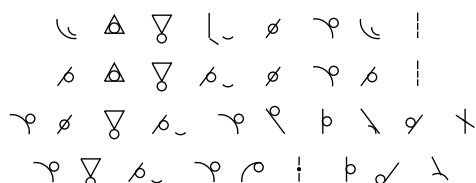




ATLAN

Introduction to a proposed philosophical
international auxiliary language

道可道，非常道。
 名可名，非常名。
 無名天地之始；
 有名萬物之母



*The Path that can be trodden,
 is not the eternal Path.*

*The Name that can be named,
 is not the eternal Name.*

*Nameless, it is the Originator
 of heaven and earth*

*Named, it is the Mother
 of ten thousand things.*

...

-Dao De Jing, chapter 1

Foreword - by Ana Bosnić

It is my distinct pleasure to write the foreword to this work by Stijn, Niek, Jarno, Jep, Jonathan and Max, students of the Humanities Honors Program at Utrecht University. As this book shows, they have attempted to create a new language (one with the potential for universal adoption), thus embarking on a linguistic adventure packed with challenges, curiosity, and a dash of audacity.

Creating a new language is a complex, interdisciplinary task requiring skills, time, effort, dedication, motivation, endless creativity, and above all – (linguistic) knowledge. There also needs to be a deeper understanding of linguistic notions and theories, complexities that underpin human communication, the cultural diversity that influences our lexicon, and the paradox of simplicity and expressiveness.

The authors have embarked on this journey to create Atlan, trying to develop a universal, neutral, and simple language; a language that should be able connect people, transcend boundaries, and be embraced by diverse communities. Ambitiously, they have attempted to craft a linguistic framework that would not favor any specific culture or group, but rather provide a neutral platform

for human communication. Needless to say, that this was a titanic task, and the mere act of attempting it is praiseworthy.

As a constructed language, Atlan benefits from the kind of tools that an organically generated language cannot. Thus, for example, the creators of Atlan have tapped into vast reservoirs of linguistic data to inform their decisions, allowing them to create Atlan's phonetic inventory, script and vocabulary. This fusion of human creativity and computational insights laid the foundation for their linguistic invention, offering a boost in their pursuit of universality, unambiguity, and parsimony. It goes without saying that this innovative approach to language creation serves as a testament to the boundless possibilities that arise when human creativity converges with computational tools.

In conclusion, this book encapsulates the result of the arduous work carried out by Stijn, Niek, Jarno, Jep, Jonathan, and Max in their quest to create a universal and neutral auxiliary language. Throughout the pages of this book, readers will bear witness to their tireless pursuit to create a language that transcends borders and fosters effective global communication. It is my hope that

their journey inspires further research into the intricacies of language, serves as a reminder that the power of human ingenuity knows no bounds, and shows that linguistics can be fun!

And can Atlan become a new lingua franca? Well, only time will tell.

Ana Bosnić

Contents

1	Introduction	1
1.1	Opening Remarks – Niek Elsinga	1
1.2	The story of King Atlas – Stijn Janssens . .	9
1.3	Need for an IAL – Jonathan Roose	13
1.4	Eco’s words – Jonathan Roose	16
1.5	Linguistic relativity – Max Geraerds . . .	18
2	Phonology – Niek Elsinga	40
2.0	Introduction	40
2.1	The Neurologic Basis of Language	46
2.2	Consonants in Atlan	49
2.3	Vowels in Atlan	53
2.4	Decisions regarding the phonetics: tone and prosody	56
2.5	Morphology in Atlan	57

CONTENTS

2.6	Concluding remarks	60
3	Writing Atlan	62
3.0	Writing system	62
3.1	Direction of reading – Stijn Janssens	77
3.2	Punctuation	78
3.3	Transliteration	78
3.4	Numerals	79
3.5	Mathematics – Stijn Janssens	83
3.6	Typography	83
3.7	Names, loanwords, and cartouches	84
3.8	On Dyslexia – Stijn Janssens	88
3.9	T _E X and Atlan – Jarno Smets	94
4	Morphosyntax	96
4.1	Unambiguous Syntax – Jarno Smets	96
4.2	Atlan’s grammar	114
4.3	Greenberg’s universals	118
5	Our Ontology	128
5.1	Parsimony in semantics	129
5.2	The anthropocentricity of language	147
5.3	Phenomenological ontology and qualia . .	152
5.4	Universal Semantics	173

5.5	Concluding remarks	175
5.6	Appendix	176
6	Lexicon	192
6.1	Introduction	192
6.2	Comparison methods and language families – Jep Antonisse	194
6.2.1	Cosine Similarity	194
6.2.2	Language families	201
6.3	Using cognates to generate words	205
6.3.1	Language selection	207
6.4	Vocabulary list – Stijn Janssens & Jep Antonisse	211
6.5	AI Generation – Max Geraerds & Jep Antonisse	211
6.6	Atlan to English	211
6.7	Translation Protocol	211
7	Pragmatics – Jonathan Roose	212
7.1	Logophilia, Wordplay and Rhythm	212
7.2	Mood-markers	212
8	Further suggestions	213
8.1	Sign Language – Stijn Janssens	213

CONTENTS

8.2	Language Variation – Niek Elsinga	213
9	Example texts	214
9.1	The Story of Babel – Jonathan Roose	214
9.2	The Epic of Gilgamesh – Niek Elsinga	214
10	Exercises – Max Geraedts	215
	Epilogue	216
10.0	Special Thanks	216
	Bibliography	217

Chapter 1

Introduction

1.1 Opening Remarks – Niek Elsinga

“The limits of my language mean the limits of my world.”

- Ludwig Wittgenstein

Language, the remarkable construct that binds humanity together, possesses an unparalleled power to shape our thoughts, connect individuals, and cultivate shared understanding. It is through language that we express our deepest emotions and convey ideas, as well as the preservation of the vastness of human knowledge. Yet, in this vast linguistic landscape barriers and borders rise,

1.1. OPENING REMARKS – NIEK ELSINGA

resulting in imperfect communication and the impediment of the exchange of ideas across cultures and nations.

Philosophically, language can be perceived as more than a mere tool for communication. It shapes our understanding of the world, influences our perspectives, and defines our cultural identities. Language is not merely a means of conveying information, but a reflection of our collective history, aspirations, and values.

Natural languages that have emerged throughout human history presents both a marvel and a challenge. While they showcase the richness and diversity of human expression, they also lead to barriers and misunderstandings amongst speakers and cultures. In an attempt to solve these problems, several languages have been constructed; so-called constructed languages, or “conlangs”. As we will see, the language we develop here is an international auxiliary conlang. *International* meaning being as inclusive and as accessible to as much languages of the world as possible. *Auxiliary* meaning that we *don’t* want our language to replace natural language. Our language should be seen as a tool for communicating clearly and internationally.

CHAPTER 1. INTRODUCTION

The creation of artificial language is a practice that is millennia old, and to this present day, still very much alive. Many attempts at a language for international communication have been made in the past. Of course, not all constructed languages are made with the purpose of becoming international auxiliary languages. Some other constructed languages are made for film, such as Klingon in Star Trek. Others are more personal. The group of so-called *Conlangers* is a flourishing community of people who create constructed languages. You might know one of these languages, such as Esperanto. The quest for a constructed international auxiliary language, however, is not new. It has its roots in the early 20th century, when linguists, philosophers, and idealists alike envisioned a language that could serve as a bridge between nations and foster understanding among diverse cultures. Their vision was grounded in the belief that a language constructed with careful consideration of phonology, grammar, and vocabulary could provide a common ground for intellectual discourse, transcending the boundaries imposed by native languages.

This book takes you on a captivating journey through the intricacies of constructing an international auxiliary

language. It explores the fundamental principles underlying language construction, delves into the complexities of phonological categories, and examines the neurologic basis of language acquisition and comprehension. Additionally, it investigates the challenges and opportunities presented by the creation of a culturally neutral and inclusive language.

As we embark on this exploration of language and its creation, we invite you to contemplate the immense potential that lies within a constructed language - a language that aspires to be a unifying force, bringing together individuals from diverse backgrounds, fostering global communication, and ultimately transcending the limitations imposed by our native tongues.

Join us on this intellectual odyssey as we delve into the realm of linguistic possibilities, guided by the belief that language, at its core, reflects our shared humanity. Through the creation of a constructed international auxiliary language, we may pave the way for a more inclusive and interconnected world.

About the project

This book is authored by Jep Antonisse (artificial intelligence), Niek Elsinga (language and culture studies –

CHAPTER 1. INTRODUCTION

linguistics), Max Geraedts (artificial intelligence), Stijn Janssens (philosophy), Jonathan Roose (literature studies) and Jarno Smets (language and culture studies – logic). It was written for the Humanities Honours course ‘Research Seminar’ at Utrecht University, under supervision of Dr. Ana Bosnić (linguistics). Our project was to create a constructed language ‘Atlan’, and write a book about it. From February until June of 2023, we met every week to work on constructing the language, writing literature review essays on the different aspects of the language, programming different tools, and finally putting together this book as the final project. We all enjoyed working on the project, and had many interesting discussions about language, philosophy, literature etc., as well as establishing informal friendships.

The language is based on sketches made by Stijn, who had made an earlier attempt at constructing a language that would fit the proposed constraints, but was dissatisfied with the final results. He collected notes and resources on different aspects that would have to be put into the language (the writing system and phonology had already been assembled), but after realising the sheer time and ambition required to attempt completing it, he put

the project on ice for a few years.

When the project for the Research Seminar was first introduced, Ana gave a short introduction of herself and her work, mentioning the practical application of linguistics seen in constructed languages. Stijn was reminded of the old project he was still intending on finishing someday, and realised that with the help of the two AI students, the project would be a lot more achievable, having the power of computation on our side. As though through serendipity, the rest of the group members happened to be standing in close proximity when Stijn pitched his idea to them, generating much enthusiasm from everyone, and thus the project was decided upon the very same day.

Before we proceed, we must mention that the current language being presented in this book should be seen as a first draft of a fundamental proof of concept. It is far from flawless, and we invite the critical enthusiastic reader to contribute to help us add to the language and revise problems or inconsistencies. The main aimed purpose of the language is to be a competitor of English as a global language. Therefore, it's main goal is to be suited better for such a role than English is. It might

CHAPTER 1. INTRODUCTION

never be finished as a final ‘perfect language’, partly since language is always alive, but if, we may hope, it is found to be more suitable than English, our goal seems to us to be achieved.

Our constructed language ‘Atlan’ is designed to be both an international auxiliary language (IAL) and a philosophical language (PhilIAL). It is built along three primary constraints:

1. Human universality / cultural-linguistic neutrality
2. Unambiguity
3. Elegance / form from function / parsimony

The first constraint covers the goal of the language to be an IAL: a truly unbiased auxiliary language does not show a disproportionate favour of one specific language over any other, as is now very much the case with English being the main IAL (the reason why this book is written in English). It cannot be a mix of a few European languages, like Esperanto for example. Nevertheless, absolute neutrality is impossible because there is no true ‘centre’ to different linguistic structures, and the number of different languages and their relative number

of speakers will also shift the balance in the total world population (this will be accounted for with the aid of AI, see chapter 6.2).

The second constraint overlaps in political and philosophical relevance: a language that is to be learned and commonly spoken by speakers of any language on Earth is intended to unify and overcome language barriers, as if to ‘undo the confusion of tongues’, and to construct a ‘modern Adamic language’. Therefore, miscommunication and ambiguity should be avoided as much as is realistically possible. Within the analytic tradition, philosophy is often regarded as the ‘perfecting of language’ through making statements logically consistent and definitions clearly defined (Stanford, 2022). These concerns together require Atlan to have an orthography that is phonologically consistent, a lack of homonyms and synonyms that do not add any meaningful nuance and a syntax that does not (easily) allow for grammatically confusing or logically ambiguous statements.

The third constraint is the most ideal and philosophical in nature. ‘Elegance’ here is meant in a similar way to how mathematicians and physicists praise simple and straightforward formulas that describe and predict a vast

set of phenomena and data. The goal is thus to have as little unnecessary parts as possible; less is more. This goal we call *parsimony*. This means that any form of complexity, be it orthographic, semantic or syntactic, should arise as an emergent property of the combination of its basic parts.

1.2 The story of King Atlas – Stijn Janssens

We have chosen to name our constructed language ‘Atlan’, which consists of the words ‘AT’, meaning all / every / universal, and ‘LAN’, meaning speak / talk / language. Therefore, the name can be understood literally to mean ‘Universal Language’. Although the majority of Atlan’s lexicon was generated by an AI programmed on natural language data, the syllables ‘AT’ and ‘LAN’ were consciously assigned their meaning as a symbolic homage to the mythical figure titan Atlas.

In Greek mythology, Atlas was said to have been condemned to by the Gods to uphold the firmament for eternity, after having lost in the Titanomachy, an epic battle between the Titans and the Gods. The Greek poet Hesiod located him at the extreme West, at the edge of

1.2. THE STORY OF KING ATLAS – STIJN JANSSENS

the known world (which back then mostly referred to the landmasses surrounding the Mediterranean¹ sea). This made him later be identified with the Atlas Mountains of Northern Morocco. This seems to coincide with a folk legend of the local Mauri people, also known as Berbers, of present-day Morocco, who to this day still tell of the legendary King Atlas of Mauritania. Because of this, a suggested etymology for the name is the local Tamazight word ‘*ádrār*’, meaning mountain.

According to Greek mythology, he was encountered by the hero Perseus. Upon arrival in Atlas’s Kingdom, he asks for shelter, claiming to be the son of Zeus. Atlas refuses, because of a prophecy that once told him that a son of Zeus would come to steal golden apples from his orchard. Because of this, Perseus turned Atlas into a mountain range, with his head at the peak with forests for hair, and his shoulders as the ridges. Perseus, however was not the prophesised apple thief. The real thief was rather his grandson and half-brother (thanks to Zeus’ incestuous practices), Heracles. When fulfilling his twelve labours, he was sent to steal some golden apples from Hera’s orchard, which was tended by Atlas’s

¹From Latin, meaning *middle earth*

CHAPTER 1. INTRODUCTION

daughters, the Hesperides. Atlas and Heracles tricked each other into carrying the firmament, until Heracles managed to escape with the apples.

King Atlas is said to have invented the celestial sphere, and perhaps even first having established the science of astronomy. He was supposedly skilled in philosophy, mathematics and astronomy. Perhaps this led to his connotation of carrying the firmament. King Atlas inspired cartographer Gerardus Mercator, famous for the Mercator projection of earth, to name his world-maps after him. The Atlantic Ocean was named after Atlas the titan, as well.

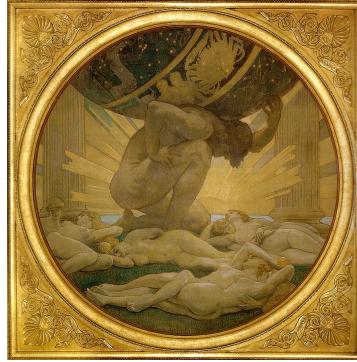


Figure 1: Titan Atlas and the Hesperides, by John Singer Sargent, ca. 1922-1925

In his late dialogue *Timaeus*, the philosopher Plato refers to King Atlas as being the first ruler of Atlantis, a city established by Poseidon. Perhaps this city might have referred to a place which is now known as the Richat

1.2. THE STORY OF KING ATLAS – STIJN JANSSENS

Structure, a geological formation of concentric circles in northern Mauritania just below the Atlas Mountains, matching the description given by Plato. During the purported existence of this city, 12.000 years ago during the African humid period, this area was a lush and fertile land, until the sudden catastrophic global warming event known as the Younger Drias took place, turning the area into the Sahara Desert we know today. Neolithic artefacts from this era have been found around the Richat Structure, as well as fluvial and torrential deposits from the time the Younger Drias is believed to have taken place. Perhaps this was the origin of the myth of its sudden destruction, having been passed on through oral tradition of North African peoples, until it reached the Egyptian priest Sonchis, who Plato claims to be the source of the story.

We have chosen to name our language after Atlas because of his legendary reputation as being the ruler of a Utopian civilisation, a symbol of knowledge, as well as his connotation with philosophy and organised knowledge about the world. It seems appropriate to us to name our language Atlan, being somewhat of an encyclopaedic philosophical language, after this ancient cross-cultural

figure representing wisdom and the bridge between heaven and earth.

1.3 Need for an IAL – Jonathan Roose

Historically the diversity of languages has been both a blessing and a curse. On the one hand has the variety of tongues been a database of ways to understand the world and human expression, on the other it has also led to barriers and in- and outgroups. This is why five of my co-students and I have taken up the ambitious task of creating a so-called International Auxiliary Language (IAL for short), a language that will allow its users to bridge language barriers and lead to mutual understanding between speakers with different mother tongues, a neutral ground on which all international communication can occur.

The lingua franca's of today's world that are used in international relations, like French, English or Swahili give hierarchical importance to the language of one particular group and/or state, these languages are bas on political power and historical conditions, they cannot be neutral, they have become international languages be-

1.3. NEED FOR AN IAL – JONATHAN ROOSE

cause of political interactions and thus are always a political matter. The aim of an IAL is to be a meeting ground of all people without it being dependent on power relations and historical animosities. This project has a lot in common with others IALs, Esperanto, for example, made by L.L Zamenhof with the hope that a collective language will lessen the violence between nations. However, why Esperanto succeeded is also why it is limited. It was made to bring speakers of European languages together and it did, for a small part, however, only speakers of European languages. Such languages are commonly referred to as ‘Euroclones’. Our goal with this project is to create a unifying international language for the whole world, which can thus not be limited to only a small set of language groups.

The ambition we have with the language Atlan is to create a language that is based on nothing more than the human condition. Later in this book Stijn will explain more how we intend to do this however, for now I would like to introduce a term that might help to better understand what we hope to achieve with Atlan. A *tertium comparationis* is a wish of many translators to have some way to compare the meaning of the original text with

CHAPTER 1. INTRODUCTION

their translation. Deriving from the Latin for ‘a third comparison’ this term describes the want for a ‘perfect language’ that could be used to completely translate a given text. Of course, the translation is meant to have the same meaning however, some meaning will always get lost, the comparison is to see whether the new meaning has not lost the essence of the original. Whether the very quiddity of the original meaning is captured in the translation. What translators want is a semiotic system that can show the essence of a message in a way that can be compared to all other languages. In Atlan we hope that we can create a linguistic code, the rules of grammar and use of a language, that can function as such a *tertium comparationis* by making it based on the essential human experiences. A language that can get to the essence of a thing by basing it on the basic human ontological being. This language will function as an IAL that is neutral and universal, it is a language based on the human condition that every human experiences.

1.4 Eco's words – Jonathan Roose

In this ambitious project we are indebted to the numerous projects that predate ours with the same or similar aims. Not only is there Zamenhof's Esperanto many more thinkers have dealt with the quest for an IAL. To name all would be too numerous however we can mention a book that has introduced many of the language projects to us. Umberto Eco's book *The Search for the Perfect Language* has been a great source of inspiration in this project. Like Esperanto the book is mostly concerned with Europe. Nonetheless to finish this introduction to Atlan we end with a passage from his book to summarise the project:

“Is it possible to reconcile the need for a common language and the need to defend linguistic heritages? Both of these needs reflect the same theoretical contradictions as well as the same practical possibilities. The limits of any international common language are the same as those of the natural languages on which these languages are modelled: all presuppose a principle of translatability. If a universal common language claims for itself the capacity to re-express a text written in any other

language, it necessarily presumes that, despite the individual genius of any language, and despite the fact that each language constitutes its own rigid and unique way of seeing, organizing and interpreting the world, it is still always possible to translate from one language to another. However, if this is a prerequisite inherent to any universal language, it is at the same time a prerequisite inherent to any natural language. It is possible to translate from a natural language into a universal and artificial one for the same reasons that justify and guarantee the translation from a natural language into another. The intuition that the problem of translation itself presupposed a perfect language is already present in Walter Benjamin: since it is impossible to reproduce all the linguistic meaning of the source language into a target language, one is forced to place one's faith in the convergence of all languages. In each language 'taken as a whole, there is a self-identical thing that is meant, a thing which, nevertheless, is accessible to none of these languages taken individually, but only to that totality of all of their intentions taken as reciprocal and complementary, a totality that we call Pure Language [reine Sprache]'" (Eco 1995:345)

1.5 Linguistic relativity – Max Geraerdt

To start I would like to explore linguistic relativity. It is an important term within the study of linguistics, and I would like to explore the possible consequences it has for a universal language. For those of you who are unfamiliar with this term, it refers to the hypothesis that Sapir and Whorf – two linguists – developed about how the structure of a language can influence our thinking. Sapir and Whorf developed two hypotheses about this presumed phenomenon. A strong and weak hypothesis, the strong one argues that language determines thought and that linguistic categories limit and determine cognitive categories. Effectively stating that the language one speaks limits their cognitive abilities. This hypothesis is now disregarded by many modern linguists. The weak hypothesis, however, is still a main point of discussion among linguists. It argues that language influences thought but does not determine it. This weaker version is much easier to accept. A good example of this is the way in which different languages have different perceptions of colors, representations of time and other elements of cognition. So, while it is safe to say that the

CHAPTER 1. INTRODUCTION

strong hypothesis is false it is difficult to deny that language does have an influence on our way of thinking. Language is our way of representing the world. A difference in language can lead to a difference in our representation of the world.

Ideas and views that would eventually go on to become to define linguistic relativity are first found in ancient philosophy. However, it only began to enter mainstream research in the eighteenth and nineteenth century, with German romantic philosophers on the forefront. (German) nationalism fuelled the discussion about language and its relationship with culture and unity at this time. Wilhelm von Humboldt – a Prussian philosopher, linguist and government functionary – stated in 1820:

The diversity of languages is not a diversity of signs and sounds but a diversity of views of the world (Traband, 2000).

After this movement in Europe, American scientists began discussing this same subject in the early twentieth century. At this time the idea that some languages were superior to others was commonly accepted. It was thought that lesser languages maintained their speakers

1.5. LINGUISTIC RELATIVITY – MAX GERAERDTS

in intellectual poverty (Migge, 2007). This caused some American linguists to seek to eradicate Native American languages, they thought that its speakers were savages and needed to speak English to become civilized. The first linguist that began refusing these beliefs was the American Franz Boas, during his studies he became fascinated with the Inuit. After learning their language and culture he began stressing the equal worth of all cultures and languages. There were no such thing as lesser languages according to Boas. Boas' student Edward Sapir went back to the Humboldtian idea that language is vital to understand the unique perception everyone has of our world (Leavitt, 2010). Sapir argued that no two languages could never be perfectly translated to each other. This dissonance in language continued in the world view of individuals according to Sapir:

No two languages are ever sufficiently similar to be considered as representing the same social reality. The worlds in which different societies live are distinct worlds, not merely the same world with different labels attached (Sapir, 1929).

This did not however mean that Sapir agreed with the

CHAPTER 1. INTRODUCTION

strong hypothesis, he did in fact disagree with it. Stating that:

It would be naïve to imagine that any analysis of experience is dependent on pattern expressed in language (Sapir, 1946).

So, it seems that in these middle stages of the development of linguistic relativism views on the subject changed dramatically over the years. As we continue through history, we arrive at the linguist Benjamin Lee Whorf. Whorf was one of Sapir's students and has been associated with linguistic relativity more than any other linguist. One of his best-known examples regards the different words the Inuit have for snow compared to the one word we have for it in English. This example showed that you could not perfectly translate even simple concepts such as snow between languages. This example was however later contested as a misinterpretation by Whorf (Pullum, 1991). Another example of Whorf's linguistic relativity was the time in Hopi. Whorf argued that the Hopi did not have countable units of time compared to the SEA – standard European languages – the Hopi instead regarded time as a single continuous concept. This notion was however also later contested by

other linguists. In the 1980's Ekkehart Malotki claimed that he had not found any evidence for the claims Whorf had made about the Hopi. This refute was then in its turn contested by relativist scholars who criticized Malotki's study for forcing the Hopi language into a grammatical model that didn't fit the data (Lee, 1996). How Whorf approached the Hopi is an example of the structure-centered approach. This approach focuses on a structural difference between languages. It then examines the possible consequences and ramifications of this structural difference. The Hopi and the peculiar structure time has in their languages is a prime example of this approach (Lucy, 1997). Whorf died at 44 and left many unpublished papers, these were eventually published in a single volume titled *Language, Thought and Reality*. Since neither Sapir nor Whorf had officially formulated a hypothesis Brown and Lenneberg – two influential linguists from the twentieth century – formulated their own in 1954:

- (i) "The world is differently experienced and conceived in different linguistic communities"
- and (ii) "Language causes a particular cognitive structure" (Brown, 1954).

These were later reformulated by Brown into the *weak*

CHAPTER 1. INTRODUCTION

and *strong* formulations:

Structural differences between language systems will, in general, be paralleled by non-linguistic cognitive differences, of an unspecified sort, in the native speakers of the language. (Weak)

The structure of anyone's native language strongly influences or fully determines the worldview he will acquire as he learns the language. (Strong)

Thus, we have arrived at the creation of the Sapir-Whorf hypothesis. Which was not created by Sapir or Whorf. What we have also seen is the difficulty of quantifying linguistic relativity. We have seen many bold claims which have all in turn been contested by others. With this reflection we arrive at the last stretch of the development of linguistic relativity. In 1996 the anthology *Rethinking Linguistic Relativity* was published. It discussed linguistic relativity that focuses more on cognition and social aspects of language. For example, men speaking Guugu Yimithirr could give directions based on a compass-like system of north, south, west and east (Levinson, 1998). This shift of focus alongside the development of better means of conducting research ushered in much new re-

1.5. LINGUISTIC RELATIVITY – MAX GERAERDTS

search seeking to not only define but quantify linguistic relativity.

