that provides information about neuronal activity in the brain. The measuring apparatus is illustrated in Fig. 6. With a sensor array brought close to the scalp, MEG detects tiny magnetic fields that result from electrical activity of neurons in cell-to-cell communication (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). The first successful measurement of brain-generated magnetic fields was demonstrated by David Cohen in 1968 (D. Cohen, 1968). Since that time, MEG has evolved into a standard method in clinical and research practice.

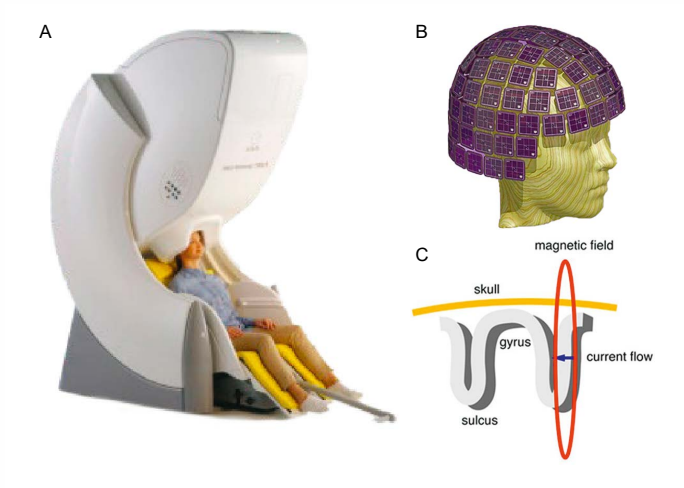

Figure 6. A) MEG apparatus with a participant. B) Diagram of sensor arrangement around the head. C) Diagram of current flow in the cortex and shape of the associated magnetic field.

MEG can measure changes in magnetic fields at a very high precision (order of 10 fT) and at temporal resolution on the order of 1 ms. Locations of neural sources can typically be determined at the centimetre to millimetre scale. Compared to other popular brain monitoring methods, MEG provides complementary information and superior spatial accuracy to electroencephalography (EEG) and superior temporal resolution to functional magnetic resonance imaging (fMRI) (Baillet, 2017). These features make MEG especially suited for studying fleeting cognitive processes such as word recognition. While MEG and fMRI measure different physical processes, one can observe a fair agreement in activated areas and their functional sensitivity in high-level cognitive tasks (Liljeström, Hultén, Parkkonen, & Salmelin, 2009), but also some intriguing differences (Vartiainen, Liljeström, Koskinen, Renvall, & Salmelin, 2011).

The precise resolution of MEG is made possible by sensors based on SQUIDs, short for Superconducting Quantum Interference Devices. The sensors leverage quantization of magnetic fields in small superconducting loops with weak links called Josephson junctions (Josephson, 1962). The SQUIDS are brought below

the critical temperature of superconducting phase transition using liquid helium. The large size of the MEG apparatus is due to the helium storage.

Because the magnetic fields of interest are tiny, on the order of one billionth of the Earth's magnetic field, special circumstances are required to measure them. To reduce outside interferences, measurements are performed in a magnetically shielded room. The room is surrounded by a metal alloy shielding with extremely high magnetic permeability. The high permeability provides a path for most flux lines of external magnetic fields to go around the room and not inside it.

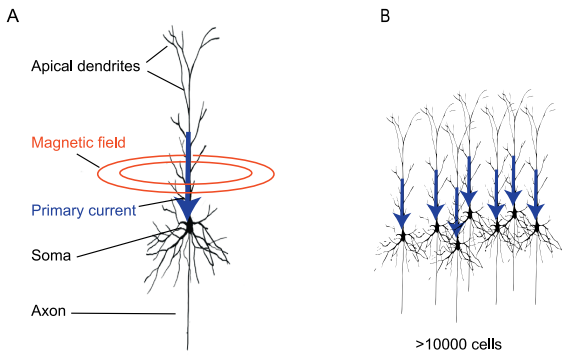

Figure 7. A) A potential difference between apical dendrites and the cell soma generates an electrical current along the dendrite. This primary current generates a magnetic field. B) Near simultaneous currents in an array of pyramidal cells can generate a magnetic field that can be detected outside of the head.

The magnetic fields are result of electric currents generated by neurons. There are two mechanisms that generate the currents: action potentials and postsynaptic potentials. Action potential is the fast depolarization of a cell membrane that propagates along the axon upon firing of the neuron. Information flow between neurons is mediated by neurotransmitters in a synapse which links axon of one neuron to a dendrite of another neuron. Postsynaptic potentials are changes in membrane potential of dendrites of postsynaptic cells that result from the synaptic actions.

The principal source of MEG signal is thought to be the primary currents due to postsynaptic potentials. There are two main reasons for this. First, the action potential is very short, lasting only 1 ms, compared to ~10 ms duration of the postsynaptic potentials. When neighboring cells generate potentials near simultaneously, the longer duration results in a longer overlap and greater summation of postsynaptic potentials. Second, the action potential generates currents that propagate both directions along the axon, resulting in an electric quadrupole. The postsynaptic potential generates an electric dipole as current flows in

one direction along the dendrite. The electromagnetic fields due to action potentials therefore diminish faster with distance than those due to postsynaptic potentials.

The main source of MEG signal is related to postsynaptic potentials in cortical pyramidal neurons (Baillet, 2017). These neurons are abundant in the human neocortex and they are oriented perpendicular to cortical surface in parallel arrays (Fig. 7). Parallel orientation of cells is essential as it results in the summation of signals in neighboring cells. Neurons of other types tend to be disarrayed resulting in more signal cancellation. A near simultaneous PSPs in at least 10000 parallel cells are required to generate a magnetic field that can be measured outside the head.

The primary currents in neurons are balanced by passive volume currents in the surrounding conductive medium. If the volume conductor is a sphere and the direction of the primary current flow is radial, volume currents cancel out the magnetic field and no signal is detected. The head is not a perfect sphere, but still, the strongest contribution to the signal outside the skull is due to primary currents that are oriented tangentially to the skull (Fig. 6c), i.e., when a population of active pyramidal cells is located in a sulcus.

3.4 MEG data analysis

The signal-to-noise ratio (SNR) of MEG is limited by the remnants of external fields, muscular activity, blinks and movements of eyes. Therefore, cleaning of the signal before the analysis is essential.

The effect of external fields was decreased by applying the time-sensitive signal space separation method, tSSS (Taulu & Simola, 2006). The magnetic field is decomposed into a basis of harmonic functions that can be separated into two spaces, corresponding to signals generated in the brain and to outside sources, by leveraging the geometry of sensor configuration. The outside sources are then discarded. The method is further improved by incorporating time-dependent statistics to disentangle sources that are very close to the sensors.

The eye movements were detected by EOG electrodes attached around the eyes. The EOG signal was used to identify eyeblinks (Study I, II and IV).

The data were analysed at the sensor level (Studies II and IV) and at the source level (Studies I and IV). To relate a magnetic field to neural sources, one must construct a model that defines the geometry of neural currents (Hämäläinen et al., 1993). A difficulty here is that there is potentially an infinite number of possible source configurations that produce a given magnetic field. To narrow down the solution space, additional constraints are imposed to select anatomically plausible source configurations.