Brown and Lenneberg thought that languages described the same objective reality. They decided to research if the difference in describing this reality could be proven to have influence on behaviour. For their experiments they decided to focus on the different descriptions of colour in different languages. For one of their first experiments, they tested whether it was easier for English speakers to remember colour shades for which there existed a specific word opposed to shades which were more difficult to describe with words. Later they also compared results between English and Zuni speakers – Zuni classifies green and blue as the same – and it was found that Zuni speakers did have more difficulty making distinctions between shades in the green/blue category (D'Andrade, 1995). These studies by Brown and Lenneberg began a tradition of investigating linguistic relativity through colour terminology. Real differences could be seen between the perception of colour by an individual and the language they speak. These studies however also received criticism because colour perception is hardwired into the brain. This causes it to be universally restricted by some

CHAPTER 1. INTRODUCTION

factors for all humans (Lucy, 1997). I however have some nuance to add to this argument. While it is true that colour perception is hardwired into our neural system, I believe linguistic relativity to be a *quale*. A relativity regarding our experiences, thoughts and inner dialogue. While it is undoubtedly true that colour perceptions are *biologically* the same for all of us, I believe the real difference lies in our *mental* representation of this biological phenomenon.

Colour research was continued by Berlin and Kay, an anthropologist and linguist respectively who are most well known for their research in colour. During their research they found clear universal conventions when it comes to colour naming. For example, they found that although different languages have different colour terminology, there are universal trends among them. Languages who only have three colour terms all have the same three colours, black, white and red (Berlin, 1969). Because colour naming was originally thought to be random, this new information was seen as a powerful argument against linguistic relativity (Grumperz, 1996). This criticism has since in turn been criticised by relativists such as Lucy who argued that the conclusions from Berlin

1.5. LINGUISTIC RELATIVITY – MAX GERAERDTS

and Kay were skewed because they insisted that colour terms only encoded colour. According to Lucy, this made them blind to instances where colour terms contained and provided other information that might be considered as linguistic relativity (Lucy, 1992). As we see and discuss more aspects of linguistic relativity it should become clear that it is a very broad and contested hypothesis.

Advances in cognitive psychology and cognitive linguistics again brought a new wave of studies that focused on linguistic relativity. George Lakoff, for example argued that language is often used metaphorically and that this metaphorical use can give us insight in the cognitive effect of language. He gave the example that in the English language time is often likened with money, a lot of metaphors including time talk about it like it can be invested, saved and spent. This cognitive relationship submerging through language can be a sign of linguistic relativity. Especially considering that other languages do not talk about time this way. Other metaphors like this that are based on human experience are languages where up is associated with good and down with bad. This association can be seen in many myths and folklore, such as heaven being high up in the skies and hell

being down. Lakoff also argued that metaphors play an important role in political debates such as the “right to life” or “right to choose” (Lakoff, 1980). Lakoff revitalized linguistic relativity not only because of his newly found results, but also because he reappraised linguistic relativity thus rendering past criticisms moot. He did this by concluding that the debate regarding linguistic relativity had been confused. To clear up this confusion Lakoff described four parameters on which researchers differed in their opinions on what constitutes as linguistic relativity. These were his four parameters:

1. The degree and depth of linguistic relativity. Perhaps a few examples of superficial differences in language and associated behavior are enough to demonstrate the existence of linguistic relativity. Alternatively, perhaps only deep differences that permeate the linguistic and cultural system suffice.
2. Whether conceptual systems are absolute or whether they can evolve
3. Whether the similarity criterion is translatability or the use of linguistic expressions
4. Whether the focus of linguistic relativity is on language or in the brain (Lakoff, 1987)

1.5. LINGUISTIC RELATIVITY – MAX GERAERDTS

Lakoff concluded based on these definitions that past critics of linguistic relativity had based their criticism on novel definitions of linguistic relativity. According to him this rendered their criticism superficial.

Up to this point we have mostly seen the broad general way linguistic relativity has developed through history. In this last part I want to focus more on some specific cases and thoughts I have about linguistic relativity. Beginning with its influence on constructed languages and literature. Because there are many instances where authors have used language – natural or constructed – in their stories. One of the best examples of this is how George Orwell showed how linguistic relativity might be exploited for political purposes. The authoritarian state in his novel 1984 created a language Newspeak which made it impossible for people to criticize them (just like Atlan, Newspeak also has some Olig synthetic features: see chapter 5.1.1). Another example is Rand's *Anthem*, a story about a dystopian communist society who erased the word "I" from their language to erase individuality. Ideas like this illustrate not only the possibility of language on us but also the fact that we can think about language in this way. The fact that we can imagine it

CHAPTER 1. INTRODUCTION

having such an influence on ourselves speaks volumes.

Looking back in history we can see the influence language has had on us and our actions. Book burning illustrates this perfectly. The earliest occurrence dates back to 600 BC. Maybe most famous example coming from the 1930's and 40's when the Nazi's burned countless Jewish books. The Nazi's sought to erase Jewish culture and saw burning their books which were written in their language about their culture to be necessary. While terrible, it does illustrate that language is inseparable from culture. Seeking to eradicate one demands eradicating the other. Which in turn means that creating one requires creating the other. Linguist and author J.R.R. Tolkien did exactly this when writing stories set in Arda, the most famous of those being *The Lord Of The Rings* and *The Hobbit*.

Others sought to create a language to enable a higher level of cognition. They believe that by speaking a new – better – language humans can reach higher levels of thought. One of these languages is Loglan and its evolution Lojban. This conlang is extremely logic based. They seek to be as logical as possible. The creators wanted to use it to test whether linguistic relativity exists. Because

the language is entirely based on logic, they thought that it would make its speakers think more logically. Speakers of Lojban reported that they did feel like they thought more logically when speaking Lojban (Nicholas, 2003). Another example of how language can influence our thoughts in a specifically directed way. Another linguist who sought to do this using her Conlang is Suzette Haden Elgin. She has invented the language Láadan which according to its creator makes it easier to express a female world view. Elgin argued that SEA languages have a male centered world view. Using linguistic relativity, she sought to counter this using language. The Toki Pona language was created with the same intent. Its creator – Sonja Lang – wanted to create a simple universal language which focused on happy thoughts. It quite literally aims to make its speakers happier (Lang, Sonja). Because of its simple nature (having only 123 words total), however, it cannot be used to express more detailed or complex meaning: its word for ‘complicated’ is even the same as the word for ‘bad’, *‘ike’*. We once again see that language can have a directed influence on our thoughts. It is not a stretch to pose that we are all confined by our language. It is our way of expressing our thoughts, desires and feel-

CHAPTER 1. INTRODUCTION

ings. The following quote by Von Humboldt illustrates this beautifully:

“...there resides in every language a characteristic worldview . . . By the same act whereby [man] spins language out of himself, he spins himself into it, and every language draws about the people that possesses it a circle whence it is possible to exit only by stepping over at once into the circle of another one (Von Humboldt, 1988).”

Throughout this chapter we have seen the evolution and creation of linguistic relativity. We have seen that it is a difficult topic to pin down and reach consensus on. We have, however, also seen that it does have a remarkable effect on our thinking and understanding of the world. All the way from colours to how we feel. We have seen that we can create languages to infuse its speakers with a certain world view. The power of language is evidently not to be underestimated and we can only guess at the future. Will there be one universal language one hundred years from now? Is one universal language desirable? One way or another, language has and always will be an integral part of our being. For without it we are left soulless.

Language & culture

Language and culture have long been inseparable. They influence each other and evolve alongside each other. Culture needs language and language needs culture. Learning a new language has made this painfully obvious to me. At one point you figure out that it is not sufficient to just learn the meaning of a word according to the dictionary. To then use grammar to construct sentences. Language is more intricate; words can mean one thing in each context only for another context to change its meaning to the complete opposite. Some words are not even in the dictionary. Some words have an entirely different meaning than the one stated in the dictionary. The meaning of some words changes dramatically over time. Some might even say that depending on which language we speak our view of the world can change. Not only this but language grows over time, it is never in a stable state. As is our world and culture.

We can see clearly that culture influences our language. When culture changes our language changes with it. A good example of this is Dutch ‘straattaal’- literal translation; street language – an unofficial dialect spoken by the youth subculture in the Netherlands. ‘Straat-

CHAPTER 1. INTRODUCTION

taal' consists of mostly normal Dutch words and sentence structure, it has however a few exceptions. It introduces new words and ignores grammar in certain situations. Hereby marking itself different from standard Dutch and dominant Dutch society. This diversion is not an accident. These youths don't want to be a part of mainstream or 'adult' society. They seek to define themselves, creating their own language plays a big role in this. It creates a very strong in-group – people who speak the language and can communicate with each other – and a very distinct out-group – people who do not speak the language – this helps in creating subcultures. The fact that a lot of subcultures have their own variations of the language of the dominant culture marks the importance of language in society and the bilateral relation between language and culture.

This makes it difficult to imagine a language without any culture attached to it. This is, however, one of the goals we have with Atlan. We want to create a language that is as universal as possible. We cannot have one dominant culture associated with Atlan as this would result in a bias for people from that culture. In this chapter I want to explore by what we really mean when we say

universal language and what our vision is of the culture that could be attached to Atlan in the future. Because a language without any culture is impossible.

Culture

As you have read in the introduction our goal with Atlan is to create a universal auxiliary language. Not based on one country, culture or region but based on human experiences. I however believe – as implied in the introduction – that a language is impossible to exist without an attached culture. I view language as I do the chicken and egg dilemma. It is impossible for one to exist without the other also existing. This conclusion seems like a problem for Atlan. Our goal is to create a universal language but at the same time it is impossible for a language to be without culture. And therefore, it is also impossible for a language to be without biases. I, however, believe to have found two possible solutions to this problem. The first option is to accept that Atlan has no culture and therefore is not a proper language. This might seem like a shocking conclusion, and I will elaborate on it later. The second option is to attempt to create a new culture attached to Atlan. A culture based on human experiences.

Language without culture

The first option I want to discuss is the language without culture option. Seeing how I have stated earlier that I believe a language without culture is impossible you might be confused by this option. Let me explain what it precisely is I mean when I say language without culture. This option originates from the dilemma of making a universal language. For this to be true it cannot depend on a culture. If it did it would not be universal anymore. But it is also true that without a culture Atlan cannot be a language. I will not go into detail on the precise conditions something has to satisfy to be a language, but I will conclude that having an associated culture is one of them. A result then of the decision to not have any culture associated with Atlan is that Atlan is not a language. Now of course it will still satisfy a lot of the conditions of being a language. It can be spoken and written, and its main purpose is to communicate with other people. But it will not be a language like we know them. It will not have a culture. It will not have a country where it is the official language of. It will not have a history. It will in some sense be more like a computer language. It will not naturally evolve over time it will instead receive

updates when deemed necessary.

This might feel like it makes Atlan a very cold and empty thing. Which it does. I think, however, that for the purpose we devised for this language this is a necessary sacrifice. Atlan will be a universal language, used for communication between people who speak different languages. Atlan does not need to be a language as we know it today because it will fulfill a different purpose. It is okay for Atlan to not have its own culture, history, country and people because we already have enough language who have those things. Atlan will be used as a worldwide communication language; it is allowed to be cold and lifeless. For those languages with identity already exist and will continue to exist in the future. The purpose of Atlan will be to bridge the gap between these cultures. It will be cold and cultureless for everyone; this will make it an even playing field for all those who speak it. This would be the main future I see for Atlan without culture.

Language with many cultures

If you strongly disagree with the idea of language without culture you are lucky because I have a second option, language with many cultures. It is impossible for Atlan

CHAPTER 1. INTRODUCTION

to have one culture because it would be biased towards that culture. It would create an ‘in’ and ‘out’ group, a fatal flaw for a supposed universal language. To avoid this problem, we could have many cultures associated with Atlan. This would create many groups who all have their own variation on Atlan. They can understand each other but they will also each have their own identity. This way Atlan can be used to communicate internationally but it will also have an identity, culture and history. In fact, it will have many different ones. This would create the option for different countries/cultures to develop their own version of Atlan with which they will build their own history with. Of course, these variations cannot be too big otherwise these different groups will not be able to understand each other anymore. But apart from this restriction this solution offers a much more alive version of Atlan than the previous one.

An obvious argument against this option would be one that argues for one universal culture that is associated with Atlan. This seems like the perfect solution. Atlan will have an associated culture and it will be a universal one. Thus, not excluding anyone and maintaining its purpose as a universal language. I don’t think this

is possible unfortunately. Creating one universal culture is a worthy ideal but I am afraid it is not yet possible. As I have said before a culture creates ‘in’ and ‘out’ groups. I believe that culture not only creates these groups it needs them. It originates from them. We can see throughout history that a common enemy brings people closer together. This is also the case for culture. The effect of trying to create a universal group with everyone is impossible. There needs to be some sort of ‘out’ group. The effect of this choice would be very similar to the language with many cultures’ choice. There would be many variations in Atlas, all of them associated with their own culture. It is best to have this view from the very beginning of Atlas’ journey, giving the speakers this freedom rather than having them take it.

I see these as the two possible solutions for the problem I stated in the introduction. It remains a fact that a universal language cannot have a culture. It would not be universal anymore. Having a language without a culture would mean it will not be a language anymore. This poses a problem. I offer two possible solutions for this problem. The first one is to create a language without a culture. The second solution is to create a language with

CHAPTER 1. INTRODUCTION

many cultures. These solutions are extremes on the same spectrum. I don't know which of these solutions is the best solution. I do think that they both solve the problem albeit in their own way. They would have massively different consequences for Atlan in the future. I look forward to seeing how Atlan will develop in our society in the future.

Chapter 2

Phonology – Niek Elsinga

2.0 Introduction

A language is a system through which an individual can communicate with others, which is structured in grammar and vocabulary. Languages are usually spoken, but can also be conveyed by signs as with sign language, or with script. The definition of language is quite a contested topic. Multiple theories about the purpose of language have been proposed. One of the first definitions of languages was put forward by Ferdinand De Saussure. De Saussure saw language as self-contained, self-regulating system, in which the elements are characterised by their

relationship with other elements in the system. De Saussure named his vision on linguistics ‘semiology’, but this was posthumously named structuralism by other linguists (Matthews, 2014).

Nowadays, linguistic scholars deem the structuralist approach outdated, and favour more recent explanations. While some linguistic scholars such as Noam Chomsky and Steven Pinker see language as a biological, formal, or ‘mathematical’ system of signs that are dictated by grammatical rules to convey an utterance (Chomsky, 2002; Pinker, 1994), other scholars such as Nicholas Evans pursue the more ‘functional’ approach and see language as a system of communication that allows for the exchange of utterances (Evans & Levinson, 2009). One other view sees language primarily and purely as a ‘tool’ that can be used for humans to undertake linguistic behaviour, in that language is solely a means of producing and understanding utterances that evolved over the course of human history (Fitch et al., 2005).

Note that these definitions more or less convey the same meaning: “a system through which an individual can communicate”. The difference between these views is not so much what language is for, but what it em-

phasises. They are not mutually exclusive to a certain degree. Nonetheless, contemporary scholars predominantly adopt Chomsky's biological approach. However, even this view has been contested, on the grounds that neuroscientific studies have found neither biological nor neurologic evidence for the existence of Chomsky's theory on the application of WH-questions, i.e., what, where, when, who(m/se), why, which, and how (Kluender & Kutas, 1993).

English is still the most spoken language of academia worldwide, and the *lingua franca* of the western world (Mauranen, 2003). It has not, however, gained this position because it is easy to speak or learn. Pronunciation of English vowels, for example, is unlike its graphemic notation, due to phonological shifts of vowels after the standardisation of English spelling in the 15th and 16th centuries (Denham & Lobeck, 2007). English did not gain its position because of the purported absence of cultural influence of English, as stated by Knapp and Meierkord (2002). English fulfils the need of a global lingua franca, as it has spread to large areas of the world due to various factors. These include the adoption of the Latin script worldwide, the invention of the internet and its

first widespread use in the United States of America.

The development of the American research university and subsequent adoption of English as *the* academic language have also been of tremendous importance its widespread use. However, there exist more sinister factors as well, such as widespread colonization brought about by the British, American cultural hegemony, and the spread of Christianity through western missionaries (Ariza & Navarro, 2006). The use of English in academic language has long been postulated by some to be ‘neutral’, i.e., free of cultural influences (House, 2003). However, as of late this claim has been challenged. Scholars such as Pölzl and Seidlhofer (2006) and Knapp and Meierkord (2002) have claimed that English is ‘imperialistic’ by definition due to the use of English by colonists. These colonists subsequently decreed that English would be the sole language to be spoken in countries which do not have English as its endogenous language, and as such was seen as a form of oppression (Macedo & Bartolomé, 2014). Other scholars have presumed that English can be ‘neutral’ to a certain degree, and that it is up to the speaker of a language to give partiality to one’s words and actions (Norton, 1997). If this view is mirrored

against the notion of the impartiality of language and that language and culture are interwoven to their very core as famously articulated by Kramsch (2014), it is possible to surmise that any language that has evolved naturally in humans through use and repetition without conscious planning or premeditation is intrinsically biased, due to the fact that culture and language are inherently linked (Lyons, 1991).

Atlan is designed to be an auxiliary constructed language, a language that is created with the purpose of facilitating communication between people who have different native languages. This decision has been made because we are of the opinion that a language that is used in academic context should be neutral. This does not imply that the language shall solely be used for academic purposes, nor does it mean that it should replace other languages.

With the creation of the language, multiple goals have been kept in mind. The primary purpose in the creation of a language is to be as culturally neutral as possible, so that no group of people will be especially favoured or disfavoured when learning the language with regards to the similarity to their own. Creating a language from

scratch can procure this cornerstone.

Another main goal is that the language should both be easy to speak and understand. The notion of unambiguity is another tenet, with the goal of reducing confusion or misinterpretation within communication as much as possible. This means being as sparse as possible, with different elements of the language, where simplicity is key, and complexity should arise from the combination of the basic elements. This is, of course, of utmost importance in phonology and morphology. If a differing consonant is used, it would change the entirety of the word. The same applies to morphology, where the distinction needs to be made between who the actor and who the recipient is.

This paper will serve as an overview regarding the phonological and morphological considerations that have been made for the language. In the first section of this paper, I will elaborate on the neurology concerning speech and language. The second section will cover the choices that have been made regarding the phonology for consonants, vowels, and prosody. Finally, I will close this paper by summarising what has been stated, and giving some concluding remarks.

2.1 The Neurologic Basis of Language

Neurolinguistics is the study of how the brain produces, comprehends, and acquires language. It combines both the framework of humanities, namely the language aspect, with a neuroscience approach. The two traditional brain areas that are correlated with the production and comprehension of language and speech with respectively Broca's area in the frontal lobe, and Wernicke's area in the temporal lobe (Geschwind, 1972), which are connected through the *fasciculus arcuatus* (Bernal & Altman, 2010). These areas are not bilaterally localized, and solely exist in the left cerebral hemisphere.

The production of speech occurs according to three main principles: conceptualization, formulation, and articulation. In the first stage, conceptualization, an individual with the intention to create speech links the desired concept to the particular spoken words. This preverbal message contains the desired to-be conveyed thoughts to be expressed. The second stage is formulation, in which the linguistic form for the desired message is formulated. Here, knowledge of grammar, phonology, and phonetics is applied to the preverbal message. The third stage

is the articulation of the message, in which motor functions are activated to produce the utterance.

The perception of language or speech begins at the level of the sound signal and the process of audition. Subsequently, speech sounds are further processed in order to gain information regarding acoustic cues and phonetics. This information can then be used for processes that are considered to be ‘higher-level’ language processes, such as word recognition (Levelt, 1999). These produced sounds are then further processed in the auditory cortex of the brain.

Research has indicated that the auditory cortex processes voiceless and voiced phonemes differently in ferrets, which have similar structures in the processing of auditive information when compared to humans (Mesgarani et al., 2008). Phonemes are, put very simply, sounds, or the smallest units of speech. Phonemes are usually divided into consonants and vowels (Yallop & Fletcher, 2007). Consonants are created by constricting the air-flow in the vocal tract when air is forced out of the lungs, and is mostly done by the tongue.

Some consonants can also be created by, among others, the nose and vocal tract. Voiced consonants are con-

2.1. THE NEUROLOGIC BASIS OF LANGUAGE

sonants that incorporate the vibration of the vocal cords when the articulation of the letter occurs. Some examples of voiced consonants are the /b/, /d/, and /g/. Voiceless consonants on the other hand do not make use of the vocal cords. Examples of voiceless include /p/, /t/, and /k/. Some languages, such as Arabic, do not have the voiceless bilabial plosive /p/ in their phonological inventory (Al-Ani, 1970). When a speaker of Arabic wants to say the word ‘pizza’, they would pronounce it as ‘bizza’, for the voiced bilabial plosive /b/ is used instead of the /p/. If an Italian on holiday in an Arabic-speaking country would order a pizza, pronouncing the word with the voiceless bilabial plosive /p/, a monolingual speaker of Arabic would not have any hindrances whatsoever with the comprehension of the utterance (Versteegh, 2014).

This can be linked to another research by Liégeois-Chauvel et al. (1999) on the inquiry of the perception of voiced and voiceless phonemes. In this research, a speaker produced voiceless and voiced phonemes, with the following vowel being /a/ (/pa/, /ta/, /ka/ for voiceless, and /ba/, /da/, /ga/ for voiced) in a random order. Neurologic tests were carried out using a tool called ‘electroencephalography’ (EEG). An EEG maps where in the

brain electrical pulses occur, i.e., where and which areas of the brain are activated when an individual is exposed to stimuli. The EEG has shown that the auditory cortex is able to process syllables with voiced consonants from syllables with voiceless consonants in the left hemisphere, however, the right hemisphere was not able to make this distinction and solely processed acoustic stimuli. Furthermore, the auditory cortex was not able to differentiate syllables with voiced consonants and voiceless consonants. The results from the EEG showed no discernible differences between syllables with voiced and voiceless consonants. However, a differential coding of voiced and voiceless syllables is preserved. This would still mean that an individual is able to distinguish these phonemes (Liégeois-Chauvel et al., 1999).

2.2 Consonants in Atlan

As previously stated, the word ‘pizza’ would be pronounced as ‘bizza’ by according to Arabic phonology (Al-Ani, 1970). The example also states that ‘pizza’ and ‘bizza’ would both be understood as the same word. This is because in Arabic, the ‘b’ and ‘p’ are variants of the same phoneme.

This is called allophony.

Furthermore, certain languages (or language families) use scripts that do not implicate the voicing of a consonant, such as Tamil. Tamil uses both voiced and unvoiced consonants, however, it is decided by context (e.g., a -linguistic- register), and not by its script. A consonant being voiced or unvoiced does not imply that a word gets a whole new meaning, but gives meaning to the context of the word. Consonant voicing thus is not contrastive in Tamil (Keane, 2004; Schiffman & Arokianathan, 1986). Regarding the phonology of our language, the decision has been made that both voiceless and voiced consonants are allophones. For example, a speaker of our language would perceive both the voiced bilabial plosive /b/ and voiceless bilabial plosive /p/ as the same phoneme.

The script is meant to reflect this, as is the case with Tamil. Furthermore, because not every language has the same set of phonemes nor the same number of phonemes, we have decided that nine distinctive categories should be made. The phonemes that belong to each respective category are allophones in our language. The categories in this were chosen according to mutual intelligibility, proximity according to the consonantal chart of the In-

ternational Phonetic Alphabet, and manner of articulation (Ladefoged, 1999). Furthermore, consideration has been given to the frequency of each phoneme and its subsequent category. Every category contains a phoneme that has a high rate of frequency in languages worldwide. In order to retrieve the information regarding salience of the phonemes, the UCLA Phonological Segment Inventory Database (UPSID) and the Phonetics Information Base and Lexicon (PHOIBLE) were used (Maddieson, 1984, 1986; Moran & McCloy, 2019). These databases document the frequency of every existing phoneme.

The categories are as follows: The first category contains the (bi-)labial plosives [b, p]. The bilabial plosives are found in 98.89% of all languages worldwide according to UPSID. The second category consists of the coronal plosives, i.e., the dental, dento-alveolar, alveolar, and retroflex plosives, [t, d, ʈ, ɖ]. The coronal plosives are found in almost every language according to Liberman *et al.* (1967), however, no exact percentage is given regarding its frequency. The third category contains the dorsal plosives and dorsal fricatives [k, g, q, ɢ] and [x, ɣ, χ, ʁ]. The dorsal plosives and fricatives are found in 99.30% of all languages worldwide according to PHOIBLE and

phonemes are found in 84.49% of languages worldwide according to PHOIBLE and UPSID. The eight category are the coronal sibilant fricatives [s, z, ʃ, ʒ, ɬ, ɮ, ɕ, ʑ, ʈ, ɟ]. According to UPSID, these phonemes are found in 88.03% of languages worldwide. The ninth category consists of the palatal consonants [ç, j, ɟ], which according to UPSID and PHOIBLE are found in 90% of all languages. A tenth quasi-category was made for glottal and pharyngeal consonants; however, we have decided to give these phonemes no meaning.

Atlan also employs a glottal stop [ʔ], however this sound is not notated in its orthography. Rather, it functions to differentiate two of the same vowels when placed next to one another. For example, ‘KA.AK’ could be confused with ‘KAK’ if there is no pronounced distinction between the two syllables, therefore the former should be pronounced as ‘KAʔAK’.

2.3 Vowels in Atlan

Categorising the vowels was considerably more difficult, considering that vowels cannot be placed on an axis of ‘place of articulation’ and ‘manner of articulation, as is

the case with consonants (Ladefoged, 1999). Vowels can be placed on a spectrum, with one axis from ‘close’ to ‘open’, and another from ‘front’ to ‘back’. The close-to-open axis refers the position of the tongue placed against the roof of the mouth. ‘Close’ in this context means that the tongue is positioned as close as possible to the roof of the mouth as it can be without creating a constriction, whereas ‘open’ means that the tongue is positioned as far as possible from the roof of the mouth. The front-to-back axis refers to the position of the tongue in the mouth. ‘Front’ in this context means that the tongue is positioned as far forward as possible in the mouth, ‘back’ means that the tongue is positioned as far backwards as possible in the mouth (Yallop & Fletcher, 2007).

Vowels considered to be close-front include [i] as in the English word ‘free’ and the Dutch ‘vieren’, and close-back include [u] as in the Dutch ‘voet’ and the English ‘boot’. Open-front vowels include [a] as in the British English ‘hat’, and open-back include [ɑ] as in the Dutch ‘bad’ (Gussenhoven, 1992; Roach, 2004).

Because the quality of vowels is a spectrum and not every vowel exists in every language, a certain degree of allophony exists in vowels as with consonants. In In-

donesian, [i] and [u] are allophones of /i/ and /u/, while in Dutch they are contrastive (Gussenhoven, 1992; Soderberg & Olson, 2008).

For our language, five categories of vowels were made. As with the consonants, these are based on the salience of the vowels and its frequency in languages worldwide. The data regarding this is based on the same tools as for consonants; UPSID and PHOIBLE. Three of these categories were easily made because most languages contain this respective vowel. These are, from high to low frequency, [i, u, a], with respectively 92, 88, and 86% occurrence in languages worldwide. For the two remaining categories, a substantial lower frequency is noted for [e, o], with respectively 61 and 60% of the languages worldwide containing the vowel according to UPSID and PHOIBLE.

These five categories were chosen because these five vowels are found in every language, and the frequency of the vowels [e, o] were found in roughly the same percentages in language families worldwide, with the exception of (some) Australian languages (Butcher, 2018; Moran & McCloy, 2019).

Another extra vowel is used in our constructed lan-

2.4. DECISIONS REGARDING THE PHONETICS: TONE AND PROSODY

guage, namely the schwa [ə], but this vowel is not notated. Its function is to differentiate two of the same consonants that occur next to one another, similar to the use of the glottal stop. For example, in spoken Atlan, the difference between 'AK.KA' would be barely distinguishable from 'A.KA', therefore the former would be pronounced 'AKəKA' to retain the distinction.

2.4 Decisions regarding the phonetics: tone and prosody

In some languages, tone (i.e., the use of pitch), is a meaning distinguishing feature. For example, Mandarin is a tonal language and depending on the pitch or variation in pitch, the word 'ma' can have five different meanings, such as horse, mother, scold, or as a marker for a question (Lee et al., 1996). Pitch can be as important as vowels themselves for comprehension of words and grammatical functions. Following on from the previous section, we decided that phonetic properties, such as tone, have no semantic nor pragmatic value. Prosody has no intrinsic value either. Prosody consists of intonation and rhythm. Intonation are the changes in pitch

used for, e.g., conveying the speaker's attitudes and emotions, or to highlight or focus an expression. The concept of rhythm in language is dubious at best, and is perhaps better explained by the notion that the perception of rhythm is based on the language that an individual already speaks, and is thus irrelevant to precisely define (Arvaniti, 2012). Likewise, lexico-semantic characteristics are solely conveyed through phonemes, not the intonation or pitch of said phonemes. This is due to the fact that only a select few languages make use of pitch differences for semantic meanings, such as Mandarin, Cantonese, Vietnamese, Yoruba, and Navajo (Bauer & Benedict, 1997; Yip, 2002).

2.5 Morphology in Atlan

What is the difference between 'walking' and 'walked'? Both words convey the meaning that something or someone is moving at a regular pace by lifting and setting down each foot in turn, never having both feet off the ground at once. However, while they convey the same meaning in movement, saying: "Stijn walking there yesterday," would be incorrect, as is "Jep is walked there".

The difference here is the suffix: a morpheme added at the end of a word to form a derivative (e.g. -ation, -fy, -ing). A morpheme, and its subject of study called morphology, examine the smallest meaningful units of language, which can be individual words or parts of words (Matthews, 1991). The main goal of morphology is to understand how words are constructed and how they convey meaning. It analyzes the various types of morphological processes, such as affixation, where morphemes are added in, around, before, or after a word, or compounding, where two or more words are combined to form a new one, and inflection, in which the form of a word is altered to indicate grammatical information like tense, number, or gender (Booij, 2007).

Some languages make heavy use of these morphemes for these context-related factors, other languages do not. Languages exist on a continuum in regard to morphology, but can be more or less categorised. On one end of the spectrum exist isolating or analytic languages, in which words are composed primarily of individual morphemes that are each distinct and carry a specific meaning, such as Vietnamese (Comrie, 1989). In isolating languages, each morpheme generally corresponds to a spe-

cific concept or grammatical function.

On the other end of the spectrum, there are polysynthetic languages, in which words are composed of multiple morphemes that are fused together to express complex ideas and convey a wealth of information within a single word. In polysynthetic languages, a single word can contain a combination of roots, affixes, and grammatical markers, allowing for the expression of entire sentences' worth of information (Baker, 1998). In polysynthetic languages, the process of word formation involves extensive morphological affixation, compounding, and incorporation. An example of a polysynthetic language is Nahuatl, spoken in Mesoamerica before the colonization by Spanish conquistadores (Rolstad, 2001; Suarez, 1983). Somewhere in this continuum, agglutinative languages exist.

In agglutinative languages, morphemes are typically added to the root or stem of a word to express various grammatical features such as tense, aspect, mood, number, case, and person. Unlike other synthetic languages like Nahuatl, agglutinative languages maintain a one-to-one correspondence between morphemes and specific grammatical functions. Generally, agglutinative languages

have a great degree of transparency in their morphological systems. The term “transparent” means that the relationship between the morphemes and their meanings is relatively straightforward and predictable. The affixes are typically added in a consistent and regular manner, allowing for clear distinctions between different grammatical features (Durrant, 2013). A wonderful example of an agglutinative language is Turkish (Lewis, 2001).

Atlan, like Turkish, makes use of an agglutinative system for morphemes. One of the core elements of Atlan is that unambiguity is a prerequisite. In choosing an agglutinative system for morphemes, we are of the opinion that this keystone has been achieved.

2.6 Concluding remarks

In this paper, we explored the neurologic basis of language and discussed the phonological considerations for the creation of an international auxiliary constructed language. We highlighted the interconnectedness between language and the brain, and the subsequent choices regarding the phonetic, phonological, and morphological system.

Language is a complex system that is not merely a tool for communication but a reflection of our culture and identity. While constructing a language that is completely devoid of bias may be challenging, striving for neutrality and inclusivity is a worthy endeavour. The creation of a neutral and accessible language has the potential to enhance global communication, foster cultural exchange, and promote inclusivity. While language will always carry cultural influences, our efforts to create a more neutral language reflect our commitment to open dialogue and mutual understanding in an increasingly interconnected world.

Chapter 3

Writing Atlan

3.0 Writing system

Jarno Smets

ATLAN'S writing system is a natural application of our philosophy: start with elementary parts, and then every complexity shall be a mere combination of those parts. Our glyphs (as we shall call them) each denote one syllable. This is always the case: they will always stand for the *same* syllable. Unlike English: in the words “tone” and “to”, the “to” is pronounced respectively [tu] and [tə].

That is the rationale behind our writing system; let us dive into the details. As told, Atlan has a set of basic

CHAPTER 3. WRITING ATLAN

lines. They are:

Consonants		Vowels	
Line	In I.P.A.	Line	In I.P.A.
	/t/	⊂	/u/
—	/k/	⊃	/i/
/	/n/	⊄	/a/
\	/m/	⊅	/o/
⌒	/j/	○	/e/
⌑	/s/		
⌐	/f/		
⌑	/l~r/		
•	/p/		
○	<nil ¹ >		

Figure 1: The basic lines of Atlan's writing system.

¹When you see this hollow circle, the other line is combined with nothing. Don't panic if you don't yet understand this; it will be explained shortly.

3.0. WRITING SYSTEM

These lines all represent a single vowel (V), or a single consonant (C). We can combine them to make syllables. By combining two consonant lines, you get a CVC-syllable, such as *loj*, *pas* or *mup*. You can also make a VC-syllable, such as *mu*, *po*, or *ji*. The vowels don't have separate lines in a CVC or VC-syllable; instead, the vowel is determined by the position of the two consonant-lines. We will go deeper into that below. First, we give the rules for the order of the consonants and vowels: what determines whether two lines make e.g. *poj* or *jop*, *mu* or *um*?

This order is determined by the manner in which the lines combine. There is always a “bigger ”line, and a smaller one. Rule of thumb: the bigger line usually is the most vertical of the two. These lines fit inside an imaginary box. The position of the smaller line relative to the bigger line, determines the order of consonants. A visual of this rule helps:

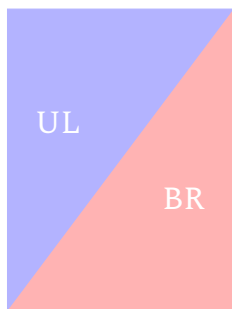


Figure 2: Box for determining consonant order.

If the smaller line is in the upper-left triangle (UL), the consonant it designates comes first. If it is in the bottom-right one, it comes second. For the rest of the explanation, it is advised to keep this box in the back of your head. An example:



As you see here, the smaller line is found on top. Hence, it is placed inside the upper-left triangle. The consonant for which the smaller horizontal line stands (the *k*), comes before the other consonant, the *s*.

3.0. WRITING SYSTEM

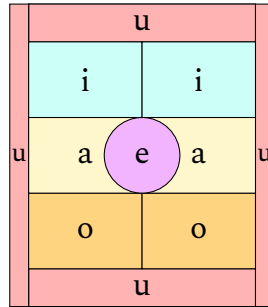


Figure 3: Location of the smaller line in relation to the vowel.

The vowel is...

- *u* if the smaller line is found at the edges. The smaller line is in its whole above, under, left or right of the main line.
- *i* if the smaller line is found on the upper-left or upper-right side of the main line. It is usually smaller than the line made for *u*, to avoid confusion.
- *a* if the smaller line is found left or right to the middle of the line.
- *o* if the smaller line is found on the bottom-left or bottom-right hand of the main line. Again, this line is smaller than the line for *u*.

- *e* if the smaller line is placed in the middle. Or, if the small line intersects with the main line at the middle. In some instances, the small line is then split up by the main line.

Then we have three exceptions to these rules. The first: you can combine two of the same letter-lines, to make syllables such as *pop*, *mum*, or *lol*. The order of these lines doesn't matter; hence we choose place the smaller line to the upper-left of the main line in such cases. For the vowel *u*, there are two small lines, split at the center. For *e*, there are either two or three small lines. At least one of those lines crosses through the center of the imaginary box.

The second exception has to do with the *k* (/k/). That letter had the base line – , right? It's a horizontal line. Because of that, we have to think of a different box than figure 2 and 3 to figure out the consonant-order and the vowel. The solution is simple: we flip the box. It then looks like this:

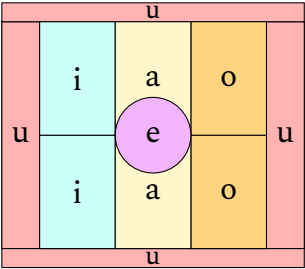


Figure 4: the imaginary vowel-placement box for /k/.

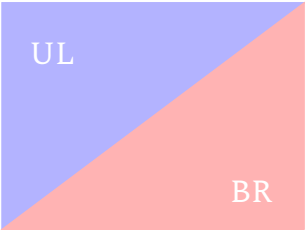
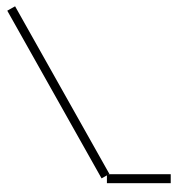


Figure 5: the imaginary consonant-order-box for /k/.

The third exception: Remember that the *p* is represented by the dot •. For clarity, we couldn’t combine simply two dots to make a full syllable. Hence, two *p*-dots combine a bit different from the rest of the lines. The *p* also can’t combine well with the circle (which designated “nothing”). They combine in the following way:

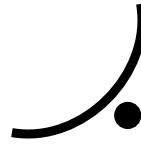
Basic line		u	i	a	o	e
	•	⦿	⊗	◁	▷	⊗
•	•	◊	△	◁	▷	△
•		⊕	▽	▷	◁	△

These were the rules for the script of Atlan. It might sound a bit cryptic, so let’s discuss some examples. If you still feel uncertain whether you understand the rules, read through them again. Personal experience tells that, after some time, recognizing letters gets more intuitive.



Let’s dissect this letter. This is the letter “mok”, or (/mok/) phonetically. First step is to discover the main line, which is the long diagonal here. This is the consonant *m* (/m/). Then there is a smaller line, found in the bottom-right corner. This is the *k* (/k/). The horizontal line is in the bottom-right of our imaginary square. Hence, the *m* comes before the *k* (see also figure one). We got the two consonants, now rests the vowel. Feel free to look back at figure two. The smaller line is found in the bottom-right corner, hence the vowel here is an *o* (/o/). The full syllable is *mok* (/mok/).

Now let's look at another one. See if you can determine the syllable yourself first. The main line is obvious: it's the big curve. This big curve is a j^2 ($/j/$). The smaller dot is a p ($/p/$). The dot is found inside the quadrant



“UL” of figure one. Hence, the dot comes first. The dot is found a bit to the left of the centre of our imaginary box. Hence, the vowel here, is the a ($/a/$). The full syllable is jap ($/jap/$).

Do you feel if you got the hang of it? Let's do a few more. To spice things up a bit, we'll have a syllable with the vowel e . Remember that this vowel had smaller lines be placed in the centre. Alternatively, the



smaller line could intersect the centre, or be split up by it. In this example, the smaller line is split up by the bigger line. The bigger line, the l ($/l/$) splits up the line for j

²Quick tip: the curve for j looks alot like the j itself, doesn't it? Look for more of these similarities in our writing system; they help!

(*/j/*). Because it does, the vowel is *e* (*/e/*), and the syllable is *lej* (*/lej/*).



Now the last example. This, we think, is the best-looking glyph in our catalogue. What does it stand for? There aren't two, three, or four separate lines here, as should be. Instead, there is a triangle with a circle inside. What do we do? Well, remember the *p* (*/p/*), which was a dot. And remember that "nothing" also has its line: the circle. There was an exception for when two p-dots combined, or a p-dot and a circle. The exception was explained a few pages back. If you go there, you encounter the same glyph. This syllable is the *pe* (*/pe/*). A tip for remembering these glyphs: if you see a glyph with a triangle and a circle, think of the *p*.

We hope the examples have made clear how our writing system works. This concludes the explanation of our writing system for syllables. Upcoming is our writing system for numbers, and for names. Before we get to the next part, a few words of advice for learning the writing system:

- On the next pages, a full list of our glyphs is added. They are 490 in number; as many glyphs as we have syllables. Don't be intimidated by the list; instead, use it wisely. Look through the list, and try to grasp the pattern of formation. Read the explanation above, and try to get a feel of how our glyphs are formed. Again: after some while, you'll have a stronger intuition.
- Try drawing some of the glyphs. It helps for getting used to the glyphs. You don't need a ruler to draw them; just make sure they can be distinguished from each other.
- Any blank space is the "nothing" we talked about above. You can see the circles appear at the rows of those blank spaces. If the half of an entire row is empty, it means that a combination's reverse is unnecessary. Think of syllables like *lol*, *nun* or *juj*. They are symmetric, so we don't need a full row for them.

Again, on the next page is a table containing all our glyphs. The left two columns contain the base lines, placed in order of combination. If you've forgot which base line























































































































































































































































































CHAPTER 3. WRITING ATLAN

stands for which consonant, return to the table at the begin of the chapter. It's a great tool for using as a reference. So come back when you need it.











































































































































































































3.0. WRITING SYSTEM

[illegible]

CHAPTER 3. WRITING ATLAN

e																				
o																				
a																				
i																				
u																				
Base lines																				
																				
e																				
o																				
a																				
i																				
u																				
Base lines																				
																				


3.0. WRITING SYSTEM


e															
o															
a															
i															
u															
Base lines															
															
e															
o															
a															
i															
u															
Base lines															
															

3.1 Direction of reading – Stijn Janssens

Out of the roughly 6.500 languages worldwide, only 12 languages read right to left. The biggest among these is Arabic, with about 1,7 billion speakers. Preferred reading direction has to do with the materials on which the language was historically written, correlating with the technique used to write or carve letters. However, since LtR languages form the majority, and writing materials don't form an obstacle anymore in our modern world, Atlan will use this reading direction as well. Nearly all human languages on earth are written in rows, stacked from top to bottom, and thus Atlan will do the same. This will also make it hospitable to digital environments and graphic design. Because the syllable glyphs can be read all at once, they could also be stacked in vertical columns, reading from top to bottom, for example when employed in calligraphy, or writing along pillars or vertical ridges.

3.2 Punctuation

Atlan has minimal punctuation, only having dedicated symbols for a comma and a full stop, and spaces are the same as in any other orthography. A comma is notated as a small half circle which is open at the top:  symbolizing an 'open' continuation of the sentence, and the full stop is notated as a half circle which is open at the bottom:

, symbolizing a closed sentence.

Other punctuation will be marked by Atlan's semantic atoms: question sentences start with the interrogative particle E, and so this eliminates the need for a question mark. Exclamative sentences start with the particle O, eliminating the need for an exclamation mark. Other examples would be '&' being 'AN' ('and'), '%' being 'EP.NO' ('per hundred'), ':' being 'I' ('relative clause') etc.

3.3 Transliteration

Atlan words will be transliterated into the Roman alphabet using the archetype letters U, I, A, O, E, P, T, K, M, N, S, F, L, J. Dots are used in between each syllable, in order to prevent confusion about where syllables are broken

up, since this could create ambiguity in meaning. Dots in between two of the same consonants (eg. AK.KA) or vowels (e.g. KA.AK) are pronounced as a glottal stop or shwa, respectively (see chapter 2.2 and 2.3).

Atlan's syllables are all (C)V(C). Some loanwords or names, however, might have two or more consonants in a row within the same syllable. In such cases, the individual letter lines that exceed the CVC limit, will stand on their own next to the syllable glyph. The name 'Stijn', for example, will then become 'S.TEJ.N'.

3.4 Numerals

Atlan also has a numeric system distinct from the familiar arabic-numerals. They look like this:

3.4. NUMERALS

1 —┐	10 ┐—	100 —┐	1000 ┐—
2 —┐	20 ┐—	200 —┐	2000 ┐—
3 —┐	30 ┐—	300 —┐	3000 ┐—
4 —┐	40 ┐—	400 —┐	4000 ┐—
5 —┐	50 ┐—	500 —┐	5000 ┐—
6 —┐	60 ┐—	600 —┐	6000 ┐—
7 —┐	70 ┐—	700 —┐	7000 ┐—
8 —┐	80 ┐—	800 —┐	8000 ┐—
9 —┐	90 ┐—	900 —┐	9000 ┐—

The bottom right corner will be the first order of magnitude (below 10), the upper-right corner the second order (tens), the bottom-left corner the third (hundreds) and the top left the fourth (thousands). This way, when reading a single numeral, one would read from left to right and top to bottom, first the thousands, then the hundreds, then the tens and then the below tens, like for example the number 2023 (┐┐┐┐). An empty line is equal to zero (—), and having one of the corners empty but others with a number attached means that that order of magnitude is zero (such as the third order of magnitude in 2023). Decimals can be made by using a comma and

adding a numeral behind it: in this case the orders of magnitude are flipped: one behind the comma is a tenth, two is a hundredth, three a thousandth and four a ten thousandth. For example, 2023,2023 would be $\text{+}\text{7}\text{ } \boxtimes \text{+}\text{7}$.

An added benefit of this numeral system, besides taking less space, is that addition could be more visually intuitive for some numbers: $\text{—}\text{1}$ plus $\text{—}\text{7}$ equals $\text{—}\text{v}$, $\text{—}\text{1}$ plus $\text{—}\text{7}$ equals $\text{—}\text{v}$, $\text{—}\text{1}$ plus $\text{—}\text{2}$ equals $\text{—}\text{3}$.

Atlan's numeral system also allows for a duodecimal base notation, with the addition of unique numerals for 10 and 11:

$$10 = \text{1—}, 120 = (\text{1—}^{\perp}), 1440 = (\text{1—}^{\text{v}}), 17280 = (\text{1—}^{\text{+}})$$

$$11 = \text{1—1}, 132 = (\text{1—}^{\text{v}}), 1584 = (\text{1—}^{\text{+}}), 19008 = (\text{1—}^{\text{+}})$$

The added benefit of a duodecimal system is that certain fractions divisible by 2, 3, 4 and 6 are more straightforward to calculate by heart, whereas 10 can only be divided easily by 2 and 5. Another benefit is that it is more suited for many numerical systems, such as time divided into 60 seconds per 60 minutes per 24 hours, the twelve months and zodiac signs, as well as rotation being divided into 360 degrees.

The duodecimal system can also be easily counted on a single hand, by using the thumb of the same hand as a pointer to count off the 3 finger segments in the 4 remaining fingers (see figure). Atlans number syllables fit this system as well:



Image taken from West (2015).

1, 2 and 3 are 'IP' \triangle 'OP' \circ ▷ 'UP' ∇ , since these all end in P they are grouped together on the first finger. 4, 5 and 6 are 'IK' \circ — , 'OK' \circ — \circ , 'UK' \circ — , following again the same vowel pattern, but with K, grouping them on the second finger. Similarly, 6, 7, and 8 are 'UK' \circ — 'IM' \circ — \searrow 'OM' \searrow \circ , and 9, 10, 11 are 'JI' \circ — \searrow 'JO' \searrow \circ — \searrow 'JU' \searrow \circ — \searrow .

Currently, Atlan does not have a standardised system to clarify beforehand whether decimal or duodecimal numerals are used, other than to spot the usage of the numerals \top and $\top\top$. Frankly, current duodecimal systems in Arabic notation don't have this either, but it could be easily stated verbally beforehand.

3.5 Mathematics – Stijn Janssens

Just as with punctuation, mathematical symbols can be approximated by semantic atoms. For example, plus + could be ‘AN’ 𐀀 (‘and’), minus - could be ‘NE’ 𐀁 (‘negative’), divided by : could be ‘EP’ 𐀂 (‘per’), equals = could be ‘ME’ 𐀃 (‘equal’). This way, speakers will not be required to learn many new mathematical symbols, but rather the glyphs could function as these, as well as carrying their own pronunciation and meaning. More complicated mathematical symbols or notations might need to be formalized and standardized by mathematicians, which might require more than one syllable.

3.6 Typography

Since Atlan’s writing system is comprised of a set of basic lines, a great degree of artistic freedom is possible in creating different fonts and calligraphy styles to write the language. Atlan typography should make sure to remain faithful to the specific orientations of the different lines as to not cause confusion between them. Since any glyph contains a minimal amount of lines, usually

3.7. NAMES, LOANWORDS, AND CARTOUCHES

two, typographic ornamentation can be added to glyphs without causing much confusion. Here we provide four examples of typographic variations on the word ‘Atlan’: a Times New Roman font, a *Comic Sans*-style font, Asian-style calligraphy and Arabic-style calligraphy.



Figure 1: Atlan in different typographic styles.

3.7 Names, loanwords, and cartouches

Words that denote certain names, such as personal names, placenames, names for institutions, as well as loanwords denoting certain culture-specific objects or concepts, like certain meals, for example, are incorporated into Atlan as phonetic approximations of these words in their source language. Because every possible syllable in Atlan already has its own designated semantic value, this could cause confusion, if the syllables are interpreted as Atlan words instead of loanwords. For this reason, names and loanwords always follow the following structure:

- denoting category, e.g.:

person ſ (EJ)

place ſ (LU)

town L (TOS)

food ſ D (MAJ.KOS)

- NA ſ (name)
- phonetic approximation

Using this formula, ‘Paris’ would become: ‘TOS.NA.

PA.LI’ L ſ D ſ , and ‘curry’ would become ‘MAJ.KOS.

NA.KA.LI’ ſ D ſ ſ ſ . If the conversational context makes it explicit enough that the subject matter concerns a loanword, and not a literal Atlan word, the markers denoting category + NA may be omitted for the sake of fluent speech.

In (formal) typography, a cartouche may be employed to encircle the phonetic approximation in order to enhance intelligibility. Cartouches originate from Ancient Egyptian hieroglyphic orthography, where they were used to encircle the names of pharaohs (see Fig. 1) (Chrisholm, 1911).

3.7. NAMES, LOANWORDS, AND CARTOUCHES



Fig.1 Cartouche of Hapshetsut, one of the few female pharaohs in Egyptian history. She was often referred to androgynously because of a lack of feminine royal nomenclature (Graves-Brown, 2010).

The cartouche originates from the hieroglyphic ‘shen’, a stylised loop of a rope, which literally means ‘to encircle’, but had come to symbolise eternal protection. Therefore, it was also believed to have apotropaic powers (warding off evil). The word ‘cartouche’ comes from the French word for ‘paper bullet cartridge’, first applied when Napoleon’s soldiers encountered the frequently occurring glyph and noticed its resemblance to such cartridges (White, 2002). Interestingly, cartouches played an important role in deciphering Egyptian hi-

eroglyphics. The early Egyptologist Thomas Young, following a suggestion from the decipherment scholar Jean-Jacques Barthélemy, compared the cartouches that appeared on the Rosetta Stone to the proper names that appeared in the Greek text (Robinson, 2009). This way,

he was able to decipher the name ‘Ptolomy’ (Young, 1823). If Atlan were ever to become widely used, and inscriptions were made employing cartouches, perhaps the archaeologists of some distant future might employ them to decipher the script in the same manner that Young did.

The 2001 conlang Toki Pona by the Canadian linguist Sonja Lang also employs cartouches in these contexts, for the same reason that Atlan does so (Lang, 2016). Just like in Toki Pona, Atlan’s cartouche shall be a simple oval circumscribing the name or loanword, without the additional straight line as used in Egyptian cartouches. One could object that employing a typographic element that derives from a specific cultural tradition, namely that of ancient Egypt, is in conflict with Atlan’s constraint of cultural neutrality. To this we object with the following justifications: the cartouche serves a practical function in preventing orthographic ambiguity; the Ancient Egyptian culture is currently extinct and thus does not advantage current day Egyptians over any other culture; the historical origin adds an extra layer of symbolism and cultural depth which in no way interferes with practical use; and frankly, we are of the opinion that it looks quite

cool.

As we will see in the next part, the glyphs can be made on the computer using T_EX³. This also is true of cartouches. A cartouche can be made by simply saying name. For example, the T_EX code `\cartouche{Utrecht}` produces Utrecht⁴

3.8 On Dyslexia – Stijn Janssens

About 3 to 7 percent of the population has some form of dyslexia (Peterson & Pennington, May 2012), (Kooij, 2013), however up to 20 percent of the population experiences some degree of its symptoms (National Institutes of Health, 2015). For Atlan's writing system, this means that a minority group faces these problems, therefore not interfering with its universality / majority constraint. However, from the perspective of the world population, this still amounts to a significant group of people that might be disadvantaged by a writing system that is hostile to dyslexia. Atlan's writing system has some aspects that work with and others that work against dyslectics.

³We will talk more of T_EX shortly.