In Study I and IV, the signal was analysed in the source space by modelling the neural sources as Equivalent Current Dipoles (ECD). In the ECD approach, the sources of magnetic fields are represented by a sparse set of point-like current dipoles located on the cortex.. In multi-ECD modelling, the location and orientation of the ECDs are kept constant, while their amplitudes can vary in time. In study I, the sources were co-registered with an anatomical image of the

participant's brain, obtained using magnetic resonance imaging (MRI) on a separate occasion. Prior knowledge about the source components consistently detected in reading studies (Pylkkänen & Marantz, 2003; Salmelin, 2007; Tarkiainen et al., 1999) was also used to identify plausible sources on the cortex. Multiple ECDs (3 - 8) were fitted for each participant to model the entire time course of the trial. In study IV one ECD in each hemisphere was fitted at the time point of interest for each participant.

3.5 Magnetic resonance imaging

MRI generates 3D anatomical images of the brain. MRI is widely used in clinical practice where it can be applied in diagnosis and staging of diseases without exposure to radiation (McRobbie, Moore, Graves, & Prince, 2017). The first MRI images were produced in 1970s by Paul Lauterbur (Lauterbur, 1973). The importance of MRI in medicine was recognized with a Nobel prize awarded to Lauterbur and his collegue Peter Mansfield in 2003.

MRI works by using an extremely high magnetic field to align spins of protons of hydrogen atoms along the field direction. An external electromagnetic pulse is then used to perturb the spins. As the spins return to the orientation of the magnetic field, they release electromagnetic radiation that is detected by the receiver coils. As the concentration of hydrogen atoms varies across different tissue types, it is possible to reconstruct the distribution of tissue and, thus, brain structure.

In study I, MRI brain images were used to define the geometry of the source space for the MEG source-level analysis. A Siemens Skyra 3T scanner was used in this study.

3.6 Eye tracking

The eye movements during reading (Fig. 8) are assumed to be guided by attention and language functions. An eye tracker allows observation of gaze position and saccades which can be used to infer information about cognitive processes during reading.

In this thesis, we used a camera-based eye tracker, EyeLink 1000 (SR Research, Mississauga, Ontario, Canada). The system consists of an infrared (IR) LED and a camera that are placed in front of the participant. The IR light is reflected from the outer layer of the eye called cornea and captured by the camera, along with pupil positions. When the eyes move, the pupils move relatively faster than the corneal reflex. The calibration phase establishes the correspondence between the gaze position and the difference of cornea/pupil positions, as the subject gazes at specific points on the screen with known locations. Thereafter, gaze position can be estimated by measuring the relative positions of pupil and corneal reflex. Saccades are identified by velocity and acceleration of eye movements.

The eye movements were measured monocularly from the right eye and sampled at 1000 Hz. The accuracy of the system is around 0.25-0.5 degrees which

is comparable to the viewing angle of a single letter (0.41 degrees) in the experiment. The principal focus was on the timing and position of fixations and saccades with respect to target words.

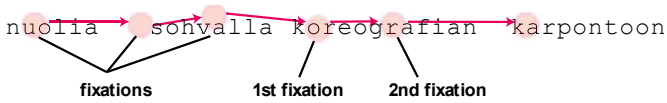

Figure 8. Typical gaze movements in a word recognition task. Fast saccadic movements intercede fixations at words.

3.7 Distributional semantic models

A central feature of words is that they mean something. One way to model a word's meaning, without referencing to objects outside of language, is to analyze how the word is used with respect to other words. Distributional semantic models attempt to quantify semantic similarities and structures based on the distributions of words in a large text corpus, for a review, see e.g., Lenci (2018).

The premise of distributional modeling is the adage, "a word shall be known by the company it keeps" (Firth, 1957). When two words share a similar textual context, they are likely related to each other in some sense. For example, because words *bird* and *feather* sometimes occur in the same paragraph, they have somewhat similar context and, hence, they are in some sense closer to each other than to some completely random word, on average. Similarity may also arise due to grammatical elements or other factors, e.g., a plural word form often has other plural forms as neighbors.

Distributional models represent words as vectors in a space that has typically a few hundred dimensions. The vector space can be thought of as a fuzzy continuum of various semantic and syntactic properties that emerge during model training but are not clearly defined or interpreted. Similarity of two words is defined as the distance between the vectors corresponding to those words. Furthermore, the distributional approach can be extended from words to sentences and longer text snippets by constructing representations for complex expressions with linear algebra operations (Bentivogli et al., 2016). In the simplest case, the phrase vectors can be constructed by summing the vectors of words that constitute the phrase. Similarly, if the entities of the vector space are individual morphemes, it should be possible to construct representations of complex words by summing the corresponding morphemes. The limitation of simple vector addition is that it is commutative, that is, the result is independent of the order in which the terms are summed. For example, the phrases "a girl eats a cookie", and "a cookie eats a girl" result in the same representation. However, in practice, more complex operations generally underperform simple addition (Blacoe & Lapata, 2012).

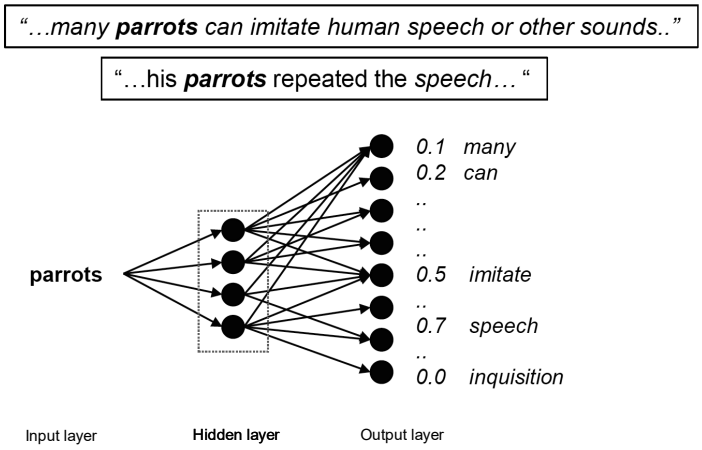

Figure 9. Training of the word2vec model. Surrounding words in the training corpus determine word's context. A neural network with a single hidden layer is trained to predict the probability distribution over words so that it approximates the word's neighborhood in the training corpus. Weights of the hidden layer are then defined as the word's vector representation.

Distributional models provide distances between lexical entities that typically correspond to measures of semantic similarity. However, because the dimensions of vector space have no clear interpretation, the model does not describe how the lexical entities are similar or dissimilar. This contrasts to feature-based models that represent words as a binary list of descriptive features (i.e., "can fly", "is red"). While readily interpretable, the feature-based models have the disadvantage that they rely on human judgment not only in assigning the features to lexical entities but also in selecting the features themselves. Studies have shown that information encoded in distributional and feature-based models is comparable (Riordan & Jones, 2011).

In Study II, we apply a distributional model, word2vec (Mikolov, Chen, Corrado, & Dean, 2013; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) to quantify semantic similarity of words and morphemes. Specifically, we use the skip-gram variant of word2vec. The model works by training a neural network with a single hidden layer (Fig. 9). The network is trained to predict the context of a word in the training corpus. That is, given a word, the output layer provides a probability distribution over neighbouring words that approximates the distribution in the learning corpus. After the model is trained, a word is assigned a vector representation that corresponds to the weights of the hidden layer.

4. Summary of Studies

4.1 Study I: Information properties of morphologically complex words modulate brain activity during word reading

Background

The word recognition process in the human brain evokes a distinct pattern of neural responses which can be functionally, temporally and spatially separated in MEG. There is evidence that these stages involve representation and processing of low-level visual features and orthography, accessing morphological and lexical units, and activation of their meaning (Salmelin et al., 2019). However, current understanding is mostly qualitative, and little is known about the specific computational properties associated with these responses.