⁴Utrecht = UT-LEK-T

Within language psychology, different orthographies are classified by their orthographic depth. Shallow orthographies are mostly phonetical, encoding all the necessary information for pronunciation in a straightforward and consistent way. Deep / opaque orthographies are on the other end of the spectrum, often deviating from literal, phonetic spellings, or omitting certain phonemes from written language (Besner & Smith, 1992). Examples of languages with shallow orthographies are Hindi, Spanish and Turkish, while examples of languages with deep orthographies are English, French and most extremely Tibetan (whose last spelling reform took place 800 years ago).

Ideally, an IAL has an orthography that is as shallow as possible, making it more accessible for dyslectic people, as well as optimising learnability for non-dyslexics. In this regard, English is quite a poor IAL candidate. Because of great phonetic shifts and many etymological spellings (see chapter 8.2.0 on language variation), English orthography is highly inconsistent, which can be shown by pushing it to some quite absurd extremes. The word ‘church’ could hypothetically be spelled ‘tolot’, by combining the ‘t’ as pronounced in the word ‘picture’

with ‘olo’ as pronounced in ‘colonel’. The word ‘fish’ could be spelled ‘ghoti’, combining the ‘gh’ from ‘enough’, the ‘o’ from ‘women’ and the ‘ti’ from ‘nation’. The word ‘what’ could be spelled ‘oed’ by combining ‘o’ as in ‘one’ with ‘ed’ from ‘hacked’, and ‘why’ could be spelled ‘ho’, using the ‘ho’ from ‘choir’. Different spelling proposals have been suggested during the past centuries, however no major reforms have been established because of low public acceptance (Wolman, 2009).

Because Atlan is spelled phonetically, it might be seen as having a shallow orthography. However, because of its syllabic nature, vowels are not directly notated but rather implied by the place of conjunction of consonant lines. Therefore technically speaking, vowels are notated, however they might require more focus and attention to be interpreted. More concerningly, many of Atlan’s glyphs have symmetrical counterparts, or are composed very similarly, only differing in orientation or place of conjunction. Since the Roman letters p, q, p and b are often confused, we might expect that such symmetries and similarities might also pose a problem for Atlan. A counterargument could be that the specific locations of line conjunction in Atlan is very regular and straightforward,

only requiring readers to identify these relative positions in order to differentiate them from one another. Since Atlan does not have any fluent speakers at the time of writing, its actual effect on dyslexia can only be guessed. Therefore, we shall briefly discuss three example orthographies that share some similarities with Atlan, with respect to dyslexia.

Most research has been done on alphabetic orthographies, most often English, this tendency is also known as *alphabetism*. Therefore, it is hard to come with concrete data on the effects of other types of writing systems on dyslexia. Hangul, a *featural* writing system used for writing Korean, has many letters that look alike, or are each other's mirror image: □, □, □, □, □, □, □, □, □, □, □, □, □. Hangul's orthography functions by combining letters into syllable blocks, therefore resulting in blocks that are sometimes barely distinguishable from one another at first glance, like □ & □, and □ & □. Surprisingly enough, many Koreans are unaware of the existence of dyslexia. A possible reason for this might be that Koreans read by identifying each syllable block all at once, rather than reading all the different letters within each block. Therefore, native speaking Koreans who learn to

read as children might not experience any symptoms of dyslexia, because they learn to recognise many different syllable block combinations as single images, while always having the phonetic cues present within the block to verify if they read it right (jreidy17, 2014). This would foster a strong association between visual cues, pronunciation and meaning. Atlan might have the same benefit in children learning the language at an early age, where they might learn to recognise each glyph as an individual image, directly associating its pronunciation with it since this information is encoded in the glyph. Since each Atlan syllable glyph has a unique meaning assigned to it, this might even result in quasi-logographic reading, where people might recognise the word, its meaning and pronunciation from its visual shape. People who learn Atlan at a later age might not have this benefit and struggle more with learning the language, as they might keep trying to read each glyph letter for letter, making possible confusion more likely.

The Inuktitut language of the Inuit people employs a writing system which, like Atlan's writing system, might be classified as a featural syllabary, or an *abugida*. It was developed by Christian missionaries in the 19th century,

and just like Atlan, it consists of basic geometric shapes and lines, with different orientations or symmetries indicating different vowels, e.g.: ◻ = pi, ◻ = pu, ◻ = pa, ◻ = pai. Research into dyslexia among Inuktitut speakers is scarce, but the small existing body of research suggests that such writing systems are not any more difficult to learn for children than other writing systems, and that they might even be easier to learn as a first writing system because of the clear parsing of syllables as linguistic units (Donovan & Tulloch, 2022).

The Amharic language of Ethiopia uses an abugida named ‘Fidel’. This writing system requires speakers to memorise many different basic shapes, some of which are very similar to one another, which are then systematically modified with small grapheme variations to indicate different vowels. E.g.: ◻ = ts’ä, ◻ = ts’a, ◻ = ts’o, ◻ = dä, ◻ = da, ◻ = do, ◻ = d. Again, research is limited, but indicates that because of the nature of the writing system, glyph-naming is crucial for all other indicators of literacy (Mekonnen, 2023). When this type of writing system maximises legibility, this creates a positive feedback loop, but when it limits reading, it does the opposite, making overall legibility harder.

3.9 T_EX and Atlan – Jarno Smets

Atlan is precise when needed, but not forcibly rigid. Being all nitty-gritty is possible, but not demanded. Hence, the typesetting language T_EX⁵ is a great fit for this book, and for Atlan’s writing system. Here I will quickly guide the reader through the uses of T_EX in this book.

First, there are our glyphs. They are hand-programmed in T_EX using Tikz. It was a strenuous effort, but worthwhile. The glyphs are high-resolution, scaleable, and they also look the part. The usage is straightforward as well.

To print any of the Atlan glyphs in T_EX you load in the package `Atlan.sty`. Most glyphs are simply transliterations of syllables, with a backslash in front. E.g. `\mum` prints mum . Some commands were already occupied, hence some of the commands are named differently, e.g. `\Atlanpi`, since simply `\pi` would print π . Next up, I plan to make a font that is available on other typesetting platforms.

Then, our numeral glyphs rely on LuaL^AT_EX a more potent version of T_EX. The command, again, is straightforward. You simply state

⁵Or its modern forms L^AT_EX and LuaL^AT_EX, as used here.

`numbr{<number>6}`. An example of the command: `\numbr{321}` produces \top .

Then, of course, this book is typeset in \TeX . We could have made it easier for ourselves. But, typesetting with \TeX was worth the effort. We are proud of what we have made; both content- and appearance-wise.

⁶Due to the nature of our numeric system, the biggest number you can fill in, is 9999.

Chapter 4

Morphosyntax

4.1 Unambiguous Syntax – Jarno Smets

AMBIGUITY is of all times and places, and natural language is rife with it. *Goal, purple people eater, John trades with Mary*; these words and expressions can all be interpreted in multiple ways. Some despise ambiguity, while others wallow in it. Whatever one thinks of ambiguity, it is a part of natural languages.

For our constructed language, we want to minimize ambiguity. This for the sake of clarity and communicability. Hence this essay.

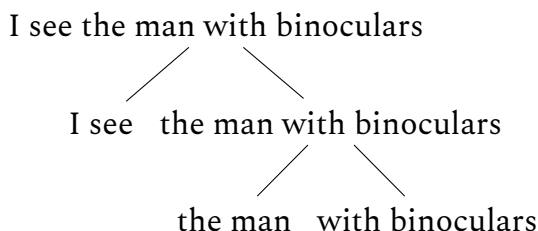
In this essay, I will cover a specific type of ambiguity,

namely *syntactic ambiguity*, also known as *structural ambiguity*. A sentence that can be interpreted in multiple ways due to its syntax, is structurally ambiguous.

My aim in this essay is twofold. First, I want to show why syntactic ambiguity is a problem, especially for the goals of our project. Then, I will propose a strategy to minimize this form of ambiguity, and argue for that strategy choice.

What is syntactic ambiguity?

Syntactic ambiguity occurs when word-order gives rise to multiple interpretations (Oaks,2012, p.16) . The sentence “I see the man with binoculars” could be parsed (split into grammatical parts) in two ways:



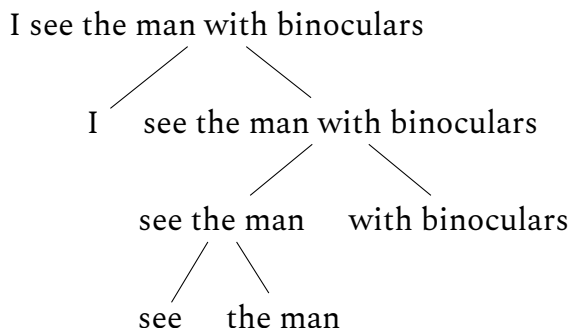


Figure 1: different syntax trees for “I see the man with binoculars”

As we see in the above syntax trees, the difference in interpretation hinges on the (de)coupling of the words *man* and *with binoculars*. You could make *with binoculars* modify *man*. One could also modify *see* via *with binoculars*. The structure of the sentence doesn’t give preference to one over the other.

For further illustration: one common type of syntactic ambiguity, is *scope ambiguity*. Scope ambiguity co-occurs mostly with logical operators such as quantifiers (for all, there exists), negation, and coordinators (and, or, but)¹. *Scope* is the part of a sentence over which such quantifier, negation, or coordinator ranges. Other in-

¹These are all operators in propositional logic. Quantifiers: \forall = for all, \exists there exists, \neg = negation, “not”. \wedge = and/but, \vee = or.

stances of scope ambiguity are seen with modifiers, which I will briefly discuss below. Scope and scope ambiguity can best be explained by example:

(1) *My cat is not grey or black*

Two readings for (1): my cat is neither grey nor black, he is red, for example. Alternatively, my cat is not grey, but is black of colour. The scope for negation is ambiguous here. The *not* either or it has scope over *grey or black*, or it only has scope over *grey*, .

Where lies the origin of such structurally ambiguous sentences? Yang (2014) discerns five major causes of structural ambiguity in English:

- A Negation scope
- B Words with special syntactic functions
- C Improper abbreviation
- D Unclear word-characteristics
- E Unclear modifier-relations

We discussed an instance of A above already. With B, Yang refers to words that generate *subordinate clauses*;

4.1. UNAMBIGUOUS SYNTAX – JARNO SMETS

subsentences. These sentences could either be the object of the bigger sentence, or be a truly subordinate clause. E.g.: *The girls reported to me when they came.* Did the girls report to me after they arrived? Or did they report their time of arrival? it is unclear, due to the meaning of the word *when* .

Now on to cause C Yang mentioned. *Improper abbreviation* is the improper shortening of a sentence. Again, think of the sentence *Mary trades cards with Joe.* I could have said *Mary trades cards together with Joe* if I wanted to convey that message. But I didn't; I left out the word *together*, making it ambiguous.

Then, an example will elucidate cause D: *drinking water is unsafe.* Is *drinking* a verb in itself, or part of the larger phrase *drinking water*? The word characteristics for *drinking* are unclear. *Drinking* can either be seen as a verb, or as a noun together with *water*.

Finally, cause E refers to a modifier. A *modifier* is a linguistic element that changes the meaning of another linguistic element. For example, *grey* modifies *dog*. With unclear modifier relations, it is not apparent which modifier modifies what. In the phrase *purple people eater*, it is unclear whether *purple* modifies *people*, or *eater*.

I propose we bring these causes down to two. Firstly, structural ambiguity is caused by unclear semantic roles. A *semantic role* of a word or sentence-part is the role it plays in the meaning of the sentence. For example, the semantic role of *the grey dog* is the same in both sentences underneath:

(2) *The cat attacked the grey dog*
The grey dog was attacked

In the example given above, *I see the man with binoculars*, the semantic role of *with binoculars* is indeterminate. Is *with binoculars* how I see the man? Or does the man have binoculars? it is precisely this indeterminacy that seems to generate the ambiguity.

The second cause I propose, is unclear word-grouping and unclear scope. To get rid of the ambiguity in phrases as *purple people eater*, or *lesbian vampire killer*, it needs to be specified which words modify which.

The issue for Atlan

In the previous part, I examined syntactic ambiguity. Now, why is this a problem for Atlan?. I will here propose three reasons for that goal. First, I will argue that structural ambiguity inhibits the parsing of language by computers. Computer-parsing could boost the spread of At-

lan. Secondly, I will show that some forms of syntactic ambiguity would endanger the communicative function of our constructed language. Atlan should be a bridge between two languages. Syntactic ambiguity can make it more difficult for two speakers from different languages to communicate. Lastly, I will argue that, in some high-stakes circumstances, syntactic ambiguity could be a great danger.

First of all, syntactic ambiguity is a problem for computers. Computers need a so-called *parser* to understand our language: The machines pick apart a sentence, in order to fully understand it (Schubert,2020). Syntactic ambiguity is a true roadblock for such parsing. Because syntactic ambiguity gives rise to multiple parsing options, a computer can't give a definite parsing of a syntactically ambiguous sentence. To circumvent, or to (partially) overcome it, multiple algorithms have been created. Yet it remains a difficult problem (Chowdhary, 2020, p.645).

For our constructed language, computer parsing and processing could be of help to the language learner. Translations would be more accurate, and practice materials can be generated more quickly. The presence of syntac-

tic ambiguity is troubling for computers to analyze natural language.

Besides, structural ambiguity endangers universality. Since our language is intended as an auxiliary constructed language, people learn our constructed language as a *second language*. Hence, learners all approach our language from the perspective of their mother tongues. Now here lies the problem: different languages have interpret scope in different ways.

This has been shown, for example, in Scontras et al. (2017). This team of researchers found out that Mandarin lacks *inverse scope*. Inverse scope can best be explained by an example: “A badger dug every hole”. In English, two readings are available for such sentence:

Surface scope - *There was one badger such that it dug every hole.*

Inverse scope - *For every hole, there was a (different) badger that dug it.*

Scontras et al. found out that the inverse scope reading is simply not available in Mandarin Chinese. Furthermore, they found out this lack of inverse scope is found in the English of native Mandarin speakers. Another

study showed similar results: Korean learners of English habitually preferred the surface-scope reading, and left the inverse-scope reading out (Seon & Shin, 2022).

So, when learning new languages, speakers have the tendency to bring their native scope-reading preferences with them. This endangers the communicative function of our conlang. If our constructed language has certain scope ambiguities in it, miscommunication can occur. Say you have speaker X, in whose language both scope-readings are available. She communicates such a scope-ambiguous sentence to speaker Y. X wants to bring across the inverse scope-reading. To speaker Y, *inverse* scope-readings are *not* available. Then X fails to bring across *her* wished interpretation of the sentence; a communicative error has occurred. Hence, structural ambiguity endangers the communicative clarity of our constructed language.

Expanding further on communicative clarity: some contexts strictly demand that there be no ambiguity. Hazardous environments, such as nuclear power plants, weapon factories and the like, should communicate in a clear, unambiguous manner. Also law practice should be ridden of ambiguity. These are high-stake-environments. Any

communication mistake could have far-stretching consequences.

Say an English nuclear-power plant has the following instructions etched into an important control panel:

(3) In case of emergency: pull the horizontal striped lever

Now, there are two levers in the control room. One is a lever you pull from north to south, and it is marked with horizontal stripes. The other lever is horizontal, but has vertical stripes instead. Which lever do you pull? I hope this example makes it clear how dangerous syntactic ambiguity can be.

Of course, this was a fabricated example. A real-life example, can be found in (Layman,1962):

(4) Serbian subjects in the United States, shall enjoy the rights which the ... laws grant ... to the subjects of the most favoured nation.

Example (4) elicits two interpretations: Serbian subjects who reside already in the United States enjoy the rights, or Serbian subjects, independent of where they remain, enjoy the rights when they are in the United States. This is syntactic ambiguity in law. Here it can have grave consequences for a large number of citizens.

With these few examples, I have shown why syntactic ambiguity is best left out in our constructed language. Firstly, it would make it hard for computers to parse our language. That while computers generally help to spread a language faster. Secondly, structural ambiguity in a language can cause miscommunication within a language. Not every language allows multiple scope readings, for example. Scope ambiguity can then lead to miscommunication in a language. Thirdly, syntactic ambiguity can be of real danger. It could cause communicative issues in high-stakes environments, such as infrastructure and law.

It must be noted, however, that syntactic ambiguity is not only a *bad* phenomenon. It can also serve poetic and humorist endeavours. For example, the structurally ambiguous sentence

(5) *Time flies like an arrow; fruit flies like a banana*

is undeniably witty². Does the fruit fly similar to a banana, or do fruit flies love a banana? The first part of (5) seems to prime the reader for the first reading.

Minimizing syntactic ambiguity

²Found in (Cornish-Bowden, 2015).

Now I will look at the efforts of other constructed language to minimize syntactic ambiguities. I will examine the benefits and downfalls of their approaches. From that examination, I will aim to distill the strategy for *our* constructed language to bring structural ambiguity to a minimum.

One of the main origins of structural ambiguity is the distance between sentence-parts. In a structurally ambiguous sentence, it becomes unclear how the words are fit into phrases, and then how phrases fit in a sentence. For example, in the noun-phrase *purple people eater*, does *purple* belong to *people*, or to *eater*? Solving structural ambiguity is then making clear which words modify what, to only give one interpretation of a phrase or sentence.

The constructed language Lojban³ indeed does this. It has two ways of specifying which words belong together. The first manner comes in the form of the structure word *bo*. *Bo* enforces scope (The Lojban Reference Grammar, 2023). To see how, let's take the English sentence "That is a big bug catcher". In English, you could

³Lojban [loʒban] is a constructed language, created by a group of people wanting to improve another constructed language, *Loglan*. One of its spear points is having an ambiguous syntactic structure. Found on: <https://mw.lojban.org/papri/Lojban>, may 23rd, 2023.

interpret this either as a big catcher of bugs, or a catcher of big bugs. In Lojban, the word *bo* makes this difference explicit:

(6) That is a bug-catcher that is big.

*Ta barda miptera bo kavbu*⁴.

(7) That is a catcher of big bugs.

Ta barda bo miptera kavbu.

As you might have guessed from the above examples, the structure word *bo* “pulls” two words together, to combine them. Since the combination of words is made explicit by *bo*, ambiguity is resolved.

There is a second way of coupling words in Lojban. The makers of Lojban decided to make rules for grouping, the so-called *brivla*. *Brivla* is an umbrella term for nouns, verbs, adjectives and adverbs (The Lojban Reference Grammar, 2023). The *left-grouping-rule* states that the two leftmost *brivla* are grouped together. So, the sentence *Ta barda miptera kavbu*, is automatically parsed equivalent to the second reading above (The Lojban Reference Grammar, 2023).

⁴*Ta* = “That is”, *barda* = “big”, *miptera* = “bug”, *kavbu* = “catcher”, and *bo* is the structure word. English translation found in (Jbovlaste: a lojban dictionary, 2023)

It seems Lojban got structural ambiguity under control with these two restrains. What are the advantages and disadvantages of this approach?

As already mentioned above, the word-groupings are made explicit, effectively removing structural ambiguity from the language. This increases the clarity of Lojban, and thereby makes the language more universal. There are some downsides however. As we saw above, some scope readings are not even available in the mother-tongue of some speakers. The left-grouping rule described above could enforce a reading upon the language learner, which the language learner is far from familiar with. Lojban then might sometimes give rise to miscommunications.

Another constructed language with the intent of minimizing (syntactic) ambiguity, is *Ithkuil*. Ithkuil marks semantic roles explicitly in noun cases (Ithkuil, Case Morphology, 2023). This is relatively similar to German, where the case *der* usually marks the (male) subject of the sentence, or *des* marks the possessor. Ithkuil has more cases, including the ones we all know (subject, object, possessor, dative). Examples are *instrument*, *force*, *agent*, and much more⁵.

⁵Readers interested in more should visit Ithkuil's website:

Ithkuil specifies the exact case of every noun. Due to that, it is clear which word plays what role in a sentence. In *purple people eater*, for example, *eater* could be nominative, while *purple people* would be marked as accusative. In that way, ambiguity is brought down to a minimum. However, there is one big downside to this approach: it is too complex. Ithkuil is very complex, and hard to learn. Even the creator, John Quijada, can't speak it fluently (Foer, 2023). Thus, the ubiquitous presence of cases seems to do more harm than good; it eliminates ambiguity, but at the cost of learning-ease and fluency.

We have seen how Ithkuil and Lojban deal with syntactic ambiguity. Taking this in account, how will Atlan deal with it?

A feature of Lojban was the explicit word-coupling with the structure word *bo* . The word directly made clear what words formed a separate noun-phrase. However, it is an extra word to remember. We believe it is a better idea to couple words in the most direct sense of the word: literally connect them to each other. This is a familiar feature of, for example, Dutch: *grijze hondentemmer* (grey hound-tamer) versus *grijze-honden tem-*

http://www.ithkuil.net/newithkuil_04_case.htm.

mer. Both in English and Dutch, the words “dog ”and “tamer ”are joined to indicate that they belong together. In speech, words that should be separated, are separated by a pause.

Now, what about scope ambiguity? For negation, for example, we will include two types: sentential and predicate negation. Sentential negation is a form of negation that spans over a whole sentence. For this we put NE in front of the sentence. E.g. *I have **not** been to school today*. Predicate negation on the other hand, only spans over a predicate. For this we put NE in front of the predicate (or noun). For example, *I’m very **un**happy at the moment*. This would fix negation scope ambiguity. Take the aforementioned example *my cat is not grey or black*. The two readings can be separated using the distinction between types of negation:

(8) *My cat is ungrey or black*
it is not the case that my cat is grey or black

The sentential negation will take the form of a distinct particle, whereas the predicate negation will be an affix. This has the following reasons. Sentential negation spans over a whole sentence. To make it immediately

apparent that a sentence is negated, it would be convenient to have a loose particle to place at the beginning of a sentence. Predicate-negation occurs within a sentence, and binds to predicates. Hence, it will be an prefix, connected to the predicate it negates.

This approach to negation doesn't make it more difficult to learn. Most languages are familiar with it: the most common types of negation are negative particles, and affixes (Martin et al., 2005, p. 454) Even if, for a learner's mother-tongue, there is a mismatch between negation type (sentential and negation) and form (particle and affix), the forms are very likely familiar. This will very likely make our approach to negation somewhat more intuitive for a language learner. Moreover, predicate negation is present in a majority of languages (Martin et al., 2005, p.467).

But what about scope ambiguity outside of negation? E.g. *The dog or the cat and the bird made a mess.* Here, we appeal to operator strength from Classical Logic. Negation comes first. Then comes conjunction ("and"). Last comes disjunction ("or") (O'Donnell et al., 2007, p.120)⁶.

⁶After that comes the conditional ("if...then", \rightarrow) and the biconditional ("if and only if", \leftrightarrow). As far as I can tell, they don't seem to generate syntactic ambiguity, hence I leave them unmentioned

In the above example, the sentence is read as: (the dog or the cat) and (the bird) made a mess. That the bird made a mess, is certain. Whether the dog or the cat made a mess is uncertain.

Now it is worth noting a few *caveats* about my approach. Firstly, I reasoned mostly from syntactic ambiguities in English and Dutch. This could leave room in my solutions for syntactic ambiguities not thought of by me. Hence, I talked primarily of *minimizing* syntactic ambiguity. Besides, it is worth noting that context will disambiguate as well. I have mostly examined structurally ambiguous phrases and sentences in isolation. Some of those phrases or sentences would not be as ambiguous in context.

In this essay, I have shown two things. First, I argued that syntactic ambiguity should be avoided when constructing a language. This because syntactic ambiguity troubles computers, endangers communicative function, and can be potentially harmful.

Secondly, I have proposed several general recommendations for battling syntactic ambiguity. This I distilled from previous attempts at constructing structurally un-

here.

ambiguous languages, such as Lojban and Ithkuil . Lojban made its structure clear, but had a redundant syntax rule. Ithkuil explicitly specified the semantic role of each word, but became extremely hard to learn and speak as a consequence.

Atlan won't be as specific as Ithkuil or Lojban. It is a balance we need to find between preciseness and learnability. Both Ithkuil and Lojban are extremely precise, but sacrifice learnability. I am confident that Atlan will find a good balance, and that the learner will profit from that.

4.2 Atlan's grammar

Atlan's grammar has the challenge of steering a middle course between minimalism of complexity, yet simultaneously allowing for unambiguity. It tries to be minimally prescriptive in its structure, allowing for more freedom for individual and cultural expression while remaining intelligible. Atlan will do this in the following way: any grammatical function that can be expressed within the language, has its own unique assigned syllable. Verbs are not conjugated in complicated arbitrary tables, and

nouns are not endlessly modified by cases, but rather specific grammatical functions are conjoined together, like legoblocks, in an entirely regular way. This allows for a lot of freedom in choosing specific grammatical forms without having to know foliage of grammar. The grammatical markers are added in the order in which they are listed in the word list provided in chapter 6.2.

Atlan's word order is both SVO (subject – verb – object, I eat fruit) and SOV (subject – object – verb, I fruit eat). This means that in every case, the first word of a sentence (apart from mood markers such as interrogative or exclamative) is the subject of the sentence. From there, the speaker is free to choose either the verb or the object to follow, depending on, for example, highlighting words, concept constructions, stream of consciousness &c. According to Kemmerer (2012), the total amount of SOV and SVO dominant languages, or in other words, languages that always put the subject first, amounts to 89% of all languages on earth. However, most languages still allow a basic degree of freedom in word order, the dominant word orders are merely tendencies, never hard rules. Therefore, having the flexibility of SVO and SOV ensures that most people on earth will be intu-

itively capable of formulating sentences in Atlan.

Different cases can be marked by adding their corresponding syllables as prefixed to the designated word. The object is marked with the accusative marker 'EK' $\text{—}\bigcirc\text{—}$, verbs with the verb marker 'TU' \bigcirc , possessives with the genitive 'TA' ᵇ &c. Plural is always marked at the very end of the word, as the only exception.

Verbs can be given tense, aspect and mood. Unmarked verbs are always present tense or infinitive, depending on whether it has a subject. A word can be made past by adding the prefix 'PA' $\text{ᵈ}\bigcirc$, future by adding 'FE' ᶑ , progressive by adding 'PO', passive by adding 'PI' &c. For the complete list of grammatical markers, see the list in chapter 6.2.