Computational models of language based on statistical learning principles have proven successful in NLP tasks (Armeni, Willems, & Frank, 2017). Study I examined whether this type of modeling may also provide a useful description of brain processes during recognition of complex and simple words. We adopted an information theoretic framework and quantified the word surprisal related to different ways of capturing the word's information content. We were specifically interested in morphologically structured representations of multimorphemic words as they play a significant role in the highly agglutinative Finnish language. The morphological representation was modeled using the Morfessor model. The performance of the Morfessor model in predicting brain responses was compared to that of surprisal measures associated with low-level visual features, orthographic features and the whole-word representation.

Experimental paradigm

Cortical responses were measured using MEG while participants (n=20) performed a visual word recognition task. The words were a random selection of Finnish nouns that had high variability across various linguistic dimensions. A significant proportion of the words were multimorphemic with at least one inflectional or derivational affix. In addition, the stimuli included symbol strings, pseudowords and stimuli embedded in Gaussian random noise. Contrasts between the stimulus categories enabled functional localization of salient reading-relevant cortical sources. Each word was shown once for each participant. In order to achieve an acceptable SNR for individual words, the word-specific re-

sponses were averaged across the participants at the source level. The relationship between the responses and language models were assessed using multiple regression models.

Results and discussion

We localized four functionally distinct source components that could be consistently identified in most (>17) participants. These were the occipital response at 100 ms, occipito-temporal response at 150 ms, and bilateral temporal cortex responses around 400 ms.

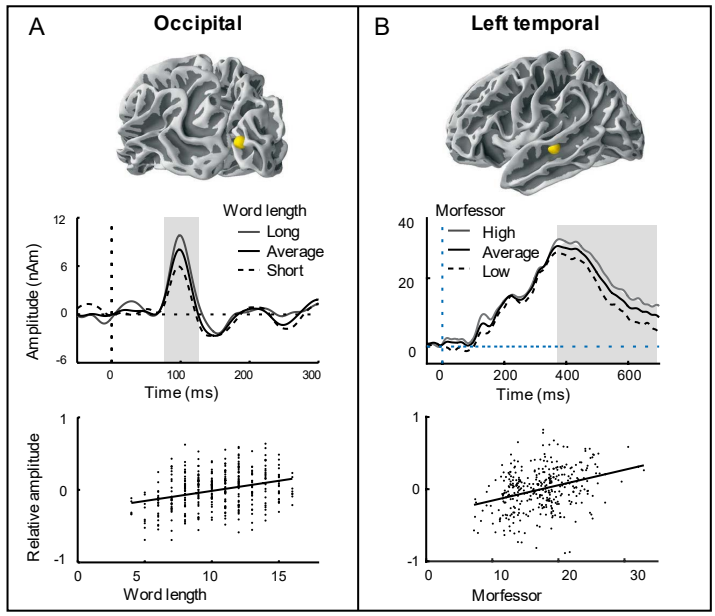

Figure 10. A) Top: Location of the occipital response. Middle: The amplitude of evoked response averaged over short, average and long words. Bottom: The response amplitude to individual words of a given length, averaged over the shaded time-window. B) Top: The location of the left temporal response. Middle: The amplitudes of evoked response averaged over low, average and high Morfessor values. Bottom: The response amplitude to individual words of a given Morfessor value, averaged over the shaded time period.

We found that the first occipital response was related to low-level visual complexity of the stimulus item, whereas the occipito-temporal response was described by a combination of visual and orthographic features, but not morphological or lexical features. In contrast, the bilateral temporal responses were best described by the morphological model and the whole-word frequency that are arguably associated with higher-level language processes (Fig. 10). The success of the Morfessor model and its partial independence of the other predictors in a multiple regression model suggest that at least some words can be modeled as a combination of morphemic units that may arise naturally from the requirement

of optimization of representation. Overall, the study is a step towards a quantitative understanding of cortical language processing.

4.2 Study II: Learned morphemic representations successfully decode brain responses to written words

Background

A morpheme is a linguistic concept that denotes the most elementary meaning-bearing unit of language. However, the morphemes that are defined by linguists are based on rules that are hand crafted to describe a given language, and they are thus not very general. The morphological model Morfessor, also used in Study I, offers one description on how morphemes can be formed from general statistical considerations. This type of description might be conceptually better suited for describing neural computations. Here we examine how the morphemic units from the Morfessor model compare to linguistically defined morphemes as models of neural processing.

The study procedure is visualized in Fig. 11. We modelled the semantic features of words and morphemes by a distributional corpus-semantic model, word2vec (Mikolov, Chen, et al., 2013; Mikolov, Sutskever, et al., 2013). This model represents words as vectors in a high-dimensional space. The distribution of words in the vector space has been shown to correspond to the semantic and syntactic relationships of words and has proven useful in predicting brain responses (Oota, Manwani, & Bapi, 2018; Xu, Murphy, & Fyshe, 2016). We trained this model for three types of morphemic units: the units from the Morfessor model, linguistic morphemes, and whole-word units. The assumption was that a collection of morphemic units that work well in the decoding task may correspond to representations used by the brain.

Experimental paradigm

In this study, the data from Study I was used. For the analysis, we selected a subset of 224 multimorphemic words with at least 50 occurrences in the corpus. We constructed separate corpus-semantic models for whole words, linguistic morphemes, statistical morphemes from the Morfessor model and random subword segments. The word representations were then mapped from the corpus-semantic space to the MEG responses in a machine learning setting (Sudre et al., 2012). The success of this mapping was evaluated by its ability to correctly identify items that were not included in the training of the mapping.

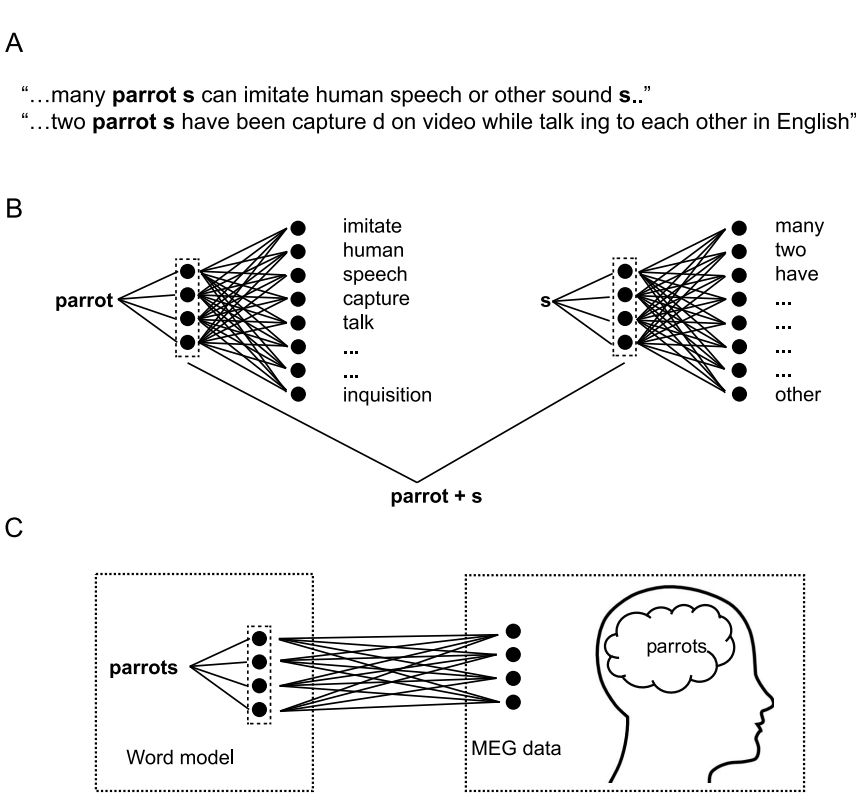

Figure 11. A schematic of the study procedure. A) Words in the learning corpus are segmented into morphemes using morphological analyzers. B) Vector representations for morphemes are created using word2vec skipgram algorithm. Vector representation for a complex word is constructed by summing the morpheme vectors. C) In the neural decoding, a mapping is learned between the word vectors and MEG data recorded during a word recognition experiment.