Predicates, in which something is said of something else, e.g. fruit is sweet, are marked with the predicate marker, where the noun (fruit) comes before, and the predicate (sweet) would come after. This would make 'FUT SI TIT' ᵀ ᶑ ᵇ . An adjective can also predicate something of a noun, meaning that the very same construction, without spaces, can create 'fruit which is sweet', which can be reformulated as 'sweet fruit'. Since sweet

describes the fruit, it is placed behind the word for fruit, since this is the basic rule of thumb for word hierarchy in Atlán.

Because usually, Atlán words are interpreted literally, metaphoric speech may be indicated by the prefix 'MU' μ .

Gender is not marked obligatory, so purely gender neutral language is entirely possible and even more straightforward, however, if the speaker so desires, the particles for 'masculine', 'MA' \searrow or 'feminine', 'FI' \circ can be added.

Atlán has three separate markers for so called 'degrees of removedness from speaker'. This means that the first degree refers to the here and now of the person uttering the language: the first person 'I' 'EJ.AM', the place 'here' 'LU.AM' $\circ \searrow$, the demonstrative 'this' ES.AM $\circ \searrow$, the time 'now' 'JA.AM' $\hookrightarrow \searrow$. The second degree is the second person, once removed from the speaker: 'you' 'EJ.UN' $\searrow \nearrow$, 'there' 'LU.UN' $\circ \nearrow$, 'that' 'ES.UN' $\circ \searrow \nearrow$, 'then' JA.UN' $\hookrightarrow \nearrow$, and the third degree is 'them' 'EJ.AJ' $\searrow \mathscr{D}$, 'yonder' 'LU.AJ' $\circ \mathscr{D}$. The demonstrative 'ES' $\circ \searrow$ with-

out marker for removedness can be understood to be equivalent to 'it'.

Finally, Atlan uses a scale degree of 'negative' - 'neutral' - 'positive'. These markers can be added as prefixes to words to create relative terms, such as cold - room / body temperature - warm. The possibilities with nuanced expression are endless as you can combine many different words and functions together, allowing for the expression of thought that might go beyond the lexical inventory of natural languages.

4.3 Greenberg's universals

The American linguist Joseph Greenberg (1963) compiled a proposed set of cross-linguistic grammatical principles. Atlan, being an IAL, should ideally comply with as much of these universals as possible, such that its grammar is as intuitive as possible to as much people as possible. Below is the full list of Greenberg's universals. If indicated with a plus '+', this means that Atlan follows this principle. If it is indicated with '~' this means that it does not apply to Atlan's structure, but therefore also doesn't break any universal. If it is indicated with a mi-

nus ' - ', however, this means that Atlan does not follow this principle, while it would have to apply. Only 4 out of the total 45 universals are not obeyed by Atlan, and 18 do not apply. This means that Atlan complies with Greenberg's universals to a satisfying degree, and in the cases in which it doesn't comply, this is for the sake of consistency and simplicity of its rules.

Typology

1. + "In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object."
2. + "In languages with prepositions, the genitive almost always follows the governing noun, while in languages with postpositions it almost always precedes."
3. ~ "Languages with dominant VSO order are always prepositional."
4. - "With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional."

5. + “If a language has dominant SOV order and the genitive follows the governing noun, then the adjective likewise follows the noun.”
6. ~ “All languages with dominant VSO order have SVO as an alternative or as the only alternative basic order.”

Syntax

7. ~ “If in a language with dominant SOV order, there is no alternative basic order, or only OSV as the alternative, then all adverbial modifiers of the verb likewise precede the verb. (This is the ‘rigid’ subtype of III.)”
8. + “When a yes-no question is differentiated from the corresponding assertion by an intonational pattern, the distinctive intonational features of each of these patterns are reckoned from the end of the sentence rather than from the beginning.”
9. + “With well more than chance frequency, when question particles or affixes are specified in position by reference to the sentence as a whole, if ini-

tial, such elements are found in prepositional languages, and, if final, in postpositional.”

10. + “Question particles or affixes, when specified in position by reference to a particular word in the
11. + “Particles do not occur in languages with dominant order VSO.”
12. ~ “Inversion of statement order so that verb precedes subject occurs only in languages where the question word or phrase is normally initial. This same inversion occurs in yes-no questions only if it also occurs in interrogative word questions.”
13. ~ “If a language has dominant order VSO in declarative sentences, it always puts interrogative words or phrases first in interrogative word questions; if it has dominant order SOV in declarative sentences, there is never such an invariant rule.”
14. + “If the nominal object always precedes the verb, then verb forms subordinate to the main verb also precede it.”
15. + “In conditional statements, the conditional clause

precedes the conclusion as the normal order in all languages.”

16. + “In expressions of volition and purpose, a subordinate verbal form always follows the main verb as the normal order except in those languages in which the nominal object always precedes the verb.”
17. - “In languages with dominant order VSO, an inflected auxiliary always precedes the main verb. In languages with dominant order SOV, an inflected auxiliary always follows the main verb.”
18. ~ “With overwhelmingly more than chance frequency, languages with dominant order VSO have the adjective after the noun.”
19. ~ “When the descriptive adjective precedes the noun, the demonstrative and the numeral, with overwhelmingly more than chance frequency, do likewise.”
20. - “When the general rule is that the descriptive adjective follows, there may be a minority of adjectives which usually precede, but when the general rule is that descriptive adjectives precede, there are no exceptions.”

21. + “When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in that order. If they follow, the order is either the same or its exact opposite.”
22. - “If some or all adverbs follow the adjective they modify, then the language is one in which the qualifying adjective follows the noun and the verb precedes its nominal object as the dominant order.”
23. + “If in comparisons of superiority the only order, or one of the alternative orders, is standard-marker-adjective, then the language is postpositional. With overwhelmingly more than chance frequency if the only order is adjective-marker-standard, the language is prepositional.”
24. ~ “If in apposition the proper noun usually precedes the common noun, then the language is one in which the governing noun precedes its dependent genitive. With much better than chance frequency, if the common noun usually precedes the proper noun, the dependent genitive precedes its governing noun.”

- 25. ~ “If the relative expression precedes the noun either as the only construction or as an alternate construction, either the language is postpositional, or the adjective precedes the noun or both.”
- 26. ~ “If the pronominal object follows the verb, so does the nominal object.”

Morphology

- 27. ~ “If a language has discontinuous affixes, it always has either prefixing or suffixing or both.”
- 28. ~ “If a language is exclusively suffixing, it is postpositional; if it is exclusively prefixing, it is prepositional.”
- 29. + “If both the derivation and inflection follow the root, or they both precede the root, the derivation is always between the root and the inflection.”
- 30. + “If a language has inflection, it always has derivation.”
- 31. + “If the verb has categories of person-number or if it has categories of gender, it always has tense-mode categories.”

- 32. ~ “If either the subject or object noun agrees with the verb in gender, then the adjective always agrees with the noun in gender.”
- 33. ~ “Whenever the verb agrees with a nominal subject or nominal object in gender, it also agrees in number.”
- 34. ~ “When number agreement between the noun and verb is suspended and the rule is based on order, the case is always one in which the verb precedes and the verb is in the singular.”
- 35. ~ “No language has a trial number unless it has a dual. No language has a dual unless it has a plural.”
- 36. + “There is no language in which the plural does not have some nonzero allomorphs, whereas there are languages in which the singular is expressed only by zero. The dual and the trial are almost never expressed only by zero.”
- 37. + “If a language has the category of gender, it always has the category of number.”

- 38. + “A language never has more gender categories in nonsingular numbers than in the singular.”
- 39. ~ “Where there is a case system, the only case which ever has only zero allomorphs is the one which includes among its meanings that of the subject of the intransitive verb.”
- 40. + “Where morphemes of both number and case are present and both follow or both precede the noun base, the expression of number almost always comes between the noun base and the expression of case.”
- 41. + “When the adjective follows the noun, the adjective expresses all the inflectional categories of the noun. In such cases the noun may lack overt expression of one or all of these categories.”
- 42. + “If in a language the verb follows both the nominal subject and nominal object as the dominant order, the language almost always has a case system.”
- 43. + “All languages have pronominal categories involving at least three persons and two numbers.”

CHAPTER 4. MORPHOSYNTAX

- 44. ~ “If a language has gender categories in the noun, it has gender categories in the pronoun.”
- 45. + “If a language has gender distinctions in the first person, it always has gender distinctions in the second or third person, or in both.”
- 46. + “If there are any gender distinctions in the plural of the pronoun, there are some gender distinctions in the singular also.”

Chapter 5

Our Ontology

The goal of this chapter is to expound on the ontology of the language, which concerns its semantics and syntactical inventory. I will achieve this by discussing the literature on various attempts and classifications proposed within the topics, and then dividing all linguistic meaning into its irreducible, unambiguous and unique atomic components, as to respect constraint 2 (unambiguity). It will use some principles of natural bifurcation of meaning (chapter 5.1.2), and built off of the universal substratum of phenomenology and qualia (chapter 5.2 and 5.3). It will also be linguistically grounded in the culturally universal basic elements of human life (chapter 5.4).

5.1 Parsimony in semantics

Oligosynthesis

Besides constraint 1 (cultural neutrality), which will be discussed in chapter 5.4, Atlan's lexicon has to follow constraint 2 (unambiguity) and 3 (parsimony), which will be discussed in this section. Ideally, the language should contain as little basic words as possible, as to reduce the time required to learn the language. Therefore, it should be sparse with its words, only adding new words when these carry a meaning that is not already covered by another word. Complex concepts should not get their own separate words, for this would add an inestimably large number of extra words, but rather be composed of more simple and universal words that constitute its meaning. Atlan shall achieve this by using a semantic system that is oligosynthetic, meaning that it has a very limited number of semantic atoms (*oligo* = few), from which more complex meaning is built by combining different atoms (*synthesis* = combining). Each semantic atom (or 'root') shall be covered by a unique one-syllable word. Atlan syllables can take four shapes (C = some consonant, V = some vowel): V, CV, VC, CVC. This is abbreviated to

(C)V(C). There are 9 consonants and 5 vowels in its phonetic inventory, yielding a total of $5 + 5 \times 9 + 5 \times 9 + 5 \times 9 \times 9 = 500$ possible combinations, however all syllables ending in *-ij* cannot be sufficiently distinguished by ear from those ending in *-i*, so all (-) *ij* will not be included. This leaves a total number of 490, and thus the challenge posed in this chapter is that of reducing all meaning to 490 atoms or a combination of these.

The term *oligosynthetic* was first coined by the linguist Benjamin Lee Whorf and is defined as having at most a few hundred-word roots. However, this seems to be extremely rare among natural languages. A possible example might be the Kalam language of the Highlands of New Guinea (Pawley A. , 1993). Two other languages previously regarded as oligosynthetic by Whorf are the Aztec language Nahuatl and the Native American language Blackfoot, but these are now commonly classified as polysynthetic (using many roots to synthesise more complicated meaning). Oligosynthesis is more popular among constructed languages, such as Sona, Ro, aUI, Ygyde and Kali-sise (FrathWiki, Oligosynthetic language, n.d.). These all have different numbers of semantic primes and methods of synthesising them, but they

commonly have the following problems (Watson, n.d.):

- I. Complicated combinatory systems
- II. Unclear word-parsing
- III. Vagueness of composite meaning

Atlan will overcome problem I by using an extremely basic manner of combinatory synthesis: the most semantically essential prime comes first and is followed by primes that hierarchically specify the meaning of the word. Grammatical functions always come in front of the semantic root as prefixes (except for the plural¹), and semantic specifications are appended as suffixes.

Atlan words composition:

grammatical.function – main.semantic.root –
semantic.specification – plural

This will also solve problem II, because syllables can be recognised to be semantic when they are CVC, and grammatical otherwise. This way, grammatical syllables are easily distinguished from semantic ones. It is therefore always audibly clear where a word begins and ends. In written text this is aided by the fact that in Atlan's own

writing system, CV and VC syllables always consist of a small circle attached to a line, while CVC syllables are always two lines connected to each other.

The 5 V syllables will be restricted to mood-markers and general sentence structuring, because of the onomatopoeic quality of these basic vocal sounds. These, however, can also be used grammatically to modify the meaning of nouns, verbs and pronouns, for example to turn 'here' into 'where?'.

- Exclamative (prosody), imperative, vocative = o
- Interrogative (question, prosody) = e
- Stress marker (prosody) = a + stress
- Relative clause = i (+ pronoun)
- Subjunctive (wish) = u

The 89 CV, VC syllables will be mostly restricted to morphological and abstract functions, indicating grammatical, syntactical or logical functions and relationships between words. Logical functions will be derived from the logical operators of predicate logic and possible worlds semantics (Priest, 2011), and grammatical functions are

derived from the Universal Networking Language (UNL) (Portal, sd), and syntax-semantically reduced where possible (see chapter 6.2). UNL was initiated by the United Nations University in 1996 and continued by the international non-profit organisation UNDL from 2001 onwards. Its goal is to function as a formalised pivot language between natural language and interlingual machine translation, therefore having formalised all grammatical functions, together with a large-scale ontology of all concepts contained within all its source languages (Universal Networking Language Portal, n.d.). Atlan does not use the latter, however, because it would break constraint 3, parsimony.

The remaining 396 CVC syllables will cover the semantic primes, and these will be systematically selected and ordered in the remainder of the current chapter and chapter 3 and 4. Because Atlan's writing system is syllabic, and each syllable has a fixed semantic value assigned to it, individual glyphs can be read both phonetically as well as ideographically/logographically. This would allow for it to be a kind of interlingual orthography.

Problem III, that of vagueness of composite meaning, will be tackled in several ways. Most importantly, some

of the most universal ‘semantic molecules’ (see chapter 3) will have their own assigned syllable. These molecules are definitions that could be reduced to more fundamental atoms but are often used to build more complex meanings, therefore being condensed into a molecule as to prevent unnecessary complexity of compound words. However, this is not an all-encompassing solution for words that have very specific and context derived definitions.

Circumlocution is the phenomenon where concepts which do not have a specific word for them in a language are described by giving a circuitous description of the intended meaning. An example of this in English would be ‘the day after tomorrow’. Atlan will never be able to fully eliminate some forms of circumlocution in its lexicon, mainly because of it being an oligosynthetic language. However, confusion around these instances can be minimalised in the following ways.

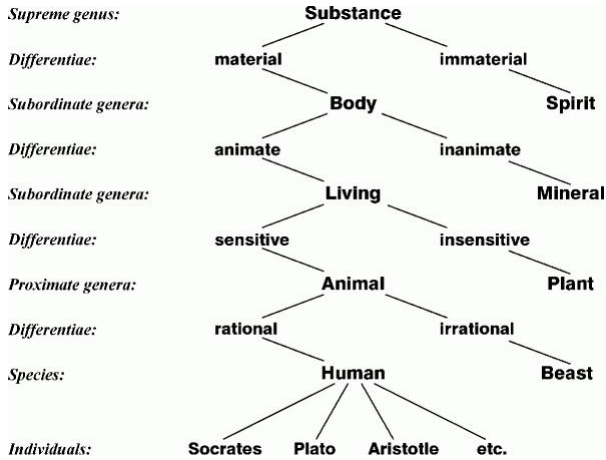
A standardized set of compound definitions can aid speakers by being a guideline to using compound words. Someone who learns the language, would have to learn the 490 semantic primes, and then be familiarized with some standardized compound words. Because the meaning of the compound word will be derived from the syllable

bles it contains, learning these compound words will be intuitive and require less mnemonic effort than learning a completely new word would. This way, when a speaker encounters a word they have never heard before, they will be able to derive, or at least estimate the intended meaning just by recognizing its syllables.

Additionally, neologisms could be created during improvised language use, as a sort of generative etymology, and be directly perceived by the listener, possibly allowing for freer linguistic cultural- and self-expression. Finally, some compound words will be systematically constructed in a taxonomical fashion, when the word's definition allows for this, such as is the case for all living creatures. Each compound word is sorted along the axis of importance, with the most fundamental semantic prime in front, and followed by other primes that hierarchically add layers of semantic precision. This creates a universal logic in the composition of compound words that allows one to identify the ontological category in which the word falls and refine the definition by the refining descriptors appended to it, as if 'zooming in' with a semantic lens.

Taxonomy

5.1. PARSIMONY IN SEMANTICS



Tree of Porphyry, taken from Cohen (2007)

Many philosophical constructed languages that came before Atlan suggested or employed a so-called taxonomical ontology. Many of these are inspired by a diagram invented by the 3rd century Neoplatonist Porphyry when explaining Aristotle's *Categories* (Porphyry, 300), called the Porphyrian tree (Franklin, 1986). In the *Categories*, Aristotle outlines a broad ontology of human apprehension, classifying everything that can be the subject or predicate of a statement into 10 categories: substance, quantity, quality, relation, place, time, position, state, action, affection (Aristoteles, 40). The Porphyrian tree (shown below) shows how a human falls into these cat-

egories, by showing different bifurcations from the first five categories (Cohen, 2007):

Before Aristotle, similar ontological categorisations were made by the Vaiseshika school (Padārtha) (Stanford University, 2019) and the Stoic school, and after him numerous other thinkers, including Plotinus (430, 2019), Kant (Categories) (1781, 1998), Hegel (1812, 1975), Peirce (1867), Husserl (1900, 1993) and Whitehead (categorical scheme) (1929, 2010). Not to mention folk ontologies, such as the Bantu nominal classes (Bleek, 1862–69) and the common distinction between animate and inanimate. All of these mostly were not taxonomical, however, and have been discussed extensively, but it is not the purpose of this paper to further investigate these discourses.

There is, however, one specific notion within these categorisations which Atlan will incorporate, namely that of the *triadic categorisation* of Being. In Hegel, this comprises 1) Being (mind, consciousness, sensation), 2) Essence (Other, duality) and 3) Notion (synthesis, reference). In Peirce, these equate to Firstness (Quality, feeling, consciousness), Secondness (Reaction) and Thirdness (Meaning, representation). Atlan incorporates this through three degrees of removing from the speaking subject (see chap-

ter 6.2). The first degree can be combined with the semantic prime for ‘person’ (‘EJ’ ʃ) to mean ‘I’ (‘EJ.AM’ ʃ ʌ), the second degree to mean ‘you’ (EJ.UN ʃ ʃ), and the third to mean ‘he / she / they’ (‘EJ.AJ’ ʃ ʃ), as well as other primes like place, time and demonstratives.

A more explicit taxonomical ontology makes use of a so-called hierarchical classification. The idea of a ‘perfect’ or philosophical language was a popular idea during the enlightenment, being discussed by Bacon, Descartes, and Newton as a part of a widespread desire for a language that does not confuse the speaker’s understanding of reality or distort the natural order present in it (Eco, 1994, pp. 209-227).²

The core idea of such a language is that it would build its words by adding different letters for every hierarchy of meaning. This way, related words would sound similar. There is, however, always a necessary degree of arbitrariness to such a system, since the arbitrary choice has to be made of what categories are specified by adding a limited set of different letters. These would have to be made from scratch, not based on previous languages, as

to avoid confusion caused by natural language. These languages are known as a priori, and Atlan will be a priori as well, even though its contents are sourced from data from many natural languages but recombined in an attempt to circumvent natural language confusion.

The first serious attempts at this ideal were made by George Dalgarno in his *Ars signorum* (1661, 1968) and John Wilkins in his *Essay Towards a Real Character and a Philosophical Language* (1668, 1968). Initially, the two collaborated on a philosophical oligosynthetic language, but they couldn't agree on whether to make the taxonomy encyclopaedic or build compound words from a small set of primes. Wilkins published his own version based on the former and Dalgarno the latter. Dalgarno's language sadly never caught on, perhaps because the explanation of the linguistic working of the language was shrouded in philosophy which explained the structure ¹.

Wilkin's language found a bit more recognition, being originally taken seriously by the Royal Society, with an attempt to finish the language after his death by a designated committee. It too, however, slowly lost interest in people and descended into oblivion (FrathWiki, *Ars*

¹We hope this is not the case for this book as well

Signorum, n.d.). Its taxonomical classification was structured to encompass every animal, plant, mineral and artifact. He achieved this by setting up an ingenious taxonomical tree to indicate the relations and bifurcations of meaning, along with a system of hierarchically adding vowels and consonants to specify differences and species within different categories.

Wilkins regarded the language presented in his essay as just a draft, although he provides 2,030 different primitives, as well as a 15,000-word list for different English words but admits that it should be worked out by different teams of scientists to work out different concepts within their respective disciplines. His collaboration with the Royal Society was largely part of that attempt. Later on, Wilkin's taxonomy went on to inspire Roget's Thesaurus (1805) and later on Diderot and d'Alembert's Encyclopaedia (1759).

The idea was met with a lot of criticism as well. Voltaire criticised the optimism of the people attempting to create such a language in the form of the character Dr. Pangloss in his satire *Candide* (1759, 1963). Jorge Luis Borges wrote an essay criticising taxonomical categorisations in general and Wilkin's language specifically (1942) in

which he mocks different instances of arbitrary classification by mentioning a fictional Chinese taxonomy called the *Celestial Emporium of Benevolent Knowledge*. This list contains some very culture-specific, arbitrary and absurd categories such as ‘those belonging to the Emperor’, ‘those that have just broken the vase’ and ‘those that from afar look like flies’. This criticism seems like a bit of a stretch, because Wilkins put in a systematic effort to make a coherent classification, and this is not as arbitrary or absurd as Borges’ fictional classification. The linguist George Lakoff supports this claim by stating that many non-western cultures use classifications similar to European ones (Lakoff, 1987). Borges does point out that a successful execution of the idea could in theory have many benefits: “Mauthner points out that children would be able to learn this language without knowing it be artificial; afterwards, at school, they would discover it being an universal code and a secret encyclopaedia” (Blevins, n.d.).

Foucault was inspired by Borges’ essay to write his book *The Order of Things* (1966, 2010, p. preface) in which he analyses the social grounding of epistemic assumptions. He argues that implicit norms within intellectual communities determine thought and influence which top-

5.1. PARSIMONY IN SEMANTICS

ics are researched, and which are not, and how the established bias influences the interpretation of the data that is found. These assumptions and norms are bound to cultural and historic settings, and periodically go through reforms as a result of paradigm shifts. This is a strong blow to the aspiration of a universal classification of the universe. Borges claims that this is because: “we do not know what sort of thing the universe is”. Metaphysicians and phenomenologists might differ on this, however, as will be discussed in chapter 3 of this essay.

Besides this, Wilkins’ system has the disadvantage of only having a limited number of differences and species that can be specified because of the limited phonology. Atlan will be more similar to Dalgarno’s language, in that it will not drive its words through a hierarchical process of taxonomy, but rather by combining primes at will, allowing for exponentially more semantic combinations than are possible in Wilkin’s system.

Another problem of Wilkin’s language is that words with similar meanings have very similar pronunciations, to the point of confusion. Modern information theory warns of this (Norman, n.d.), and Eco even identified that Wilkins himself made such a mistake, confusing *Gade*

(barley) for *Gape* (tulip) (1994, p. 249). This would hinder the language's intelligibility when mishearing can easily change important nuances in definition, as well as making is harder to speak fluently, because any speaker would have to work through tables and flowcharts in their minds while simultaneously talking, without making any mistakes. The language would also be very intolerant of subtle shifts in pronunciation and phrasing that tend to occur naturally within languages over time, because this would cause the whole encyclopaedic house of cards to come crashing down.

Atlan will not have this problem, because its semantic primes are syllables instead of phonemes, and Atlan's phonemic inventory is built to accommodate variation in phonetic approximation and sound shifts within its 14 archetype letters, without causing confusion or ambiguity.

The philosopher Deleuze and psychoanalyst Guatarri proposed an alternative to arborescent (tree-like, hierarchical) epistemic networks like employed by Wilkins and Dalgarno, namely that of a *rhizome*, an analogy with a decentralised plant root network (1980, 2019). Such a model does account for bifurcations and conceptual re-

latedness but is more modal and allows for more complex interlinking than mere hierarchy. It would also fit the philosopher Quine's idea of an interrelated epistemic 'web of beliefs' (Ney, 2014), as well as Wittgenstein's claim that concepts are not clearly delineated, but rather surrounded by a 'corona' of associated concepts (1953, 2010, p. p. 181).

Because of this, the main ontology seems to be better off using a combinatory system, which would allow for endless recombination and web-like relationships between similar words. This however doesn't mean arborescent taxonomy should be completely abandoned. The most relevant modern case of taxonomical classification is that of natural species, although individual species don't have clear demarcations and are loosely defined by their ability to produce fertile offspring (Nature Publishing Group, n.d.). Genetic diversification happened through bifurcation, known as *speciation*. The reverse, different species merging into one through hybridisation, called *despeciation*, does sometimes occur (including among early hominins), but is exceedingly rarer (Junior, 2018). Because of this, the evolutionary tree of life is primarily arborescent.

CHAPTER 5. OUR ONTOLOGY

Modern biological taxonomy employs the following hierarchical classification: life – domain – kingdom – phylum – class – order – family – genus – species (Biology Dictionary, 2017). Atlan’s biological lexicon is constructed along this framework, using semantic primes to describe different bifurcations, inspired by the Latin etymologies employed in binomial nomenclature (a hang-over from Latin being the academic *lingua franca*). These binomial nomenclatures only mention the genus and species names of an organism, and words to designate species can re-occur in other genera to identify other species, adding to the parsimony of terms required to name all creatures within this system. Atlan shall have separate primes for the categories: *virus*, *bacteria*, *archaea*, *plants*, *amoebas*, *fungi*, *animals*. In chapter 4 of this essay, a few culturally universal animal terms are identified, and these will be reduced to: *mammal*, *fish*, *bird*, *worm*, *reptilian*, *insect*. Other animals could be reduced to their descriptions: sponges could be designated as foam-animals, starfishes to star-animals, snakes to legless-reptiles, amphibians to mucus-reptiles, molluscs to shell-animals, jellyfish to mucus-animals etc.

Chemical molecules could be named by formalising a

translation of the IUPAC nomenclature of organic chemistry from the originally Greek roots (IUPAC, 2021). Just like Wilkin's language, Atlan will be dependent on scientists and specialists of different kinds of professions to add formalisations of their respective jargon nomenclatures to the lexicon in order to fully flesh out its lexicon.