Results and discussion

We found that the decoding accuracy of statistical Morfessor morphemes and linguistic morphemes were close to equal, whereas completely random subword units performed worse (Fig. 12). The combination of whole-word and morphemic models performed better than either model alone. This may indicate that the brain represents complex words as compositions of morphemes that are not restricted to those defined by linguistics. A compact corpus-semantic model of morphemes obtained via a general optimization principle is thus as successful in describing brain responses as the more complex linguistically defined rule-based model.

When decoding accuracy of individual morphemes constituting the word was tested, we found that the first morpheme, corresponding the word lemma, was the principal driving force for the performance. In the linguistic model, the non-lemma morphemes generally had no additional predictive power. However, in the models using Morfessor morphemes, the predictive accuracy was more evenly distributed among the first and later subword units.

The decoding accuracy reached a level of about 68% using morphological and whole-word models. This is on par with the results of previous studies that have used distributional corpus-semantic models to decode MEG responses evoked by non-inflected simple nouns (Derby, Miller, Murphy, & Devereux, 2018; Hulten et al., 2018; Simanova, van Gerven, Oostenveld, & Hagoort, 2014). The performance of the model with random units was lower than that of the other models but above the significance threshold. This is expected as at least some portion of random units approximate actual morphemes, whatever those may be.

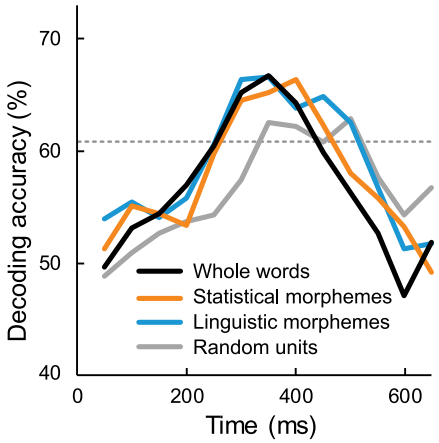

Figure 12. Decoding accuracies as a function of time for corpus-semantic models with different units of representation. Dashed line represents p<0.05 significance threshold.

4.3 Study III: Statistical models of morphology predict eye-tracking measures during visual word recognition

Background

A central issue in studies of human morphological processing has been the nature of the unit of representation in the hypothesized mental lexicon (Schreuder & Baayen, 1995). Morphologically rich languages, such as Finnish, have a large number of possible word forms, and storing all these forms as a separate entries in the mental lexicon may be a suboptimal strategy for a neurocognitive system. An alternative approach is to decompose complex words into morphological building blocks, thereby reducing the number of required units. However, the decomposition of words may also entail processing cost. If the units of mental lexicon are determined by optimization between these two costs, it may be possible to address the question using computational models incorporating similar optimization.

Statistical models of word processing can be based on different types of units of representation, such as single letters, n-grams or morphemes. The choice of units determines the balance between the cost of storage of lexicon and the computational cost of combining these units into words (Virpioja et al., 2018). A minimal lexicon consists of just the letters that need to be combined to form the words. At the other end of the spectrum there are whole-word units that minimize the need of computations, but the size of the required lexicon is large. In a balanced approach, some words are stored as wholes while others are decomposed into units that may resemble the linguistic notion of morphemes. The morphemic models may also differ on whether the morphemes are treated as independent units or whether information of the preceding morphemes is used to predict the subsequent morphemes.

Eye movements during reading, particularly timing of fixations to target words, provide information about automatic aspects of word processing, with a better temporal resolution than behavioural reaction times that are measured only at the end of the whole process. In this study we investigated how statistical models of morphology, reflecting different balance between storage and computation, describe eye-movements during word recognition. We also reached beyond the classical single-word reading task to a more natural experimental setup that is assumedly less affected by decision processes inherent in reaction times. In addition, the initial and subsequent fixation on a target word were used to disentangle early and late processes during word recognition.

Experiment

Participants' eye movements were recorded during word recognition experiments. The stimuli were Finnish nouns consisting of 1-5 morphemes. Multimorphemic words consisted of the stem and one or more inflectional and derivational suffixes. In addition, a set of pseudowords were included. The first experiment was a standard lexical decision task similar to the one used in Studies I & II. Trials started with presentation of a fixation cross for 500 ms, followed by a stimulus word. The task was to determine whether the word was a real Finnish word or a pseudoword. The second experiment was a multiword reading task in which the target word was presented with other unrelated words in a row. In some cases, the row of words also included a pseudoword, and the task was to detect those trials.

In both experiments, the first and subsequent fixations to target words were analysed separately. We constructed linear mixed effects models to predict the fixation durations using the word processing cost from the various morphological models as the independent variable.

Results and discussion

The models based on statistical optimization were successful in predicting eye movements in both tasks. The first fixation duration, thought to correspond to early processing, was best predicted by morphological models that decomposed

all words into morphemes. The late measures, in turn, were best predicted by models that segmented some words into morphemes while keeping others unsegmented, supporting dual route architectures of word recognition that feature simultaneous processing of whole words and decomposed segments (Baayen & Schreuder, 1999; Schreuder & Baayen, 1995). Overall, the models based on whole words performed well for short words, but did not fare well for long words, especially in the second experiment. This suggests that the predictive accuracy of the whole-word models in lexical decision (Virpioja et al., 2018) may partially stem from the decision-making processes that are less emphasized in a more naturalistic task. The models that used predictive information of preceding morphemes to predict subsequent morphemes performed well for the long words.

4.4 Study IV: Two distinct auditory-motor circuits for monitoring speech production as revealed by content-specific suppression of auditory cortex

Background

Speech production allows the brain to make explicit forward predictions about incoming sensory information. The response to spoken stimuli on the auditory cortex is modulated if the listener himself is the one speaking. In addition to overt speech, also imagined speech has been shown to reduce the auditory cortex activation to speech stimulus (Sams et al., 2005). In this study, we used MEG to examine whether the forward prediction and auditory cortex modulation is sensitive to context at subword level, i.e., whether the brain tries to match the content of the phonological working memory to the individual syllables of speech stimulus. If the attenuation is context-specific at the word level, it should be observed only when the imagined word matches exactly the spoken word input. If the attenuation is observed when the imagined word has only a partial match with the spoken word, one may conclude that the forward prediction must operate at a subword level.

Experimental paradigm

The stimuli were sequences of 5 spoken pseudowords that participants listened to while MEG was recorded. In addition, unrelated visual symbols were presented on the screen simultaneously with the spoken words. The symbols were used in a control condition.

The experiment consisted of two conditions with identical auditory and visual stimuli but different task instructions: the memory task and the control task. In the memory task, the participant was instructed to memorize the first pseudoword and repeat it at the end of the trial. The pseudoword thus had to be kept in short-term memory during listening of the subsequent pseudowords. In some cases, the sequence of pseudowords included items that matched the memorized pseudoword either fully or had a single matching syllable. One aspect of the study was to determine whether the type of match is reflected in the brain responses. In the control task, the instruction was to disregard the spoken input and count the number of repeating geometric symbols. Hence, the control task provided a baseline condition without the imagined speech.

Results and discussion

We analyzed the MEG responses to the final pseudoword of each trial. The independent variables were the experimental condition and the relationship of the final pseudoword to the first pseudoword in the trial.

In the memory condition, we found that the cortical N1m responses were significantly attenuated when the pseudoword was concordant with the item main-

tained in the working memory. The attenuation was evident in bilateral temporal regions in both sensor-level and source-level analysis. No such attenuation was observed in the control condition. In contrast, when the beginning of the pseudoword matched but its ending differed from the rehearsed pseudoword, there was a significant enhancement of the N1m response on the left hemisphere but not on the right. The result supports distinct circuits for subword and whole-word level processing. The left temporal cortex was especially sensitive to partial item matches whereas the whole-item match was reflected on both left and right hemispheres.

5. General Discussion

Efforts to link brain activity to quantitative models are essential to gain a more detailed understanding of the computational properties underlying the language ability. Models developed for NLP applications have shown promise in describing morphological or subword aspects of human word processing, although these models have not been developed particularly for neurocognitive tasks (Chater & Manning, 2006; Virpioja et al., 2018). Similar approaches may be applicable to other domains of language function as well.

A central feature of these models is the information theoretic concept of surprisal (Shannon, 1948). The surprisal is also the essence of the so-called Bayesian brain hypothesis, which proposes that many brain functions can be understood in a common and very general framework, that of minimization of free energy (Friston, 2010, 2012). The surprisal based models have intriguing feature that the predictions which are calculated using linguistic information may be connected to physical phenomena that can be studied in neuroscience. Therefore, this type of models offer one possibility to bridge the conceptual gap that exist between linguistics and neuroscience.

Study I of this thesis was aimed to examine how surprisal-based measures, capturing different levels of linguistic abstraction, predict brain responses during visual word recognition. In addition, Study IV examined whether the brain makes explicit predictions at subword level that modulate responses to spoken words. Studies II and III were aimed at seeking the optimal morphemic units for decoding brain responses and describing eye-movements during word recognition, and comparing these subword units to the morphemes that are defined by linguistic analysis.

5.1 Expectations and information in predicting brain responses

MEG responses during visual word recognition are thought to reflect distinct stages of language processing (Salmelin et al., 2019). In Study I, we found that the first cortical stage of this process, the occipital response at 100 ms, could be well modeled with surprisal associated with overall visual complexity. Importantly, two conceptually different ways to calculate the visual surprisal produced predictions that worked equally well. The first method used the number of letters to index complexity. The second method calculated the compressibility of the visual image of the word. These predictions were able to quantitatively predict the amplitude of the response at a single-item level. High image complexity was associated with an enhanced response.

For the second activation peak at the occipito-temporal area, we found, again, that image-complexity/word length acted as a predictor, but in addition, the word bigram frequency significantly improved the prediction when used in the same regression model. This counts as evidence that the response is sensitive to learned statistical features of language. However, we were not able to connect the response to any morphological model. Taken at face value, this finding is in line with proposals suggesting that word recognition proceeds in a hierarchical fashion from single letters to bigrams to more complex units in the ventral visual stream (Dehaene, Cohen, Sigman, & Vinckier, 2005).

The sustained N400m type activation in the bilateral temporal cortices was well predicted by lexicality, indexed by word frequency, as well as surprisal given by the Morfessor model. In a multiple regression setting, we found that these measures explained unique portions of variance.

Study IV explored the role of expectation in a more immediate fashion. That study showed that active maintenance of syllables in the phonological working memory directly affects the brain response related to sensory input. When the input matched the content of the working memory, the neural responses were attenuated as would be expected if the role of prediction is to minimize the energy expenditure required to process incoming information.

Taken together, our results show that the concept of surprisal is well suited for tracking the word recognition process in the human brain. Both long-term predictions based on statistical features of language environment and immediate-term expectations based on memory content seem to attenuate the responses to incoming language content.

5.2 Neural correlates of morphological or subword information

Studies I and II were focused on brain responses related to morphological aspects of complex words. In a highly synthetic language such as Finnish, the reader frequently encounters novel word forms, and it is essential for the neurocognitive system to appreciate the regularity in the word structures. Study I showed that the N400m-type evoked response on bilateral temporal areas was correlated with surprisal values from the whole word and the Morfessor model. Because the two models simultaneously predicted aspects of the response, it may indicate simultaneous representation of whole-word form and the morphemic components. This result would be in line with several hybrid models of word recognition proposing that the full-word route operates in parallel with a decomposition route (Baayen & Schreuder, 1999; Caramazza et al., 1988; Frauenfelder & Schreuder, 1992).

There is some disagreement on when exactly morphemic representations may be accessed. A sublexical hypothesis, adopted by many decomposition models (e.g., Taft, 1994), is a claim that morphemes are accessed at early stages of word recognition. However, there are also arguments claiming that morphological information is considered only after the whole word has been represented (Giraudo & Grainger, 2001).

Some studies, mostly using English language (Solomyak & Marantz, 2009a; Zweig & Pylkkänen, 2009) have found evidence that the response in the occipital-temporal area, preceding the N400m response, is sensitive to visual form of morphemes, supporting the sublexical hypothesis. However, when controlling for visual forms by examining English heteronyms, Solomyak and Marantz (2009b) found that the lexical properties did not affect the early processing before 300ms. In study I, we found that this response was correlated with open-bigram frequency, but the Morfessor model had no unique predictive power in the multiregression model. Other studies in Finnish have also failed to associate the occipito-temporal response to morphological complexity (Vartiainen, Aggujaro, et al., 2009). One reason why the Morfessor model does not correlate with this response can be that the morphemes of the model correspond to lexical units rather than visual forms, and these are not always identical. For example, the letter pattern 'tea' is a morpheme in 'teapot' but not in 'teacher', this situation occurs quite often in Finnish words. Hence, if the occipito-temporal response relates mostly to visual forms, it does not necessarily reflect morphological complexity of Finnish words.

In addition to words, we also examined pseudowords in Study I. Because the Morfessor model is not dependent on the lexicality of the word, the surprisal values can also be calculated for pseudowords. Pseudowords with lower surprisal, i.e., those with typical word-like segments, were associated with attenuated neural response, but a longer reaction time. This pattern of behavioural and brain responses is similar to those previously associated with orthographic neighbourhood of pseudowords (Holcomb, Grainger, & O'Rourke, 2002). The morpheme-frequency effect with pseudowords suggests that in some cases, morphemes can be processed also when they are not part of a meaningful whole word.

Study II examined the idea that morphemes are the minimal meaning-bearing units. That is, in order to be considered morphemes, the statistical morphemes should have some semantic properties, like linguistic morphemes. We found that the brain responses to morphologically complex Finnish words could be predicted by a corpus-semantic word model. Importantly, similar predictive accuracy was achieved by training the distributional models for individual morphemes. We show that the set of morphemes from the Morfessor model and those determined by linguistic rules perform equally well in the brain decoding task. This result does not support the idea that linguistic morphemes form an exclusive club of possible meaning-bearing subword segments.