The reason why such a systematic description of reality claiming to be universal might be problematic, could be that it claims to have an objective image of reality, while all human thought, individual or collective, will always be fundamentally sourced from subjective experience. The philosopher Thomas Nagel described this by saying that there is no *view from nowhere* (1989), while such a taxonomy appears to claim this anyway. The fact is that human concepts always have a necessary degree of arbitrariness because of the limited resolution of our conceptual boundaries. A taxonomical ontology implies the possibility of grasping some fundamental reductionist principle inherent to reality, while failing to see that many human concepts are emergent phenomena. The linguistic concept of an 'organism' cannot be reduced to a collection of organic molecules, because it is their complicated interplay that generates the multiple properties

that pertain to an emergent system of self-preservation that humans call a single organism (Brigandt & Love, 2017). On a microscopic level however, the clear boundaries of any single individual become fuzzy. It is because human life does not generally take place within a microscopic paradigm, that our concepts don't have this level of detail. Reality is never described objectively, but always relative to the individual(s) observing and describing their reality. Humans appear to be 'at the centre' of their own language and understanding of reality.

5.2 The anthropocentricity of language

Language and ontology are strongly entwined with one another: an ontological system is dependent on the words available to name its parts, and likewise a language is built from the set of concepts, relations, abstractions and 'things' that are captured by its lexicon (Moltmann, 2017) (Boucon, 2019). Though by far not being the only elaboration on this idea, the Sapir-Whorf hypothesis is the most well-known inference that has been drawn from it. In short, this idea postulates that the range and limits of a person's thought are determined by the language

5.2. THE ANTHROPOCENTRICITY OF LANGUAGE

they speak. The strong version of this claim, linguistic relativism contends that all of human thought is fundamentally determined by language (linguistic determinism), resulting in some thoughts being lost and modified through translation or even untranslatable. It treats language as a fixed set of cognitive tools that acts as a constraint on the individual. This view, however, doesn't enjoy a scholarly consensus (Whorf, 1956).

However, this conception of language seems to be monolithic and extra personal instead of dynamic and dependant on individuals and their fluid interactions. It is challenged when presented with the fact that people within the same language can have vastly different ontologies, philosophies and vocabularies, depending on their individual personalities, interests and social environment, and the fact that individual people can learn multiple different languages and express their thoughts through them, nonetheless. Multilingualism is, however, noted to create within an individual, multiple linguistic 'personæ' for the different spoken languages, where one's way of formulating thoughts and uttered sentences are altered by the individual characteristics of the different languages (Pavlenko, 2006).

CHAPTER 5. OUR ONTOLOGY

Perhaps the influence that language has on the thoughts of its speaker can be likened to how putting on different glasses can alter one's perception but does not change the fundamental scene being perceived through them. Sunglasses block out UV light, tinted glasses block out certain colours, different lenses shift the focus to what is near and other to what is far etc.: they all suppress some elements and amplify others, but they never change the basic composition of what is being perceived (if we discount virtual reality glasses).

Using this metaphor, the purpose of Atlan is somewhat similar to being a clear, untainted, undeformed, unbiased pair of linguistic glasses, for as far as this is possible. Every single person's eyes are different, and the glasses of their native language might be more or less similar to Atlan's. The question then becomes: what constitutes this clear human experience that becomes tainted by language?

First, we must realise that language itself is ontologically dependant on the total sum of living speakers. A dead, forgotten or undeciphered language cannot be said to currently exist in the same way that a living language like English exists. It might be revived in the fu-

5.2. THE ANTHROPOCENTRICITY OF LANGUAGE

ture through the reconstruction of its linguistic information, but it only comes back into being when living humans are again able to read, write or speak the language. Furthermore, Wittgenstein's private language argument states that a language is a fundamentally social thing, and that a purely personal language is therefore by definition impossible (1953, 2010, pp. §243-271). Language is a complicated system of communicating all kinds of mental information, like thoughts, feelings, intentions, physical data &c., for all kinds of different purposes, like cooperation, social bonding, problem solving . In individual growing up in solitude or alongside animals never has the need nor possibility to learn and use a language, because there are no other humans around to converse with. After having passed the critical period of language acquisition without ever having learned a human language, an individual will never again be able to do so later in life (Robson, 2002).

Therefore, language is an inherently human thing, that emerged from the transferring of information from one person's individual experience to another's. Phenomenal cues, like the sound of words, the rhythm of speech, facial expressions and gestures are used as an in-

terpersonal bridge between the private mental worlds of the separate individuals. Someone can both hear themselves talking, as well as someone else: language exists in a shared phenomenal space, whereas inner thought is private. Language then becomes a highly codified system of phenomenal metaphors. The sound of a specific word is not the same as the information it codifies, but is consistently associated with the referred phenomenon, in the form of an abstracted ‘concept’.

Atlan should thus have an ontology that is built off the subjective human experience, when regarded in a social context and in direct contact with its physical environment. This immediately brings a degree of anthropocentricity with it, because words relating to the human psyche, body, daily life, social environment etc. will be given higher priority than the myriad of concepts and jargon within specialized disciplines that are less directly related to the everyday human experience. Moreover, humans should be able to think, talk and understand fluidly in a language, and not be required to consciously perform complicated linguistic computations within their heads while using the language.

5.3 Phenomenological ontology and qualia

Subjective experience is ultimately prior to any claim, idea, observation, connection &c that can be communicated about reality. Anything in the world has to first present itself to us humans through phenomenal, subjective experience, before we can abstract it and understand it as an 'objective' phenomenon. However, the mainstream scientific metaphysical framework has, up to now, consistently been materialist, physicalist and reductionist. When it does acknowledge the existence of mind, it often does so in greatly unsatisfactory manner by employing some version of dualism, with the mind being metaphysically separate from the physical, but somehow miraculously still having epistemological and sensory access to it and the ability to manipulate it (moving one's body at will) and be manipulated by it (physical alterations to the body or the ingestion of physical substances can alter the mental perception). When confronted with these problems, science will often try to explain the connection through a functionalist account of the neural network and a materialist explanation of the composition of our neurons, but always failing to close the explanatory

gap to how this material process constitutes phenomenal experience. This is most painfully brought to light by the Hard Problem of Consciousness, the insurmountable chasm between a mechanistic description of neurology and the subjective experience of what it ‘feels’ like to exist as a conscious entity. Heidegger, building off the first phenomenological philosophy of Husserl, already warned of this in his own time halfway the 20th century, he called it *Seinsvergessenheit*, the ‘forgottenness of Being’ (1962, 2019). Somehow the abstractions of reality that were derived from experience have gotten a higher ontological priority than the original experience itself, the ‘objective’ is regarded with a higher esteem than the ‘subjective’. It is beyond the purposes of this essay to explain why this happened and why it is metaphysically self-contradictory. Therefore, building off the premise of linguistic anthropocentricity established in the previous chapter, I shall relate the relevance of the subjective phenomenal experience to the construction of a universal human ontology in the current chapter.

A famous thought experiment regarding the irreducibility of experience is called ‘Mary’s room’ (Jackson, 1982). It imagines a hypothetical scientist named Mary who (dis-

5.3. PHENOMENOLOGICAL ONTOLOGY AND QUALIA

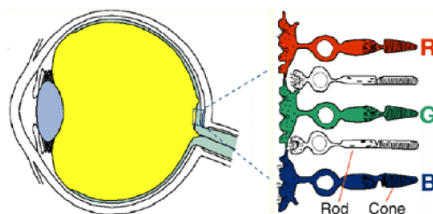
regarding ethical concerns for the sake of the thought experiment) is raised in an exclusively black and white environment for her whole life, and educated about the science of colour perception, without ever seeing colour herself. She would have learned all there is to know about the physics of light, the biology of light receptors in the eye and the neural processing of visual information in the brain. The thought experiment then asks us: if Mary was to then leave her black and white environment and step outside and see colours for the first time in her life, would she learn anything new from experiencing, for example, the colour red for the first time? Could she have known its qualitative experience before she left the black and white room? Philosophers generally agree that she could not have known (the ‘knowledge argument’) (Nida-Rümelin & Conaill, 2019). The same thought experiment could be extended to other subjective sensory perceptions like smell, taste, touch and sound and by extension even emotions and altered states of consciousness.³ Therefore, these ‘subjective’ qualitative aspects of experience appear to be fundamental and irreducible, modern philosophers call them ‘qualia’ (Tye, 2021). Since the coining of the term qualia in 1929 by C.I. Lewis, the con-

cept has remained mostly confined to longwinded debate within the philosophy of mind.

The formalisation of qualia is done by taking an ‘objective’ scale such as the spectrum of light frequencies, and then mapping phenomenal experience onto this by taking as a fundamental unit the smallest perceptible difference. Classically, qualitative experience is divided into the five physical senses: *vision, hearing, smell, taste and touch*. In this essay I will supplement these with affective emotional experience and altered states of consciousness (the subjective experience of being stoned, drunk or tripping seems to be irreducible, they can be vaguely described when compared to sober consciousness, but to know the qualia of the experiment, one must take these substances personally). Because the language is oligosynthetic (see chapter 1 of the essay), the main purpose of this chapter is identifying the basic building blocks of the different types of qualia (vision, sound, taste, scent, physical sensation, emotion and consciousness states), which can then be combined into more nuanced qualitative descriptions.

Vision

5.3. PHENOMENOLOGICAL ONTOLOGY AND QUALIA



Cones of the human eye. From Mafalda (2017)



Wheel of colour. From Judge (2012)

Firstly, colour is the most straightforward, because it is already commonly divided into primary, secondary and tertiary colours. In psychophysical colorimetry the primary colours red, green and blue are regarded as being *complete*, that is, constituting all human perceptible colours when combined along an axis of light to dark (additive light mixing), yet also being *imaginary*, meaning only existing subjectively as qualia, while not being distinguishable as primaries through ‘objective’ measurement (MacEvoy, 2007). A standard human eye contains three types of light receptor cones, one for each of the three colours.

All qualia that exist as polarities on a spectrum will be reduced to a semantic prime for the spectrum, combined with the particles for *positive/high*, *neutral* and *neg-*

CHAPTER 5. OUR ONTOLOGY

ative/low. Translating this into an oligosynthetic semantic colour inventory, we would get the following (capitalised letters representing some as of yet undetermined but fixed assigned syllable):

SHADES:

- Black/dark: JE.LAS ʃ ʎ (= negative + brightness)
- White/light: FO.LAS ɓ ʎ (= positive + brightness)

PRIMARY:

- Red: EL ʎ
- Green: OS ɓ
- Blue: UL ɓ

SECONDARY AND TERTIARY (RGB is chosen as fixed order):

- Orange: EL.OS ʎ ɓ
- Cyan: OS.UL ɓ ɓ

5.3. PHENOMENOLOGICAL ONTOLOGY AND QUALIA

- Magenta: UL.EL ୧ ୨
- Orange: EL.OS.EL ୨ ୩ ୨
- Chartreuse green: EL.OS.OS ୨ ୩ ୩
- Spring green: OS.UL.OS ୩ ୧ ୩
- Azure: OS.UL.UL ୩ ୧ ୧
- Violet: UL.EL.UL ୧ ୨ ୧
- Rose: UL.EL.EL ୧ ୨ ୨

In 1969, anthropologist Brent Berlin and linguist Paul Kay published their book '*Basic Color Terms*' in which they proposed their research concerning the prevalence and development of different colour terms in languages from around the world (Berlin & Kay, 1969). They proposed a chronological scheme of seven evolutionary stadia through which languages generally add colour terms to their lexicon. These are as follows:

- Stage I: dark-cool (>'black') & light-warm (>'white')
- Stage II: red

CHAPTER 5. OUR ONTOLOGY

- Stage III: green OR yellow
- Stage IV: green AND yellow
- Stage V: blue
- Stage VI: brown
- Stage VII: purple, pink, orange or gray

Atlan, needing to conform to constraint 1, cultural neutrality, should thus contain these colours in its lexicon. Stage I-V have already been accounted for, and stage VI and VII can be covered by the combination of the colours from the earlier stages (following the order of shade-XYZ and primary-secondary-tertiary):

- Brown: red + green + blue = 'EL.OS.UL' ʼ ɔ ɔ ɔ
- Pink: white + red ʼ = 'FO.LAS.EL' ɔ ɔ ʼ
- Gray: colour + brightness + neutral = 'KAL.UJ.LAS' ʼ ɔ ɔ

All other colours and shades can be achieved using this combinatory system. One might argue that not all

5.3. PHENOMENOLOGICAL ONTOLOGY AND QUALIA

languages have the same lexical colour inventory, and some make more or less distinctions than English, but it should be noted that having a word for a specific colour is not the same as being able to perceive these different colours and their (subtle) differences (weak linguistic determinism, see chapter 1). Within the line of thought of linguistic determinism, one could argue that learning to speak Atlan, as having this colour system, would gift the speaker with an intuitive understanding of the composition of phenomenal color.

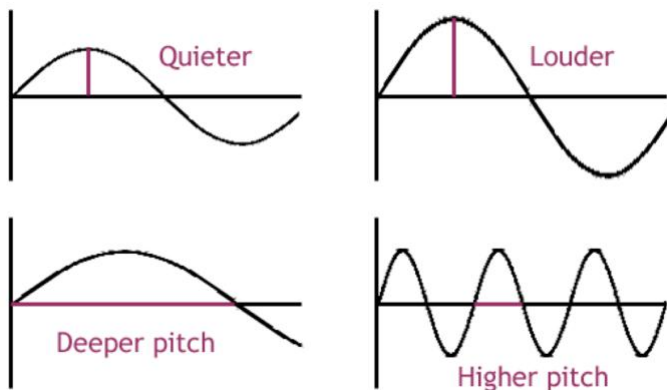
Besides colour, three-dimensional shape is the other primary irreducible element within vision, constituting what in cognitive science is known as Gestalt (Rollinger & Ierna, 2019). This term is also applicable to proportional ‘shapes’ or patterns within other qualia, like a musical melody. Visual shape can be geometrically reduced to lines/sides (1D), corners, surfaces (2D), angles and volumes (3D), making use of numerals to specify the quantities of these elements, as well as spatial prepositions to indicate relative location. This, however, will be further expounded upon in chapter 4 of the book, on Atlan’s numerals and mathematics.

Sound

CHAPTER 5. OUR ONTOLOGY

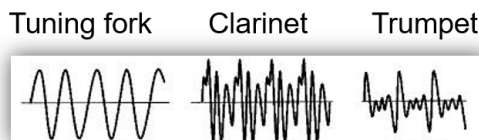
Sound is commonly divided into three elements:

- The frequency of the soundwave, phenomenally corresponding with *pitch*;
- The amplitude of the soundwave, phenomenally corresponding with *volume*;
- The shape of the soundwave, phenomenally corresponding with *timbre*.



Frequency visualised. From Mata (2015).

5.3. PHENOMENOLOGICAL ONTOLOGY AND QUALIA



Timbre. From SimplifyingTheory (n.d.).

Music theory almost universally divides pitch into a scale of notes, which repeats at the octave, where the ratio to the first note of the previous scale is 2:1. Western music uses the *diatonic* scale, comprised of specifically alternating whole and half steps, and which is *heptatonic*

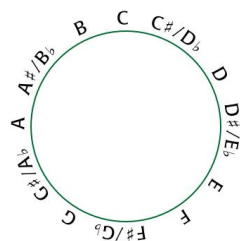


Figure 3: The Chromatic Circle. From Wikimedia (2022).

because it is comprised of 7 notes, before repeating at the octave (8th). There are a total of 12 half note steps within an octave, which comprise the *chromatic* scale. Chromatic notes are specified relative to the *natural* (♮) heptatonic scale, by indicating a half step lower by a *flat* (♭) and a half step higher by a *sharp* (♯).

The diatonic scale is not the most prevalent musical scale in the world, however, but rather the *pentatonic*

scale (five notes) (Encyclopædia Britannica, inc., n.d.). However, a pentatonic scale fits neatly within a heptatonic scale, for example: C–D–E–G–A is a pentatonic scale which only omits the notes F and B. Because of this, Atlan shall use a system based on the Western heptatonic scale, while remaining culturally neutral because this system also accommodates the more common pentatonic scale. If it were to use the pentatonic scale as the standard, the heptatonic scale would be an extended modification, which would be quite clumsy seeing as Western music is so vast and globalised.

Notes will be indicated with a context-dependent optional semantic prime for *note/pitch* in front of the note name. The combination *note* + the first 7 numerals (see chapter 6.2) can be taken to represent the 7 notes (heptatonic: *IP, OP, UP, IK, OK, UK, IM*, pentatonic: *IP, OP, UP, OK, UK*). This will also make calculating musical intervals more intuitive because it will only require simple mental subtraction.

Sharp will be indicated by *pitch-positive* (also designating *high* in the context of sound frequency) and flat (also indicating *low* in the context of sound frequency) by *pitch-negative*. Double sharps (!musDoubleSharp) and

5.3. PHENOMENOLOGICAL ONTOLOGY AND QUALIA

double flats (!musDoubleFlat) could be created by reduplicating *positive/negative* respectively. Microtonal notes (which fall in between the chromatic notes) could be accounted for as follows: half-sharp (half sharp) -PN, half-flat (half flat) -NP, three-quarter-sharp (three quarter sharp) -PNP, three-quarter-flat (three quarter flat) -NPN. Major and minor could be designated by the semantic primes for *happy* and *sad* (see the subchapter Emotions), and the seven scale modes could simply be numbered. Harmonic music theory is way too elaborate and complicated to cover fully in this paragraph, however it could be fairly easily constructed from the building blocks presented here.

Volume is a lot simpler: within music theory it is known as dynamic, and divided into loud (Italian: *forte*, *f*) and soft (Italian: *piano*, *p*), further nuanced by the prefix medium (Italian: *mezzo*), and by introducing a three-degree scale of intensity (*ppp*, *pp*, *p*, *mp*, *mf*, *f*, *ff*, *fff*). Adopting this system, Atlan can specify volume by combining the semantic prime for *volume* with *positive*, *medium*, *negative*, and a *comparative/superlative* system, which will also be applied in other places of the language: *X*, *more X*, *most X*. Changes in dynamic (growing louder, *crescendo*,

or softer, *decrescendo*) can be described by the semantic primes for *becoming* combined with *more-volume-positive/negative*.

Finally, timbre, corresponding with the specific shape of the soundwave pattern, has an almost infinite range of possible combinations. This is why in language, terms that denote timbre are always metaphoric approximations, describing the sound with words that denote phenomena unrelated to sound when taken literally, but have a similar phenomenal quality to the sound (e.g., piercing, empty, dark, bright), or are related to the origin of the sound (e.g., nasal, metallic). P. Sesuni analysed 45 studies on different timbre terms in English, Japanese, French, Czech, Swedish, Dutch, Finish, Spanish and German, and from these, identified 59 different descriptors (see appendix 1) (Susini, Carron, Rotureau, Dubois, & Misdariis, 2017).

Because these are all semantically reducible to non-sound-related terms, Atlan will not have any timbre-specific semantic primes, but rather use these and other timbre-descriptors, preceded by the semantic prime denoting *sound*, combined with an adjective-marker. When occurring in a clearly sound-related context, the *sound* prime

5.3. PHENOMENOLOGICAL ONTOLOGY AND QUALIA

may even be omitted, as there might not be any semantic confusion when it is already obvious that the adjective refers to a sound.

Taste and Scent

Taste and scent are strongly correlated because both rely on chemoreceptors (molecule-detectors) (Reina, 2022), with the main difference between the two that taste is perceived by the tongue and concerns solid and liquid matter, while smell goes through the nose and pertains to gaseous matter. Because many tastes have an olfactory counterpart, tastes may be marked by adding the semantic prime *taste* and scents with *smell*.

The five main taste categories are *sweet*, *sour*, *bitter*, *salt*, *umami* (Deutsch, 2019), which will be separate semantic primes in Atlan. These may double as scents when marked with *smell* instead of *taste*. The true range of scents and tastes, just like timbre, is overwhelmingly complex, thanks to the many different possible molecules and combinations between these. Thus, further nuances in flavour and aroma may be constituted in similar fashion to timbre, referring either to the source of the smell or taste (e.g., floral, alcoholic, vanilla-like), or a comparable quality (e.g., harsh, sharp, mellow). Spice is not a taste,

but rather a form of phenomenal pain, because it is registered by chemical nociceptors (molecule-pain receptors). Spice can thus be constituted by combining the semantic prime for *pain* with *taste/smell*. Of course, different taste or scent designators can be combined at will to create even more nuanced descriptions.

5.3.4. - *Physical sensation*

The physical senses cover a broad range of different sensations in different parts of the body (Reina, 2022). All terms for physical sensation shall be preceded by the semantic prime for *feeling*. This prime might also be used psychologically/affectively in other contexts. Mechanoreceptors sense physical deformation like pressure, touch, stretch and motion but also sound, corresponding to vibrations of the ear drum. Because sound has already been covered, it will not be counted among physical sensation. Combined with the range *positive-neutral-negative*, these will yield the following terms:

- *Contact*: pressure – touch – barely touch
- *Tension*: stretched – relaxed – contracted
- *Texture*: rough – normal – smooth

5.3. PHENOMENOLOGICAL ONTOLOGY AND QUALIA

More detailed textural descriptions can be made by combining with different material primes mentioned in the conclusion of this essay. Terms for motion of touch will be mentioned in chapter 5.4.

Thermoreceptors report *temperature* and nociceptors pain, as stated earlier regarding the taste of spice. *Pleasure* does not have a specific receptor but is rather constituted by neurotransmitters fired in the brain in reaction to certain perceptions, however, is often seen as the opposite polarity of pain (Kringelbach & Berridge, 2009). These will thus we constituted as follows:

- *Temperature*: hot – tepid – cold
- *Valence*: pleasure – neutral – pain
- *Wetness*: wet – moist/damp – dry

Somatic sense refers to the outside of the body, and visceral sense to internal organs. These will be codified by the semantic primes for *outside* and *inside* respectively.

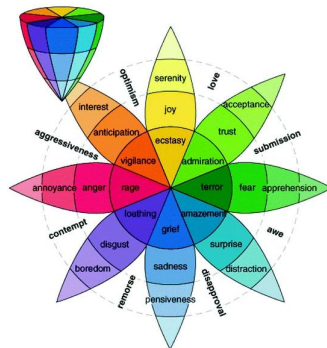
- Internal *tension*: bloating/swollen – normal pressure – cramp

CHAPTER 5. OUR ONTOLOGY

Proprioceptors sense the relative position of body parts, and the vestibular system registers the orientation of the entire body in space, perceived as *balance* (Proske & Gandevia, 2012).

- *Balance*: grounded – balanced – out of balance.
- Dizziness: *balance-negative + turning*.
- Motion sickness: *balance-negative + motion*
- Seasickness *balance-negative + sea*.

5.3.5. - Emotions



Plutchik's wheel of emotions. From Gkonou et al. (2015)

Plutchik's wheel of emotions with basic emotions and derivative... | Download Scientific DiagramIn 1980, the

5.3. PHENOMENOLOGICAL ONTOLOGY AND QUALIA

American psychology professor Robert Plutchik published a geometric model of the different basic emotions (Plutchik, 1980). Each pair of geometrically opposing sections constitute emotional antipoles. The justification for his classification is psychoevolutionary and traces the origins of the different affects to behavioural responses that would emerge naturally in reaction to different challenges and attractors encountered by humans and other animals (see appendix 2) (Plutchik & Kellerman, Theories of emotion, 1980).

Again, using the grid of positive – neutral – negative, the eight basic emotions and their respective three degrees of intensity can be covered:

1. Anger
2. Anticipation
3. Joy
4. Trust
5. Fear
6. Surprise
7. Sadness

8. Disgust

Intermediary emotions can be achieved by combining with other emotions or semantic primes (see appendix 3).

5.3.6 - *Consciousness states*

The different consciousness states of the human psyche are the hardest qualia to pin down. The NOBA Project, an online Psychology teaching platform, identifies the following (Teener & Teeny, n.d.):

1. Consciousness
2. Awareness
3. Hypnosis
4. Dissociation
5. Trance / depersonalisation
6. Sleep
7. Hallucination
8. Depression
9. Stimulation / excitation

5.3. PHENOMENOLOGICAL ONTOLOGY AND QUALIA

To complete the pallet of consciousness, I shall add the following:

1. Reason / abstract logos / logic
2. Memory
3. Desire / will
4. Conscience (moral)
5. Intuition / instinct
6. Imagination / creativity
7. Understanding

These can be combined with various other semantic primes to constitute various psychological terms (see appendix 4).

To sum up this chapter, to cover most all human qualia, Atlan will employ a set of qualia-specific atoms combined with some more general semantic atoms (see appendix 5).

5.4 Universal Semantics

Now that the basic constituents of human psychic experience have been accounted for, the different concepts within mankind's material understanding of the world have been largely neglected. This is what this chapter aims to rectify. In needing to satisfy Atlan's first constraint (human universality / cultural-linguistic neutrality), the semantic inventory would need to contain semantic primes or syntheses of these that cover a set of concepts shared cross-linguistically.

The comparative linguist Morris Swadesh published his commonly used list of 207 lexicostatistically universal concepts in 1952 (see appendix 6) (Swadesh, 1952), after making a series of revised versions. It is used to compare relatedness between languages by analysing the quantitative overlap in their words for the different Swadesh terms. Atlan builds on this list as a guideline, grouping related concepts, reducing some and adding other terms where possible (see chapter 6 of the book).

Another, more abstracted summary of the basic semantic elements of human language is given by the Natural Semantic Metalanguage developed by the cross-cultural

linguist Anna Wierzbicka (Goddard & Wierzbicka, *Meaning and Universal Grammar: Theory and Empirical Findings.*, 2002). Though different languages might express concepts in different ways, the semantic content of NSM is divided into 65 semantic primitives, spread over 16 categories (see appendix 7) (Levisen & Waters, 2017).

Atlan uses this classification as a guideline, and the semantic primitives will be synthesised with the Swadesh list. Within NSM, semantic ‘molecules’ are terms that can be reduced to the 65 primitives, or ‘atoms’, but are often used to build more complicated meanings. Where the atoms are abstract, the molecule are more concrete. There is, therefore, an added value of listing such molecules to minimise complexity, because molecule-composite words would be a lot more longwinded when all their semantic atoms had to be stated individually. The research into this topic is still underdeveloped, but a few sets of supposedly universal semantic molecules have been proposed. The website of the university of Griffith mentions the following universal semantic molecules (see appendix 8) (Griffith University, n.d.).