5.3 Optimal subword units for word representation

The question we are faced with morphological decomposition models is what exactly are the morphemic segments and situations that trigger decomposition. With network models we can inquire what the most relevant subword cues are that activate the correct word representations. Studies II and III examined the predictive performance with different types of statistical morphemes as well as linguistic morphemes.

In the context of decomposition models, the cost of word processing is composed of the storage cost of morphemic units and possible additional cost of combining distinct units to form unified representations. In the Morfessor model, this balance can be adjusted with a hyperparameter. In Study III, we found that the total gaze duration during word recognition was best predicted by a balanced model in which some words are decomposed but also many multimorphemic words were represented as whole units. The best performing model used fewer and longer segments than the linguistically structured model.

Indicative of optimization, in Study II, the semantic models using statistical morphemes were an order of magnitude smaller than the models using linguistically structured units or whole words: the number of required units in the statistical morphemic model was $1*10^5$ compared to $9*10^5$ units in the linguistic model. The number of unique whole-words exceeded 10^7. In addition, a significant number of whole words were found only once or a few times in the corpus, preventing accurate model construction and highlighting the importance of morphological considerations in the highly synthetic Finnish language.

In addition, we found that linguistic affixes had no predictive power in the semantic model, but some Morfessor affixes did (here affix refers to any subword unit that is not the first one in the word). This may result from the fact that the optimal Morfessor morphemes are longer and, therefore, can contain more meaning or semantic information than the less specific linguistic morphemes. For example, longer affixes can incorporate several grammatical markers in a single subword unit. These results suggest that, at least in the case of Finnish, the brain may prefer a more coarse-grained image of language than that defined by linguistics.

5.4 Future directions

The studies reported in this thesis provide a basis for future research in several areas. Studies I and III employed the Morfessor model to describe aspects of human word recognition and contrasted those to traditional psycholinguistic variables. While the results suggest that Morfessor accounts for a unique proportion of variance in reaction times, eye-movement measures and brain responses, and in general outperforms simpler models, the results are nevertheless quite novel, and their reliability would be greatly enhanced by replication with different word sets and participant populations. As the differences between different models are relatively modest, future studies should pay detailed attention to selecting optimal statistical methods for reliable model comparison.

In addition, while the Morfessor model has proven to be a very useful description of highly synthetic languages such as Finnish, there is nothing language-specific in the model. On the contrary, the whole approach is based on a general idea of optimization of representation that should be shared by wide variety of languages, albeit perhaps to a different degree. Therefore, replication using different languages would be paramount. The simplest way to proceed might be to

leverage already existing datasets such as lexical decision megastudies (Keuleers, Diependaele, & Brysbaert, 2010; Keuleers, Lacey, Rastle, & Brysbaert, 2012). A success in explaining behavioural reaction times would then warrant further studies using more involved measures such as eye movements and brain responses. All studies employed in this thesis employed experimental setups that required recognition of words in isolation. As a central aspect of morphology is marking the grammatical status of words in sentences, any experiment using isolated words cannot be expected to fully capture the phenomena. Therefore, future studies will also need to consider more naturalistic reading situations. Furthermore, if the brain does operate by trying to predict incoming linguistic content, the mechanisms would likely employ variety of information that is available in the particular context and environment. Consequently, models employing only statistical properties extracted from a general text corpus can most likely describe only a portion of the situation.

Study II suggests that the brain might be sensitive to letter-strings that do not always coincide with linguistic morphemes and it may associate them with semantical context. This hypothesis should be examined with a variety of experimental approaches to increase the confidence. One possibility could be traditional priming experiments. It would be interesting to see whether the distributional models for morphemic units could provide predictions for possible non-obvious priming effects beyond semantically transparent prime-target pairs, or perhaps provide a unifying description for form- and morphology-based priming results.

5.5 Conclusions

A portion of the neural and behavioural responses can be attributed to processing or representation of morpheme-like subword units, without invoking any explicit linguistic rules on what constitutes a morpheme. The statistical models of morphology provided equal or superior performance to linguistically structured models. Morphemic representations may emerge organically from requirement of optimization of representation.

Morphological processing was dissociated from that of low-level visual and orthographic features and was linked primarily with sustained activation of the bilateral temporal regions from about 250 ms onwards after stimulus presentation.

Models emerging from computational linguistics provide quantitative predictions that can be tested against neuroscientific data. These models can drastically improve neurocognitive study of language by leveraging the fast-paced ongoing development in machine learning methods.

References

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00232

Armeni, K., Willems, R. M., & Frank, S. L. (2017). Probabilistic language models in cognitive neuroscience: Promises and pitfalls. *Neuroscience & Biobehavioral Reviews*, *83*, 579–588. https://doi.org/10.1016/j.neubiorev.2017.09.001

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*(3), 438–481. https://doi.org/10.1037/a0023851

Baayen, R. H., & Schreuder, R. (1999). War and peace: Morphemes and full forms in a noninteractive activation parallel dual-route model. *Brain and Language*.

Baayen, R. H., & Smolka, E. (2019). *Modeling morphological priming in German with naive discriminative learning* [Preprint]. https://doi.org/10.31234/osf.io/nj39v

Baayen, R. H., Wurm, L. H., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon*, *2*(3), 419–463. https://doi.org/10.1075/ml.2.3.06baa

Baillet, S. (2017). Magnetoencephalography for brain electrophysiology and imaging. *Nature Neuroscience*, *20*(3), 327–339. https://doi.org/10.1038/nn.4504

Bentivogli, L., Bernardi, R., Marelli, M., Menini, S., Baroni, M., & Zamparelli, R. (2016). SICK through the SemEval glasses. Lesson learned from the evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment. *Language Resources and Evaluation*, *50*(1), 95–124. https://doi.org/10.1007/s10579-015-9332-5

Bertram, R., Schreuder, R., & Baayen, R. H. (2000). The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 489–511. https://doi.org/10.1037/0278-7393.26.2.489

Bigler, E. D., Mortensen, S., Neeley, E. S., Ozonoff, S., Krasny, L., Johnson, M., … Lainhart, J. E. (2007). Superior temporal gyrus, language function, and autism. *Developmental Neuropsychology*, *31*(2), 217–238. https://doi.org/10.1080/87565640701190841

Blacoe, W., & Lapata, M. (2012). A comparison of vector-based representations for semantic composition. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 546–556. Jeju Island, Korea: Association for Computational Linguistics.

Brennan, J. (2016). Naturalistic sentence comprehension in the brain. *Language and Linguistics Compass*, *10*(7), 299–313. https://doi.org/10.1111/lnc3.12198

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*(3), 297–332. https://doi.org/10.1016/0010-0277(88)90017-0

Cavalli, E., Colé, P., Badier, J.-M., Zielinski, C., Chanoine, V., & Ziegler, J. C. (2016). Spatiotemporal dynamics of morphological processing in visual word recognition. *Journal of Cognitive Neuroscience*, *28*(8), 1228–1242. https://doi.org/10.1162/jocn_a_00959

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, *13*(11), 1428–1432. https://doi.org/10.1038/nn.2641

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, *10*(7), 335–344. https://doi.org/10.1016/j.tics.2006.05.006

Chomsky, N. (2014). *Aspects of the Theory of Syntax* (Vol. 11). MIT press.

Cohen, D. (1968). Magnetoencephalography: Evidence of magnetic fields produced by alpha-rhythm currents. *Science*, *161*(3843), 784–786. https://doi.org/10.1126/science.161.3843.784

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *NeuroImage*, *22*(1), 466–476. https://doi.org/10.1016/j.neuroimage.2003.12.049

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204.

Creutz, M., & Lagus, K. (2002). Unsupervised discovery of morphemes. *ArXiv:Cs/0205057*. Retrieved from http://arxiv.org/abs/cs/0205057

Creutz, M., & Lagus, K. (2007). Unsupervised models for morpheme segmentation and morphology learning. *ACM Transactions on Speech and Language Processing*, *4*(1), 3:1–3:34. https://doi.org/10.1145/1187415.1187418

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature*, *380*(6574), 499–505. https://doi.org/10.1038/380499a0

Dehaene, S. (2009). *Reading in the Brain: The New Science of How We Read*. Penguin.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, *15*(6), 254–262. https://doi.org/10.1016/j.tics.2011.04.003

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences*, *9*(7), 335–341. https://doi.org/10.1016/j.tics.2005.05.004

del Prado Martin, F. M., Kostić, A., & Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, *94*(1), 1–18. https://doi.org/10.1016/j.cognition.2003.10.015

Derby, S., Miller, P., Murphy, B., & Devereux, B. (2018). Using sparse semantic embeddings learned from multimodal text and image data to model human conceptual knowledge. *ArXiv Preprint ArXiv:1809.02534*.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis*.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1), 3–71. https://doi.org/10.1016/0010-0277(88)90031-5

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006

Frauenfelder, U. H., & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In *Yearbook of morphology 1991* (pp. 165–183). Springer.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. (2012). The history of the future of the Bayesian brain. *NeuroImage*, *62*(2), 1230–1233. https://doi.org/10.1016/j.neuroimage.2011.10.004

Friston, K., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, *159*(3), 417–458. https://doi.org/10.1007/s11229-007-9237-y

Geschwind, N. (1970). The organization of language and the brain. *Science*, *170*(3961), 940–944. Retrieved from JSTOR.

Graves, A., Wayne, G., & Danihelka, I. (2014). Neural turing machines. *ArXiv Preprint ArXiv:1410.5401*.

Graves, W. W., Grabowski, T. J., Mehta, S., & Gupta, P. (2008). The left posterior superior temporal gyrus participates specifically in accessing lexical phonology. *Journal of Cognitive Neuroscience*, *20*(9), 1698–1710. https://doi.org/10.1162/jocn.2008.20113

Gwilliams, L., Lewis, G. A., & Marantz, A. (2016). Functional characterisation of letter-specific responses in time, space and current polarity using magnetoencephalography. *NeuroImage*, *132*, 320–333. https://doi.org/10.1016/j.neuroimage.2016.02.057

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal*

Society B: Biological Sciences, *363*(1493), 1055–1069. https://doi.org/10.1098/rstb.2007.2159

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage*, *17*(3), 1101–1116. https://doi.org/10.1006/nimg.2002.1268

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, *65*(2), 413–497. https://doi.org/10.1103/RevModPhys.65.413

Hari, R., Kaila, K., Katila, T., Tuomisto, T., & Varpula, T. (1982). Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: Implications for their neural generation. *Electroencephalography and Clinical Neurophysiology*, *54*(5), 561–569. https://doi.org/10.1016/0013-4694(82)90041-4

Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport*, *17*(13), 1375–1379. https://doi.org/10.1097/01.wnr.0000233102.43526.e9

Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1998). Distinct time courses of word and context comprehension in the left temporal cortex. *Brain*, *121*(6), 1133–1142. https://doi.org/10.1093/brain/121.6.1133

Henderson, J. M., Choi, W., Lowder, M. W., & Ferreira, F. (2016). Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in reading. *NeuroImage*, *132*, 293–300. https://doi.org/10.1016/j.neuroimage.2016.02.050

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, *14*(6), 938–950. https://doi.org/10.1162/089892902760191153

Hulten, A., van Vliet, M., Lammi, L., Kivisaari, S., Lindh-Knuutila, T., Faisal, A., & Salmelin, R. (2018). Cracking the problem of neural representations of abstract words: Grounding word meanings in language itself. *BioRxiv*. Retrieved from https://www.biorxiv.org/content/biorxiv/early/2018/08/13/391052.full.pdf

Jonas, E., & Kording, K. P. (2017). Could a neuroscientist understand a microprocessor? *PLOS Computational Biology*, *13*(1), e1005268. https://doi.org/10.1371/journal.pcbi.1005268

Josephson, B. D. (1962). Possible new effects in superconductive tunnelling. *Physics Letters*, *1*(7), 251–253. https://doi.org/10.1016/0031-9163(62)91369-0

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice Effects in Large-Scale Visual Word Recognition Studies: A Lexical Decision Study on 14,000 Dutch Mono- and Disyllabic Words and Nonwords. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00174

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304. https://doi.org/10.3758/s13428-011-0118-4

Klein, M., Grainger, J., Wheat, K. L., Millman, R. E., Simpson, M. I. G., Hansen, P. C., & Cornelissen, P. L. (2015). Early activity in Broca's area during reading reflects fast access to articulatory codes from print. *Cerebral Cortex*, *25*(7), 1715–1723. https://doi.org/10.1093/cercor/bht350

Kostic, A. (2013). Information load constraints on processing inflected morphology. In Laurie Beth Feldman (Ed.), *Morphological aspects of language processing* (p. 317). Psychology Press.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Lau, E., Almeida, D., Hines, P. C., & Poeppel, D. (2009). A lexical basis for N400 context effects: Evidence from MEG. *Brain and Language*, *111*(3), 161–172. https://doi.org/10.1016/j.bandl.2009.08.007

Lauterbur, P. C. (1973). Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature*, *242*(5394), 190–191. https://doi.org/10.1038/242190a0

Leminen, A., Lehtonen, M., Bozic, M., & Clahsen, H. (2016). Editorial: Morphologically complex words in the mind/brain. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00047

Leminen, A., Smolka, E., Duñabeitia, J. A., & Pliatsikas, C. (2019). Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding). *Cortex*, *116*, 4–44. https://doi.org/10.1016/j.cortex.2018.08.016

Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics*, *4*(1), 151–171. https://doi.org/10.1146/annurev-linguistics-030514-125254

Liljeström, M., Hultén, A., Parkkonen, L., & Salmelin, R. (2009). Comparing MEG and fMRI views to naming actions and objects. *Human Brain Mapping*, *30*(6), 1845–1856. https://doi.org/10.1002/hbm.20785

McRobbie, D. W., Moore, E. A., Graves, M. J., & Prince, M. R. (2017). *MRI from Picture to Proton*. Cambridge university press.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006–1010. https://doi.org/10.1126/science.1245994

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *ArXiv Preprint ArXiv:1301.3781*.

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*, 3111–3119.

Moerel, M., De Martino, F., & Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00225

Ojemann, G. A. (1991). Cortical organization of language. *Journal of Neuroscience*, *11*(8), 2281–2287. https://doi.org/10.1523/JNEUROSCI.11-08-02281.1991

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Oota, S. R., Manwani, N., & Bapi, R. S. (2018). FMRI semantic category decoding using linguistic encoding of word embeddings. In L. Cheng, A. C. S. Leung, & S. Ozawa (Eds.), *Neural Information Processing* (pp. 3–15). Springer International Publishing.

Papadimitriou, C. H. (1994). *Computational Complexity*. Addison-Wesley.

Parviainen, T., Helenius, P., & Salmelin, R. (2005). Cortical differentiation of speech and nonspeech sounds at 100 ms: Implications for dyslexia. *Cerebral Cortex*, *15*(7), 1054–1063. https://doi.org/10.1093/cercor/bhh206

Pathria, R. K., & Beale, P. D. (2011). *Statistical Mechanics*. Academic Press.