This lexicon already bears a striking similarity to the Swadesh list. Cliff Goddard identifies a few additional

semantic molecules in English (Goddard, 2012). These will be incorporated because of the globalised distribution of Anglo-American culture and language. In a video-essay he adds several other mostly culture-bound molecules (NSMLab, 2021). Atlan will synthesise these lists into the semantic inventory (see appendix 9).

5.5 Concluding remarks

In this chapter, I have introduced the problem of an oligosynthetic ontology to constitute Atlan's semantics in a way that respects the constraints of cultural neutrality, unambiguity and form from function. I discussed previous attempts at a system like this and other systems of classification, and the critical discourse around these. I discussed the role of the human perspective within human language, and the relationship between language ontology and subjective thought. I dove into the philosophical movement of phenomenology and argued for the irreducibility of qualitative experience (qualia). I went on to map out the state-space of qualitative experience, and mapped out how different qualia could be reduced into an oligosynthetic combinatorial system of word-generation.

I then referenced different academic projects that mapped out a minimal account of cross-culturally universal irreducible concepts, which will be added to the language in order to respect the constraint of cultural neutrality and the premise of linguistic anthropocentricity. Chapter 6.2 of the book contains the final semantic inventory of Atlan's 490 semantic syllable-primers, sorted into conceptually related semantic categories. These are sorted into having maximally similar initial letters and modelled by AI to be as similar as possible to various different natural languages, weighed by their linguistic genealogy and total amount of speakers, making the lexicon semi-a priori sourced.

5.6 Appendix

2: Psycho-evolutionary classification of animal emotions

The Complex, Probabilistic Sequence of Events Involved In the Development of an Emotion				
Stimulus event	Inferred cognition	Feeling	Behaviour	Effect
Threat	"Danger"	Fear, terror	Running, or flying away	Protection
Obstacle	"Enemy"	Anger, rage	Biting, hitting	Destruction
Potential mate	"Possess"	Joy, ecstasy	Courting, mating	Reproduction
Loss of valued person	"Isolation"	Sadness, grief	Crying for help	Reintegration
Group member	"Friend"	Acceptance, trust	Grooming, sharing	Affiliation
Gruesome object	"Poison"	Disgust, Loathing	Vomiting, pushing away	Rejection
New territory	"What's out there?"	Anticipation	Examining, mapping	Exploration
Sudden novel object	"What is it?"	Surprise	Stopping, alerting	Orientation

3. Combinatorics of Atlan's emotional and semantic atoms

- Optimism = anticipation + joy
- Love = joy + trust
- Shame embarrassment = fear + disgust
- Thoughtfulness = serene + interest
- Thankfulness = serene + acceptance
- Pride = admire + self
- Faith / belief = trust + know
- Extravagance = ecstasy + distracted
- Daringness = trust + anticipation
- Rejection / refusal = not + accepting
- (In)security = (not +) trust + self
- Discouraged = passive + not + trust + anticipation
- Bewildered = surprise + apprehension
- Critical / sceptical = not + trust + know

CHAPTER 5. OUR ONTOLOGY

- Frustration = anger + distraction
- Jealousy = desire + annoyed

5. Combinatorics of Atlan's consciousness and semantic atoms

- Unconscious = not + consciousness
- Sub-conscious = below + consciousness
- Ego = feeling + self
- Self-consciousness = consciousness + self
- Narcissism = admiration + self
- Selfishness / egotism = interest + self
- Depersonalisation = feeling + not + self
- Ego-death = feeling + self + death
- Derealisation = feeling + not + contact + reality
- Libido = desire + sex
- Arousal = stimulation + sex
- Orgasm = ecstasy + sex
- Deep sleep = sleep + not + consciousness
- Dreaming = sleep + consciousness
- Lucid dreaming = dreaming + con-

- | | |
|---|---|
| consciousness + awareness | • Dysphoria = feeling + bad |
| • Enlightenment = consciousness + light / bright | • High = feeling + cannabis + excitement + positive |
| • Bliss = ecstasy + peace | • Stoned = feeling + cannabis + excitement + negative |
| • Mystical experience = consciousness + God | • Tipsy = feeling + alcohol + neutral |
| • Sensory overload = feel + excitement + positive | • Drunk = feeling + alcohol + positive |
| • Peace = excitement + neutral | • Understand = reason + grasp |
| • Numbness = feel + excitement + negative | • Aha-Erlebnis = feeling + understanding |
| • Euphoria = feeling + good | • Empathy = feeling + other |
| | • Social awareness = awareness + social |

CHAPTER 5. OUR ONTOLOGY

- Intelligent = reason + positive
 - Dumb = reason + negative
 - Guilt = conscience + bad
 - Know-how = understanding + action
 - Wisdom = understanding + life
5. Overview of qualia-related atoms
- Negative
 - Neutral
 - Positive
 - Colour
 - Brightness
 - Red
 - Yellow
 - Blue
 - Sound
 - Volume
 - Become, transform
 - Note/pitch
 - 7 note-names corresponding with the numbers 1-7
 - (...)
 - More (comparative)
 - Most (superlative)
 - Smell
 - Taste
 - Sweet

- Sour
- Bitter
- Salty
- Umami
- Feeling / affect
- Contact
- Tension
- Texture
- Temperature
- Valence
- Wetness
- Outside
- Inside
- Balance
- Turn
- Move
- Sea
- Anger
- Anticipation
- Joy
- Trust
- Fear
- Surprise
- Sadness
- Disgust
- Consciousness
- Awareness
- Hypnosis
- Dissociation
- Trance/depersonalisation

CHAPTER 5. OUR ONTOLOGY

- Sleep
- Hallucination
- Depression
- Stimulation / excitation
- Reason / abstract logos
- Memory
- Desire
- Conscience (moral)
- Intuition / instinct
- Below
- Self
- Death
- Reality
- Sex
- Not
- Light
- God
- Value judgement
- Grasp
- Social

6. Swadesh 207 list of universal human concepts

1. I
2. you (singular)
3. they (singular)
4. we
5. you (plural)
6. they (plural)
7. this
8. that
9. here

- | | |
|-----------|-----------------------|
| 10. there | 25. four |
| 11. who | 26. five |
| 12. what | 27. big |
| 13. where | 28. long |
| 14. when | 29. wide |
| 15. how | 30. thick |
| 16. not | 31. heavy |
| 17. all | 32. small |
| 18. many | 33. short |
| 19. some | 34. narrow |
| 20. few | 35. thin |
| 21. other | 36. woman |
| 22. one | 37. man (adult male) |
| 23. two | 38. man (human being) |
| 24. three | 39. child |

CHAPTER 5. OUR ONTOLOGY

- | | |
|-------------|----------------------|
| 40. wife | 55. seed |
| 41. husband | 56. leaf |
| 42. mother | 57. root |
| 43. father | 58. bark (of a tree) |
| 44. animal | 59. flower |
| 45. fish | 60. grass |
| 46. bird | 61. rope |
| 47. dog | 62. skin |
| 48. louse | 63. meat |
| 49. snake | 64. blood |
| 50. worm | 65. bone |
| 51. tree | 66. fat (noun) |
| 52. forest | 67. egg |
| 53. stick | 68. horn |
| 54. fruit | 69. tail |

70. feather	85. belly
71. hair	86. guts
72. head	87. neck
73. ear	88. back
74. eye	89. breast
75. nose	90. heart
76. mouth	91. liver
77. tooth	92. to drink
78. tongue (organ)	93. to eat
79. fingernail	94. to bite
80. foot	95. to suck
81. leg	96. to spit
82. knee	97. to vomit
83. hand	98. to blow
84. wing	99. to breathe

CHAPTER 5. OUR ONTOLOGY

- | | |
|---------------|-----------------------------|
| 100. to laugh | 115. to split |
| 101. to see | 116. to stab |
| 102. to hear | 117. to scratch |
| 103. to know | 118. to dig |
| 104. to think | 119. to swim |
| 105. to smell | 120. to fly |
| 106. to fear | 121. to walk |
| 107. to sleep | 122. to come |
| 108. to live | 123. to lie (as in a bed) |
| 109. to die | 124. to sit |
| 110. to kill | 125. to stand |
| 111. to fight | 126. to turn (intransitive) |
| 112. to hunt | 127. to fall |
| 113. to hit | 128. to give |
| 114. to cut | 129. to hold |

- | | |
|-----------------|----------------|
| 130. to squeeze | 145. to freeze |
| 131. to rub | 146. to swell |
| 132. to wash | 147. sun |
| 133. to wipe | 148. moon |
| 134. to pull | 149. star |
| 135. to push | 150. water |
| 136. to throw | 151. rain |
| 137. to tie | 152. river |
| 138. to sew | 153. lake |
| 139. to count | 154. sea |
| 140. to say | 155. salt |
| 141. to sing | 156. stone |
| 142. to play | 157. sand |
| 143. to float | 158. dust |
| 144. to flow | 159. earth |

CHAPTER 5. OUR ONTOLOGY

160. cloud	175. white
161. fog	176. black
162. sky	177. night
163. wind	178. day
164. snow	179. year
165. ice	180. warm
166. smoke	181. cold
167. fire	182. full
168. ash	183. new
169. to burn	184. old
170. road	185. good
171. mountain	186. bad
172. red	187. rotten
173. green	188. dirty
174. yellow	189. straight

190. round	199. right
191. sharp (as a knife)	200. left
192. dull (as a knife)	201. at
193. smooth	202. in
194. wet	203. with
195. dry	204. and
196. correct	205. if
197. near	206. because
198. far	207. name

7. NSM atoms

Category	Primes
Substantives	I, YOU, SOMEONE, PEOPLE, SOMETHING/THING, BODY
Relational Substantives	KIND, PART
Determiners	THIS, THE SAME, OTHER ELSE ANOTHER
Quantifiers	ONE, TWO, SOME, ALL, MUCH/MANY, LITTLE/FEW
Evaluators	GOOD, BAD
Descriptors	BIG, SMALL
Mental predicates	THINK, KNOW, WANT, DON'T WANT, FEEL, SEE, HEAR
Speech	SAY, WORDS, TRUE
Actions, Events, Movement	DO, HAPPEN, MOVE
Existence, Possession	BE (SOMEWHERE), THERE IS, BE (SOMEONE/SOMETHING), (IS) MINE
Life and Death	LIVE, DIE
Time	WHEN/TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT
Space	WHERE/PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE, TOUCH (CONTACT)
Logical Concepts	NOT, MAYBE, CAN, BECAUSE, IF
Intensifier, Augmentor	VERY, MORE
Similarity	LIKE/AS/WAY

8. Proposed universal NSM modules

CHAPTER 5. OUR ONTOLOGY

Category	Primes
Body parts	hands, mouth, eyes, head, ears, nose, face, teeth, fingers, breast, skin, bones, blood
Physical	long, round, flat, thin, hard, soft, sharp, smooth, heavy
Spatial / physical	be on something, at the top, at the bottom, in the middle, in front of, around
Environmental	sky, the Earth, sun, moon, stars, ground, during the day, at night
Times	day
Fire and water	water, fire
Biological	creature, grow, egg, tail, wings, feathers
Biosocial	children, men, women, be born, mother, father, wife, husband
Materials	wood, stone
"Knowing" and "naming"	know (someone), be called
"Doing"	hold, make, kill, breathe, sleep, sit, lie, stand, play, laugh, sing

9. Culture-specific NSM modules

Category	Primes
Anatomy	voice, to grow, sick/ill, soul
Physical acts	pick up
Familial	brother, sister
Topology	edge, hole
Places	country, home, school, village/town/city
Material	metal, glass, paper
Mechanical	wheel, pipe, wire, engine, electricity, machine
Major cultural concepts	money, book, read, write, number, science, law, king/chief, teacher, doctor

Chapter 6

Lexicon

6.1 Introduction

According to research by First Education (2022), the Dutch are positioned as the most proficient English-speaking population globally. Other countries that achieved a position in the top 10 ranking were Denmark, Belgium, Sweden, Finland, and Germany. While a proud Dutchman may attribute this success to hard work and dedication, it is worth considering other factors that could be at play here. Notably, these countries with commendable English proficiency are all located in North Europe and speak very similar languages.

Linguistics offers a unique perspective on the relationships between different languages. By comparing the vocabulary, grammar and sound systems of various languages, researchers have identified related language families and have constructed language family trees to illustrate the evolution and divergence of languages over time. Thanks to that research, we know that many countries with notable position in the English Proficiency Index share a common linguistic background. Such a linguistic background, or language family, thus may provide a foundation for proficiency in a new language.

This would imply that certain languages are easier to learn for certain population groups. For Atlan, it is deemed important that it will become a language that is easy and quick to learn for *everybody*. This is a challenging task but might be achievable if we find some sort of shared background between almost every natural language. If it is possible to find words that look similar in different languages, which are known as cognates, the translation for those words in Atlan can be designed to resemble them as much as possible. With a model that can do this on a large scale, Atlan will become easy, neutral, and global.

6.2. COMPARISON METHODS AND LANGUAGE FAMILIES – JEP ANTONISSE

To achieve this, it is first key to create some understandings of what methods are used to compare different languages. Therefore, we will take a closer look at the so-called cosine similarity. Thereafter, it is necessary to conduct an examination of the existing language families that exist. In that way, we gain a deeper insight into the connections between existing natural languages. In addition, with that gained understanding, it is possible to decide which language we will make available in contributing to the process of cognate finding. The most spoken languages are weighed against each other to create a dataset that is representative of the real world. All these pieces of the puzzle come together in the final part of this chapter, where the computer program that we used to generate words in Atlan will be discussed.

6.2 Comparison methods and language families – Jep Antonisse

6.2.1 Cosine Similarity

According to research by First Education (2022), the Dutch are positioned as the most proficient English-speaking

population globally. Other countries that achieved a position in the top 10 ranking were Denmark, Belgium, Sweden, Finland, and Germany. While a proud Dutchman may attribute this success to hard work and dedication, it is worth considering other factors that could be at play here. Notably, these countries with commendable English proficiency are all located in North Europe and speak very similar languages.

Linguistics offers a unique perspective on the relationships between different languages. By comparing the vocabulary, grammar and sound systems of various languages, researchers have identified related language families and have constructed language family trees to illustrate the evolution and divergence of languages over time. Thanks to that research, we know that many countries with notable position in the English Proficiency Index share a common linguistic background. Such a linguistic background, or language family, thus may provide a foundation for proficiency in a new language.

This would imply that certain languages are easier to learn for certain population groups. For Atlan, it is deemed important that it will become a language that is easy and quick to learn for *everybody*. This is a chal-

lenging task but might be achievable if we find some sort of shared background between almost every natural language. If it is possible to find words that look similar in different languages, which are known as cognates, the translation for those words in Atlan can be designed to resemble them as much as possible. With a model that can do this on a large scale, Atlan will become easy, neutral, and global.

To achieve this, it is first key to create some understandings of what methods are used to compare different languages. Therefore, we will take a closer look at the so-called cosine similarity. Thereafter, it is necessary to conduct an examination of the existing language families that exist. In that way, we gain a deeper insight into the connections between existing natural languages. In addition, with that gained understanding, it is possible to decide which language we will make available in contributing to the process of cognate finding. The most spoken languages are weighed against each other to create a dataset that is representative of the real world. All these pieces of the puzzle come together in the final part of this chapter, where the computer program that we used to generate words in Atlan will be discussed.

Text	Frequency “Merry”	Frequency “christmas”
“Merry christmas ”	1	1
“Christmas”	0	1

Table 6.1: Word-appearance in “Merry” and “Merry christmas”.

This table can be visualized in a two-dimensional array, where on each axis the count of a word is represented. Now both texts can be placed as a dot on this grid accordingly. Drawing two lines from each point to the origin of the grid creates an angle between those lines. This angle at the origin can be calculated, in this case it would be 45° . To finish the cosine similarity, all that is needed now is to take the cosine of this angle, in this example $\cos(45) = 0.71$.

If the compared sentences are identical, the two dots would be placed on the same place in the grid. Thus, the lines towards the origin would fall precisely over each other. Therefore, the angle between both lines would be 0° , resulting in a cosine score of $\cos(0) = 1$. On the other hand, if both sentences have not a single element in common, the lines would be perpendicular to each other. With an angle of 90° , the cosine similarity would return $\cos(90) = 0$. Thus, in any case, the cosine similar-

ity gives a score between 0 and 1, showing the degree of similarity.

As another example, let's compare two words: 'Bert' to 'Ernie'. Instead of words, the vector now can be made up of letters. In a table, this would look like this:

Name	E	R	N	I	B	T
Ernie	2	1	1	1	0	0
Bert	1	1	0	0	1	1

Table 6.2: Letter frequency in the words "Bert" and "Ernie".

With six different letters occurring, it would be possible to place both 'Bert' and 'Ernie' in a six-dimensional grid and draw the lines to the origin. However, it is impossible for humans to visualize a six-dimensional graph. Thus, we need a new way to calculate the angle between the two vectors. Luckily, there exists a formula to compute the cosine similarity.

$$\textit{Cosine similarity } A, B = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

6.2. COMPARISON METHODS AND LANGUAGE FAMILIES – JEP ANTONISSE

In less mathematical terms, what this means is that each element of the two vectors A and B are compared. The products of all the elements are then summed up and divided by the length of both the vectors together. If we fill in the numbers for our Bert-Ernie-comparison, it will look like this:

$$\frac{1 \times 2 + 1 \times 1 + 1 \times 0 + 1 \times 0 + 0 \times 1 + 1 \times 0}{\sqrt{2^2 + 1^2 + 1^2 + 1^2 + 0^2 + 0^2} \sqrt{1^2 + 1^2 + 0^2 + 0^2 + 1^2 + 1^2}} = 0.56 \quad (6.1)$$

Nevertheless, converting the words based solely on letter frequency inadvertently results in losing vital information about the arrangement of the letters. This preservation is of utmost importance for our project since we try to find patterns and therefore adjacent letter combinations. To address this concern, we introduced a slight adjustment to the regular cosine similarity, where each index with the same character in both words also scores a small point. In this way, our cosine similarity tries to reward words that have the same letters on the same place.

6.2.2 Language families

The use of various comparison methods, similar to the cosine similarity, allowed linguists to identify several groups of related languages. These groups, or language families, were categorized based on common linguistic features and a shared common ancestor (Campbell, 2018). Such a common Proto-Language allows researchers to trace origins of various languages to a single root. However, this does not necessarily need to be the case. There exist languages for which it is impossible to classify them as part of a language family, such as Basque (Campbell, 2010). Researchers speculate that it might be possible that these languages, known as language isolates, might have had related languages in the past, that went extinct unrecorded. Therefore, these languages now form their own language family, with them being the only member. On the other hand, not only genetic proximity between languages is enough to be placed in the same language family. Languages that are constructed instead of naturally developed cannot be considered part of any language family, since they do not have a shared ancestor with any other language (Campbell, 2018).

6.2. COMPARISON METHODS AND LANGUAGE FAMILIES – JEP ANTONISSE

Although this means that the total number of language families in the world might be in the hundreds, not all are equally relevant today. To begin with, 94 language families are extinct, meaning there is a lack of any surviving speakers (Campbell, 2018). In addition, the number of languages and the number of speakers differ largely. There are five language families that can be considered as the main language families of the world. Every single one of these languages contains at least 5

The most widely spoken language family, with over 3 billion speakers worldwide today, is Indo-European. When Sir William Jones first spoke of this family, he proposed there were several branches with related languages (Fortson, 2011). First, there is *Indo-Iranian*, spoken in the middle east, where the languages Sanskrit, Persian and Pashto are placed. Secondly, *Italic*, with languages such as Latin, Italian and Spanish. Third, the languages of the Northern parts of Europe were placed in the *Germanic* branch. A fourth branch called *Celtic* housed the languages of the island of Great Britain, such as Irish and Welsh. Lastly, Jones portrayed one branch on its own for *Greek*.. Only after thirty years would this division be altered, when researchers added three more branches to

CHAPTER 6. LEXICON

the family. The largest new branch was called the *Balto-Slavic* branch, including languages such as Russian, Ukrainian and Czech. The two remaining branches both only contained one language, *Armenian* and *Albanian*. This family tree remained the same to this day, except for the addition of two branches with extinct languages discovered in the first half of the 20th century, called *Anatolian* and *Tocharian*.

The Sino-Tibetan language family is, even though it has more daughter languages than Indo-European, the second most spoken family, with around 1.3 billion speakers. This family can be split up into two major subgroups: *Chinese* and *Tibeto-Burman* (Shafer, 1955). The Chinese subgroup is like a family on its own, made up of different but related dialects. The largest and best-known ones are Mandarin and Cantonese. Tibeto-Burman can be branched into a few branches on its own: Tibetic (Tibetan), Burmese-Lolo (Burmese and various Lolo-languages) and Karenic.

Niger-Congo is the third most spoken family, containing the highest number of languages: over 1,500 languages are known ancestors of the Proto-Niger-Congo language. Although this number is nearly twice that of

6.2. COMPARISON METHODS AND LANGUAGE FAMILIES – JEP ANTONISSE

Indo-European, it is spoken by 600 million people, due to the immense language diversity in the Sub-Saharan Africa region (Heine et al., 2000). The largest branch in the Niger-Congo family is called the *Atlantic-Congo* branch. Herein are numerous languages spoken in West Africa, such as Yoruba and Igbo. Also, Swahili, mostly spoken in the Eastern part of Africa, falls into this category. The languages spoken in the Central and Southern parts of Africa are mostly from another branch, called the *Bantu-Congo* branch. These are languages such as Zulu, Xhosa and Shona. Other branches are the *Kordofanian* branch (Katla, Moro and Talodi) and *Mande* branch (Bambara, Mandinka and Soninke)

Austronesian, with a similar high diversity as Niger-Congo, covers the languages found in the region that stretches from Southeast-Asia to the Pacific Island. In total this family contains more than 1,200 languages and is spoken by approximately 326 million speakers, primarily in countries such as Indonesia, Malaysia and the Philippines. The most important subgroup within this family is the *Formosan* branch, forming a total of nine distinct branches (Tryon, 1995). These branches are all made up of the different indigenous languages of Taiwan. None

of them are, however, the most widespread or diverse branch of the family. This is namely the tenth branch, known as the *Malayo-Polynesian* branch, encompassing Indonesian, Javanese and Sundanese.

Lastly, the languages mostly spoken in the North and the Horn of Africa and Southwest Asia are grouped in the Afroasiatic language family. This family consists of several branches (Huernergard, 2004). The one that houses the best-known languages is the *Semitic* branch, including Arabic, Amharic and Hebrew. Another large branch is *Berber*, with languages such as Tamazight and Kabyle. Smaller branches are the *Cushitic* branch, which comprises of languages such as Oromo, Somali and Afar, and the *Chadic* branch, with as largest language Hausa. The languages in the Afroasiatic family combined are spoken worldwide by almost 600 million people.

6.3 Using cognates to generate words

A study by Otwinowska and Szewczyk (2018) argued that cognates, similar sounding words with the same meaning in different languages, are the easiest words to learn when learning a new language. The resemblance with

6.3. USING COGNATES TO GENERATE WORDS

your mother tongue makes the words much easier to remember and use than non-cognates words. By designing Atlan in a way that it has a lot of these cognates, we try to keep the trouble of learning Atlan as low as possible. To achieve this goal, it is key to make new Atlan words resemble existing words, or patterns in existing words, as much as possible.

The idea of using cognates to generate new words for vocabulary is also used in the creation of the constructed language Lojban (Cowan, 1997). Lojban proposed new words, or ‘gismu’s’ and looked for words that looked similar to it in the languages Chinese, English, Spanish, Hindi, Russian and Arabic. If three or more letters were the same and in the same order as a word in the source language, the gismu would score points. For resemblance with larger language a gismu could score more points, meaning that large languages were viewed as more important. The amount of influence each language had, in other terms the ‘weight’, was solely based on the number of speakers in 1985.

For Atlan we have built a similar program, which we will call *Lexi* from now on. To understand how Lexi works, it is wise to split the process into three parts: the lan-

guage selection, the weights, and the program itself.

6.3.1 Language selection

The choice made by the developers of Lojban to use the six largest languages was good in terms of significance. Their language set closely resembles the set of UN languages: Chinese, English, Spanish, French, Russian and Arabic. These languages are already for 49,6% of all people either their mother tongue or second language and form an official language for more than half the states in the world, according to Ethnologue. However, the developers failed to take language families into account. This results in the facts that four out of the six languages used, or two thirds, are a descendent of Proto-Indo-European, while other large families such as Niger-Congo or Austronesian are not represented at all. Distributions so far away from the real world might make the result very Eurocentric. This creates a large group of language learners unable to match any words to their native language. For Atlan to improve on this, the number of languages that is used as a source must be increased.

Also, if the desired distribution should resemble the

6.3. USING COGNATES TO GENERATE WORDS

distribution of the real world, we need to know what the distributions in the real world *are*. The frequency of each language family in the 100 most spoken languages according to Ethnologue (2022) can provide a target percentage of how big the part of each language family should be in our program.

Now we will create a *language set* or *data set*, with in it all the languages we want to find cognates in. It is important that the cognate and the Atlan word has the same meaning in all these languages: otherwise, it might find similar looking words, but with different meanings in different languages, which are known as *false cognates*. These cognates are not a sign of a common ancestor but rather a display of randomness and luck. Also, these false cognates are the hardest words to learn in a new language, even harder than non-cognate words (Otwinowska et al., 2018), thus we should avoid creating those in Atlan if we can. Therefore, we should be able to control the meaning of the words in other languages.

Hence, we need translation software. We will use the public available library called Googletrans (3.0.0). This software supports translation into 107 different languages. Since we desire the same significance the language set

CHAPTER 6. LEXICON

of Lojban had, we can analyze which of these languages are present in the list of the 100 most spoken languages. The result can be viewed in table 6.3, along with some information about each language.