Pellegrino, F., Coupé, C., & Marsico, E. (2011). Across-Language Perspective on Speech Information Rate. *Language*, *87*(3), 539–558. https://doi.org/10.1353/lan.2011.0057

Pinchot, J., Douglas, D., Paullet, K., & Rota, D. (2012). Talk to Text: Changing Communication Patterns. *Journal of Information Systems Applied Research*, *5*(2), 42.

Pinker, S. (2003). *The language instinct: How the mind creates language*. Penguin UK.

Poeppel, D., & Embick, D. (2017). Defining the relation between linguistics and neuroscience. In *Twenty-first century psycholinguistics* (pp. 103–118). Routledge.

Pylkkänen, L., Feintuch, S., Hopkins, E., & Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: An MEG study. *Cognition*, *91*(3), B35–B45.

Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences*, *7*(5), 187–189. https://doi.org/10.1016/S1364-6613(03)00092-5

Riordan, B., & Jones, M. N. (2011). Redundancy in Perceptual and Linguistic Experience: Comparing Feature-Based and Distributional Models of Semantic Representation. *Topics in Cognitive Science*, *3*(2), 303–345. https://doi.org/10.1111/j.1756-8765.2010.01111.x

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, *14*(5), 465–471. https://doi.org/10.1016/0005-1098(78)90005-5

Salmelin, R. (2007). Clinical neurophysiology of language: The MEG approach. *Clinical Neurophysiology*, *118*(2), 237–254. https://doi.org/10.1016/j.clinph.2006.07.316

Salmelin, R., Kujala, J., & Liljeström, M. (2019). Magnetoencephalography and the cortical dynamics of language processing. In G. I. de Zubicaray & N. O. Schiller (Eds.), *Handbook of Neurolinguistics*. Oxford University Press.

Sams, M., Möttönen, R., & Sihvonen, T. (2005). Seeing and hearing others and oneself talk. *Cognitive Brain Research*, *23*(2), 429–435. https://doi.org/10.1016/j.cogbrainres.2004.11.006

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological Aspects of Language Processing*. Psychology Press.

Service, E., Helenius, P., Maury, S., & Salmelin, R. (2007). Localization of syntactic and semantic brain responses using magnetoencephalography. *Journal of Cognitive Neuroscience*, *19*(7), 1193–1205. https://doi.org/10.1162/jocn.2007.19.7.1193

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Siegelmann, H. T., & Sontag, E. D. (1991). Turing computability with neural nets. *Applied Mathematics Letters*, *4*(6), 77–80. https://doi.org/10.1016/0893-9659(91)90080-F

Simanova, I., van Gerven, M. A. J., Oostenveld, R., & Hagoort, P. (2014). Predicting the semantic category of internally generated words from neuromagnetic recordings. *Journal of Cognitive Neuroscience*, *27*(1), 35–45. https://doi.org/10.1162/jocn_a_00690

Simon, D. A., Lewis, G., & Marantz, A. (2012). Disambiguating form and lexical frequency effects in MEG responses using homonyms. *Language and Cognitive Processes*, *27*(2), 275–287. https://doi.org/10.1080/01690965.2011.607712

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22. https://doi.org/10.1037/0033-2909.119.1.3

Solomyak, O., & Marantz, A. (2009a). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience*, *22*(9), 2042–2057. https://doi.org/10.1162/jocn.2009.21296

Solomyak, O., & Marantz, A. (2009b). Lexical access in early stages of visual word processing: A single-trial correlational MEG study of heteronym recognition. *Brain and Language*, *108*(3), 191–196. https://doi.org/10.1016/j.bandl.2008.09.004

Sudre, G., Pomerleau, D., Palatucci, M., Wehbe, L., Fyshe, A., Salmelin, R., & Mitchell, T. (2012). Tracking neural coding of perceptual and semantic features of concrete nouns. *NeuroImage*, *62*(1), 451–463. https://doi.org/10.1016/j.neuroimage.2012.04.048

Supek, S., & Aine, C. J. (2016). *Magnetoencephalography*. Springer.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 638–647.

Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, *122*(11), 2119–2132. https://doi.org/10.1093/brain/122.11.2119

Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine and Biology*, *51*(7), 1759. https://doi.org/10.1088/0031-9155/51/7/008

Tiitinen, H., Sivonen, P., Alku, P., Virtanen, J., & Näätänen, R. (1999). Electromagnetic recordings reveal latency differences in speech and tone processing in humans. *Cognitive Brain Research*, *8*(3), 355–363. https://doi.org/10.1016/S0926-6410(99)00028-2

van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*(2), 271–282. https://doi.org/10.1016/j.neuron.2004.06.025

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, *97*(3), 279–293. https://doi.org/10.1016/j.bandl.2005.11.003

Vartiainen, J., Aggujaro, S., Lehtonen, M., Hultén, A., Laine, M., & Salmelin, R. (2009). Neural dynamics of reading morphologically complex words. *NeuroImage*, *47*(4), 2064–2072. https://doi.org/10.1016/j.neuroimage.2009.06.002

Vartiainen, J., Liljeström, M., Koskinen, M., Renvall, H., & Salmelin, R. (2011). Functional magnetic resonance imaging blood oxygenation level-dependent signal and magnetoencephalography evoked responses yield different neural functionality in reading. *Journal of Neuroscience*, *31*(3), 1048–1058. https://doi.org/10.1523/JNEUROSCI.3113-10.2011

Vartiainen, J., Parviainen, T., & Salmelin, R. (2009). Spatiotemporal convergence of semantic processing in reading and speech perception. *Journal of Neuroscience*, *29*(29), 9271–9280. https://doi.org/10.1523/JNEUROSCI.5860-08.2009

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., … Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, *30*(4), 1414–1432. https://doi.org/10.1016/j.neuroimage.2005.11.002

Virpioja, S., Lehtonen, M., Hultén, A., Kivikari, H., Salmelin, R., & Lagus, K. (2018). Using statistical models of morphology in the search for optimal units of representation in the human mental lexicon. *Cognitive Science*, *42*(3), 939–973. https://doi.org/10.1111/cogs.12576

Wilson, S. M., Bautista, A., & McCarron, A. (2018). Convergence of spoken and written language processing in the superior temporal sulcus. *NeuroImage*, *171*, 62–74. https://doi.org/10.1016/j.neuroimage.2017.12.068

Wydell, T. N., Vuorinen, T., Helenius, P., & Salmelin, R. (2003). Neural correlates of letter-string length and lexicality during reading in a regular orthography. *Journal of Cognitive Neuroscience*, *15*(7), 1052–1062. https://doi.org/10.1162/089892903770007434

Xu, H., Murphy, B., & Fyshe, A. (2016). BrainBench: A brain-image test suite for distributional semantic models. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2017–2021. https://doi.org/10.18653/v1/D16-1213

Zipf, G. K. (1935). *The Psychobiology off Language.* Houghton Mifflin, Boston, MA.

Zipf, G. K. (2016). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Ravenio Books.

Zweig, E., & Pylkkänen, L. (2009). A visual M170 effect of morphological complexity. *Language and Cognitive Processes*, *24*(3), 412–439. https://doi.org/10.1080/01690960802180420

Aalto-DD 7/2020

ISBN 978-952-60-8911-9 (printed)
ISBN 978-952-60-8912-6 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Neuroscience and Biomedical Engineering
www.aalto.fi

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL DISSERTATIONS