6.3. USING COGNATES TO GENERATE WORDS

Language	Number of Nativespeakers in Millions	Num
1	English	
2	Mandarin Chinese	
3	Hindi	
4	Spanish	
5	French	
6	Standard Arabic	
7	Bengali	
8	Russian	
9	Portuguese	
10	Indonesian	
11	Urdu	
12	Standard German	
13	Japanese	
14	Swahili	
15	Marathi	
16	Telegu	
17	Western Punjabi	
18	Tamil	
19	Turkish	
20	Korean	
21	Vietnamese	
22	Javanese	
23	Italian	
24	Hausa	
25	Thai	
26	210 Kannada	
27	Filipino	
28	Polish	

6.4 Vocabulary list – Stijn Janssens & Jep Antonisse

6.5 AI Generation – Max Geraerds & Jep Antonisse

6.6 Atlan to English

6.7 Translation Protocol

Chapter 7

Pragmatics – Jonathan Roose

7.1 Logophilia, Wordplay and Rhythm

7.2 Mood-markers

Chapter 8

Further suggestions

8.1 Sign Language – Stijn Janssens

8.2 Language Variation – Niek Elsinga

Chapter 9

Example texts

9.1 The Story of Babel – Jonathan Roose

9.2 The Epic of Gilgamesh – Niek Elsinga

Chapter 10

Exercises – Max Geraedts

Epilogue

10.0 Special Thanks

Bibliography

%nociteTowardstheDynamicArtofPartialComprehension

Bibliography

- Aitchison, J. (1994). *Language Change: Progress or Decay?* Cambridge University Press, pp. 627–637.
- Al-Ani, S.H. (1970). *Arabic Phonology: An acoustical and physiological investigation*. De Gruyter.
- Arivati, A. (2012). “The usefulness of metrics in the quantification of speech rhythm”. In: *Journal of Phonetics* 40.3, pp. 351–373.
- Ariza, M.A. and F.A. Navarro (2006). *Medicine: use of English*. Elsevier EBooks, pp. 752–759.
- Arouet, F.M. (1963 [1759]). *Candide, ou l’Optimisme*. Dieter Hildebrandt.
- Austin, J.L. (1975 [1955]). *How to Do Things With Words: The William James Lectures delivered at Harvard University in 1955*. Oxford University Press.
- awful, adj. and adv.* (Mar. 2023).

BIBLIOGRAPHY

- Baker, M. (1998). "The Polysynthesis Parameter". In: *Language* 74.1, p. 129.
- Bauer, R.S. and P.K. Benedict (1997). *Modern Cantonese Phonology*. Walter de Gruyter.
- Berlin, B. and P. Kay (1969). *Basic Color Terms: Their Universality and Evolution*. University of California Press.
- Bernal, B. and N. Altman (2010). "The connectivity of the superior longitudinal fasciculus: a tractography DTI study". In: *Magnetic resonance imaging* 28.2, pp. 217–225.
- Bieswas-Teener, R. and J. Teeny (n.d.). *States of consciousness*. Retrieved: April 9, 2023. URL: %5Curl%7Bhttps://nobaproject.com/modules/states-of-consciousness%7D.
- Bleek, W.H. (1862). *A Comparative Grammar of South African Languages*. Trubner & Co.
- Blevins, A.D. (n.d.). *The Analytical Language of John Wilkins*. Retrieved: April 8, 2023. URL: %5Curl%7Bhttps://andrewblevins.github.io/JohnWilkins/%C2%A0%7D.
- Booij, G. (2007). *The Grammar of Words: An Introduction to Linguistic Morphology*. OUP Oxford.

- Boroditsky, L. and M. Liberman (2010). *For and Against Linguistic Relativity*. Retrieved: 4 March 2023. URL: %5Curl%7Bhttps://web.archive.org/web/20120215074211/http://www.economist.com/debate/overview/190%7D.
- Boucon, Jean-Louis (2019). “The Philosophy of Language and the Ontology of Knowledge”. In: p. 18.
- Brown, R. and E. Lenneberg (1954). “A study in language and cognition”. In: *Journal of Abnormal and Social Psychology* 49.3, pp. 454–462.
- Burde, J. (2004). *Rituals, Mantras and Science: An Integral Perspective*. Motilal Banarsidass Publishers.
- Butcher, A.R. (november 2018). “The special nature of Australian phonologies: why auditory constraints on human language sound systems are not universal”. In: *Proceedings of Meetings on Acoustics* 176ASA 35.1.
- Bybee, J. (2015). *Language Change*. Cambridge University Press.
- Campbell, L. (2010). “Language isolates and their history, or, whats weird, anyway?” In: *Annual Meeting of the Berkeley Linguistics Society* 36.1, pp. 16–31.

BIBLIOGRAPHY

- (2018). “How many language families are there in the world?” In: *Anuario del Seminario de Filología Vasca “Julio de Urquijo”* 52.1/2, pp. 133–152.
- Carron et al., (2017). *Speaking about sounds: a tool for communication on sound features*.
- Chambers, J.K., P. Trudgill, and N. Schilling-Estes (2004). *The Handbook of Language Variation and Change*. Blackwell Publishing Ltd eBooks.
- Chomsky, Noam (2002). *Syntactic Structures*. Mouton de Gruyter.
- Chrisholm, H.E. (1911). *Cartouche*. 11th ed. Cambridge University Press.
- Cohen, M. (n.d.). *Definitions and the tree of Porphyry*. Retrieved: April 8 2023. URL: <http://faculty.washington.edu/smcohen/433/PorphyryTree.html%C2%A0%7D>.
- Comrie, B. (1989). *Language Universals and Linguistic Typology: Syntax and Morphology*. University of Chicago Press.
- Coupland, N. (1985). “Hark, hark, the lark’: Social motivations for phonological style-shifting”. In: *Language & Communication* 5.3, pp. 153–171.

- Cowan, J.W. (1997). *The Complete Lojban Language*. Vol. 15. Logical Language Group.
- Cruttenden, A. (1997). *Intonation*. Cambridge University Press.
- D., Crystal (2018). *The Cambridge Encyclopedia of the English Language*. Cambridge University Press.
- DAlembert, J.L. and D. Diderot (2020). *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers*. Books on Demand Ltd.
- Dalgano, G. (1968 [1661]). *Ars signorum*. Scolar P.
- DAndrade, R.G. (1995). *The Development of Cognitive Anthropology*. Cambridge University Press.
- Davis, W.A. (2016). *Implicature*. Vol. 6. Springer.
- De Saussure, F.[Translated by W. Baskin] (1959). *Course in General Linguistics*. Philosophical Library.
- Deleuze, G. and F. Guattari (2019 [1980]). *A Thousand Plateaus: Capitalism and Schizophrenia*. Bloomsbury.
- Denham, K. and A. Lobeck (2012). *Linguistics for everyone: An introduction*. Cengage Learning.
- Deutsch, L. (2019). *The Five Basic Tastes*. Retrieved: April 9, 2023. URL: %5Curl%7Bhttps://www.psychologytoday.com/us/blog/memory-catcher/201903/the-five-basic-tastes%C2%A0%7D.

BIBLIOGRAPHY

- Eco, U. Translated by J. Fentress (1994). *The Search for the Perfect Language*. Blackwell Publishers Inc.
- Education, First (2022). *EF EPI: EF English Proficiency Index: A Ranking of 111 Countries and Regions by English Skills*. EF Education First.
- Encyclopaedia Britannica (n.d.). *Common scale types*. Retrieved: April 9, 2023. URL: %5Curl%7Bhttps://www.britannica.com/art/scale-music/Common-scale-types%C2%A0%7D.
- Evans, N. and S.C. Levinson (2009). “The myth of language universals: Language diversity and its importance for cognitive science”. In: *Behavioral and brain sciences* 32.5, pp. 429–448.
- Fantini, A.E. (2020). “Language; An essential component of intercultural communicative competence”. In: *Routledge Handbook of Language and Intercultural Communication*. Routledge, pp. 267–281.
- Fitch, W.T., M.D. Hauser, and N. Chomsky (2005). “The evolution of the language faculty: Clarifications and implications”. In: *Cognition* 97.2, pp. 179–210.
- Fortson, IV and W. Benjamin (2011). *Indo-European Language and Culture: An Introduction*. John Wiley & Sons.

- Foucault, M. (2010 [1966]). *The Order of Things: An Archaeology of the Human Sciences*. Routledge.
- Franklin, J. (1986). "Aristotle on Species Variation". In: *Philosophy* 61.236, pp. 245–252.
- Gennaro et al., (2012). "The grammatical view of scalar implicatures and the relationship between semantics and pragmatics". In: *Handbook of Semantics*. Ed. by Portner et al. Mouton de Gruyter, pp. 2297–2332.
- Gerson, L.P. (2019 [430]). *Plotinus: The Enneads*. Cambridge University Press.
- Geschwind, N. (1972). "Language and the brain". In: *Scientific American* 226.4, pp. 76–83.
- Goddard, C. and A.E. Wierzbicka (2002). *Meaning and Universal Grammar: Theory and Empirical Findings*. John Benjamins Publishing Company.
- Goddard, Cliff (2012). "Semantic primes, semantic molecules, semantic templates: Key concepts in the NSM approach to lexical typology". In.
- Googletrans (n.d.). *Googletrans 3.0.0. documentation*. Retrieved February 20th, 2023.
- Gooskens, C. (2007). "The Contribution of linguistic factors to the intelligibility of closely related languages".

BIBLIOGRAPHY

- In: *Journal of Multilingual and multicultural development* 28.6, pp. 445–467.
- Graves-Brown, C. (2010). *Dancing for Hathor: Women in Ancient Egypt*. Continuum.
- Grenoble, L.A. and L.J. Whaley (2005). *Saving languages: An introduction to language revitalization*. Cambridge University Press.
- Griffith University (n.d.). *Semantic Molecules*. Retrieved: April 9, 2023. URL: %5Curl%7Bhttps://intranet.secure.griffith.edu.au/schools-departments/natural-semantic-metalanguage/what-is-nsm/semantic-molecules%C2%A0%7D.
- Group, Nature Publishing (n.d.). *Nature news - species*. Retrieved: April 8, 2023. URL: %5Curl%7Bhttps://www.nature.com/scitable/definition/species-312/%7D.
- Gumperz, J. and S. Levinson (1996). *Rethinking Linguistic Relativity*. Cambridge University Press.
- Gussenhoven, C. (1992). “Dutch”. In: *Journal of the International Phonetic Association* 22.1-2, pp. 45–47.
- Hanne et al., (2019). “The reniform body: An integrative lateral protocerebral neuropil complex of Eumalacos-

- traca identified in Stomatopoda and Brachyura”. In: *Journal of Comparative Neurology* 528.7, pp. 1079–1094.
- Harris, R. (1987). *The Language Machine*. Cornell University Press.
- Hegel, G.W. (1975 [1812]). *Wissenschaft der Logik*. F. Meiner.
- Heidegger, M. (2019 [1962]). *Being and Time*. Martino Fine Books.
- Heine, Bernd and D. Nurse (2000). *African Languages: An Introduction*. Cambridge University Press.
- Hiksey, D. (June 2015). *How Gay’ Came to Mean ‘Homosexual’*.
- History Computer (2022). *The history of Gottfried Leibniz*. Retrieved: April 8, 2023. URL: %5Curl%7Bhttps : / / history-computer . com / the-history-of-gottfried-leibniz / %C2%A0%7D.
- Hofstadter, D.R. (2021). *G”odel, Escher, Bach: an Eternal Golden Braid*. Basic Books Inc.
- House, J. (2003). “English as lingua franca: A threat to multilingualism?” In: *Journal of Sociolinguistics* 7.4, pp. 556–578.
- (2010). *Overt and Covert Translation*. John Benjamins Publishing Company, pp. 245–246.
- Husserl, E. (1993 [1900]). *Logische Untersuchungen*. Niemeyer.

BIBLIOGRAPHY

- IUPAC (2021). *Nomenclature - IUPAC: International Union of Pure and Applied Chemistry*. Retrieved: April 9, 2023. URL: %5Curl%7Bhttps://iupac.org/what-we-do/nomenclature/%C2%A0%7D.
- Jackson, F. (1982). “Epiphenomenal Qualia”. In: *The Philosophical Quarterly* 32.127, pp. 127–136.
- Jakobson, R. (2018). “From Linguistics and Poetics”. In: *The Norton Anthology of Theory and criticism*. 3rd ed. W.V. Norton & Company, pp. 1067–1074.
- Junior, V.L. (2018). *What is despeciation?* Retrieved: April 8, 2023. URL: %5Curl%7Bhttps://www.worldatlas.com/articles/what-is-despeciation.html%C2%A0%7D.
- Jurafsky, D. and J.M. Martin (2023). *Speech and Language Processing: 3rd edition*. Published online.
- Kant, I. (1998 [1781]). *Kritik der Reinen Vernunft*. Wissenschaftliche Buchgesellschaft Darmstadt.
- Ke, J., T. Gong, and W.S. Wang (2008). “Language change and social networks”. In: *Communications in Computational Physics* 3.4, pp. 935–949.
- Keane, E. (2004). “Tamil”. In: *Journal of the International Phonetic Association* 34.1, pp. 111–116.

- Kecskes, I. (2008). "Dueling context: A dynamic model of meaning". In: *Journal of Pragmatics* 40.3, pp. 385–406.
- Kerswill, P. (1996). "Children, adolescents, and language change". In: *Language Variation and Change* 8.2, pp. 177–202.
- Kluender, R. and M. Kutas (1993). "Subjacency as a processing phenomenon". In: *Language and cognitive processes* 8.4, pp. 573–633.
- Knapp, K. and C. Meierkord (2002). *Lingua Franca Communication*. Peter Lang EBooks.
- Knapp, S. and W.B. Michaels (1985). "Against Theory". In: *Against Theory: Literary Studies and the New Pragmatism*. University of Chicago Press.
- Kramsch, C. (2014). "Language and Culture". In: *Aila Review* 27, pp. 30–55.
- Kringelbach, M.L. and K.C. Berridge (2009). "Towards a functional neuroanatomy of pleasure and happiness". In: *Trends in Cognitive Sciences* 13.11, pp. 479–487.
- Kushniarevich et al., (2015). "Genetic heritage of the Balto-Slavic speaking populations: a synthesis of autosomal, mitochondrial and Y-chromosomal data". In: *PloS one* 10.9, e0135820.

BIBLIOGRAPHY

- Labov, W. (1990). "The intersection of sex and social class in the course of linguistic change". In: *Language variation and change* 2.2, pp. 205–254.
- (1994). *Principles of Linguistic Change, Internal Factors*. Wiley-Blackwell.
- Ladefoged, P. (1999). *American English: Handbook of the International Phonetic Association*. Cambridge University Press.
- Ladefoged, P. and I. Maddieson (1996). *The Sounds of the World's Languages*. Wiley-Blackwell.
- Lakoff, G. (1980). *Metaphors we live by*. University of Chicago Press.
- (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. University of Chicago Press.
- Lang, S. (2016). *Toki Pona: la langue du bien*. Tawhid, p. 6.
- Leavitt, J.H. (2010). *Linguistic Relativities: language diversity and modern thought*. Cambridge University Press.
- Lee, P. (1997). *The Whorf Theory Complex: A Critical Reconstruction*. John Benjamins Publishing Company.
- Lee, Y.S., D.A. Vakoch, and L.H. Wurm (1996). "Tone Perception in Cantonese and Mandarin: A cross-linguistic

- comparison". In: *Journal of Psycholinguistic Research* 25, pp. 83–122.
- Levelt, W. (2000). "Producing Spoken Language". In: *The Neurocognition of Language*. Ed. by C.N. Brown and P. Hagoort. Oxford University Press, pp. 83–122.
- Levinson, S.C. (1998). "Studying Spatial Conceptualization across Cultures: Anthropology and Cognitive Science". In: *Ethos* 26.1, pp. 7–24.
- Levisen, C. and S.E. Waters (2017). *Cultural Keywords in Discourse*. John Benjamins Publishing Company.
- Lewis, G.L. (2001). *Turkish Grammar*. Oxford University Press.
- Liberman, A.M. et al. (1967). "Perception of the speech code". In: *Psychological Review* 74.6, p. 431.
- Lidell, H.G. (1894). *A greek-english lexicon*. Harper.
- Liegeois-Chauvel, C. et al. (1999). "Specialization of left auditory cortex for speech perception in man depends on temporal coding". In: *Cerebral Cortex* 9.5, pp. 484–496.
- Lucy, J.A. (1997). "The Linguistics of Colour". In: *Colour categories in thought and language*. Ed. by C.L. Hardin and L. Maffi. Cambridge University Press, pp. 320–436.

BIBLIOGRAPHY

- (1992a). *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis*. Cambridge University Press.
- Lyons, J. (1968). *Introduction to Theoretical Linguistics*. Cambridge University Press.
- (1991). *Natural Language and Universal Grammar: Volume 1: Essays in Linguistic Theory*. Cambridge University Press.
- Maass, A. (1999). “Linguistic Intergroup Bias: Stereotype Perpetuation Through Language”. In: *Advances in Experimental Social Psychology*. Elsevier B.V., pp. 79–121.
- Macedo, D. and L.I. Bartolome (2014). “Multiculturalism permitted in English only”. In: *International Multilingual Research Journal* 8.1, pp. 24–37.
- MacEvoy, B. (2007). *Do Primary Colors Exist?* Retrieved: April 9, 2023. URL: %5Curl%7Bhttps://www.handprint.com/HP/WCL/color6.html%7D.
- Maddieson, I (1984). *Patterns of sounds*. Cambridge University Press.
- (2013b). *Structural Linguistics*.
- Mata, D. (n.d.). *Amplitude, wavelength, and frequency*. Retrieved: April 9, 2023. URL: %5Curl%7Bhttps://djmdme110.

- wordpress.com/2015/11/07/amplitude-wavelength-and-frequency/%C2%A0%7D.
- Matisoff, J.A. (1991). "Sino-Tibetan linguistics: present state and future prospects". In: *Annual review of anthropology* 20.1, pp. 469–504.
- Matthes, P.H. (1991). *Morphology*. Cambridge University Press.
- Mauranen, A. (2003). "The Corpus of English as Lingua Franca in Academic Settings". In: *TESOL Quarterly* 37.3, p. 513.
- Meecham, M. and J. Rees-Miller (2001). "Language in Social Contexts". In: *Contemporary Linguistics*. Ed. by M. Aronoff and J. Rees-Miller. 4th ed. Bedford/St. Martins.
- Mesgarani, N. et al. (2008). "Phoneme Representation and classification in primary auditory cortex". In: *The Journal of the Acoustical Society of America* 123.2, pp. 899–909.
- Migge, B. and I. Leglise (2007). "Language and Colonialism". In: *Handbook of Language and Communication: Diversity and Change*. Ed. by M. Hellinger and A. Pauwels. De Gruyter Mouton.

BIBLIOGRAPHY

- Milward, C. and M. Hayes (2011). *A Biography of the English Language*. Cengage Learning.
- Moltmann, Friederike (Mar. 2017). *Natural Language Ontology*. DOI: 10.1093/acrefore/9780199384655.013.330. URL: <https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-330>.
- Moran, S. and D. McCloy (2019). *PHOIBLE 2.0*. Max Planck Institute for the Science of Human History.
- Nevalainen, T. and E.C. Traugott (2012). *The Oxford Handbook of the History of English*. Oxford University Press.
- Ney, A. (2014). *Metaphysics: An Introduction*. Taylor & Francis.
- Nida-Rümelin, Martine and Donnchadh O Conaill (2021). “Qualia: The Knowledge Argument”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2021. Metaphysics Research Lab, Stanford University.
- Norman, J.M. (n.d.). *John Wilkins creates a universal language based on a classification scheme or ontology, and a universal system of measurement*. Retrieved: April 8, 2023. URL: <https://historyofinformation.com/detail.php?entryid=1874%C2%A0%7D>.

- Norton, B. (1997). "Language, Identity and the Ownership of English". In: *TESOL Quarterly* 31.3, p. 409.
- NSMLab (2021). *The Molecules Project: The Building Blocks of Meaning*. Retrieved: March 30, 2023. URL: %5Curl%7Bhttps://www.youtube.com/watch?v=hODJ7mr1rTQ&ab_channel=NSMLab%7D.
- O'Grady, W.D. et al. (2001). *Contemporary Linguistics: An Introduction*. Palgrave MacMillan.
- Ottenheimer, H. (2009). *The anthropology of language: an introduction to linguistic anthropology*. 2nd ed. Wadsworth, pp. 33–34.
- Otwinowska, A. and J.M. Szewczyk (2019). "The more similar the better? Factors in learning cognates, false cognates and non-cognate words". In: *International Journal of Bilingual Education and Bilingualism* 22.8, pp. 974–991.
- Pavlenko, A. (2006). *Bilingual Minds: Emotional Experience, Expression, and Representation*. Multilingual Matters.
- Pawley, A. (1993). "A language which defies description by ordinary means". In: *The Role of Theory in Language Description*. De Gruyter Mouton, pp. 87–130.
- Perry, J.A. (2005). *A Tajik-Persian reference grammar*. Brill.

BIBLIOGRAPHY

- Pinker, S. (1994). *The Language Instinct: How the mind creates language*. Penguin UK.
- Piron, C. (1989). "Who are the speakers of Esperanto". In: *Interlinguistics: Aspects of the science of planned languages*, pp. 157–172.
- Plutchik, R. (1980). "Emotion: Theory, research, and experience". In: *Theories of emotion*. Vol. 1. Academic Press New York.
- Plutchik, R. and H. Kellerman (1980). *Theories of Emotion*. Academic Press New York.
- Pölzl, U. and B. Seidlhofer (2006). "In and on their own terms: the "habitat factor" in English as a lingua franca interactions". In: *International Journal of the Sociology of Language* 2006.177, pp. 151–176.
- Porphyry translated by Owen, O.F. (1853 [300]). *Isagoge*. Retrieved: April 9, 2023. URL: https://www.tertullian.org/fathers/porphyry_isagogue_02_translation.htm.
- Posner, R. and M. Sala (n.d.). *Romance Languages*. accessed may 2023.
- Priest, G. (2011). *An Introduction to Non-Classical Logic: From if to is*. Cambridge University Press.

Proske, U. and S.C. Gandevia (2012). “The Proprioceptive Senses: Their Roles in Signaling Body Shape, Body Position and Movement, and Muscle Force”. In: *Physiological Reviews* 92.4, pp. 1651–1697.

Pullum, G. (1991). *The Great Eskimo Vocabulary Hoax and other Irreverent Essays on the Study of Language*. Chicago University Press.

Reina, O. (2022). *The Sights and Sounds of Life: What are Sensory Neurons?* Retrieved April 9, 2023. URL: %5Curl%7Bhttps://www.tempobioscience.com/the-sights-and-sounds-of-life-what-are-sensory-neurons/%7D.

Rhiemeyer, J. (n.d.[a]). *Ars Signorum*. Retrieved: April 8, 2023. URL: %5Curl%7Bhttps://www.frathwiki.com/Ars_signorum%C2%A0%7D.

— (n.d.[b]). *Oligosynthetic language*. Retrieved: April 8, 2023. URL: %5Curl%7Bhttps://www.frathwiki.com/Oligosyntheticlanguage%C2%A0%7D.

Roach, P.J. (2004). “British English: Received Pronunciation”. In: *Journal of the International Phonetic Association* 34.2, pp. 239–245.

Robinson, A. (2009). *Lost Languages: the enigma of the worlds undeciphered scripts*. Thames & Hudson, p. 61.

BIBLIOGRAPHY

- Robson, A. (2002). "Critical/Sensitive Periods". In: *Child Development*. MacMillan Reference USA, pp. 101–103.
- Roget, P.M. (2019 [1805]). *Thesaurus*. Retrieved: April 9, 2023. URL: %5Curl%7Bhttps://web.archive.org/web/20130502153549/http://www.rain.org/~karpeles/rogetfrm.html%7D.
- Rollinger, Robin and Carlo Ierna (2022). "Christian von Ehrenfels". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2022. Metaphysics Research Lab, Stanford University.
- Rolstad, K. (2001). "Language Death in Central Mexico: The Decline of Nahuatl and the New Bilingual Maintenance Programs". In: *The Bilingual Review* 26.1, p. 3.
- Sapir, E. (1929). "The status of linguistics as a science". In: *Language* 5.4, pp. 207–214.
- Saville-Troike, M. and Karen Barto (2017). *Introducing second language acquisition*. Cambridge University Press.
- Schiffman, H. and S. Arokianathan (1986). "Diglossic Variation in Tamil film and fiction". In: *South Asian languages: Structure, convergence, and diglossia*. Vol. 10, pp. 371–381.
- Sherwood, B.A. (1982). "Variation in Esperanto". In: *Studies in the Linguistic Sciences* 12.1, pp. 183–196.

- Silverman, D. (2012). *Neutralization*. Cambridge University Press.
- Simplifying Theory (n.d.). *Timbre*. Retrieved: April 9, 2023.
URL: %5Curl%7Bhttps://www.simplifyingtheory.com/timbre/%C2%A0%7D.
- Soderberg, C. and K.S. Olson (2008). "Indonesian". In: *Journal of the International Phonetic Association* 38.2, pp. 209–213.
- Sparks, R. et al. (2009). "Long-term crosslinguistic transfer of skills from L1 to L2". In: *Language Learning* 59.1, pp. 203–243.
- Suarez, J.A. (1983). *The Mesoamerican Indian languages*. Cambridge University Press.
- Swadesh, M. (1952). *Lexico-Statistical Dating of Prehistoric Ethnic Contacts*.
- Takagi, N. and V. Mann (1995). "The Limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English". In: *Applied Psycholinguistics* 16.4, pp. 380–406.
- Trabant, J. (2000). "How relativistic are Humboldts "Weltansichten"?" In: *Explorations in Linguistic Relativity*. Ed. by M. Pütz and M. Verspoor. John Benjamins Publishing Company, pp. 25–44.

BIBLIOGRAPHY

- Tryon, D. (1995). "Proto-Austronesian and the major Austronesian subgroups". In: *The Austronesians*, pp. 19–42.
- Tynyanov, Y. (1979). "About Khlebnikov". In: *The Futurists, The Formalists, and the Marxist Critique*. Ed. by C. Pike. Ink Links, pp. 144–158.
- UNDL Foundation (n.d.[a]). *Relations*. Retrieved: April 8, 2023. URL: %5Curl%7Bhttp://www.undl.org/unlsys/unl/UNL2005/Relation.htm%7D.
- (n.d.[b]). *What is UNL*. Retrieved: April 8, 2023. URL: %5Curl%7Bhttp://www.undl.org/index.php?option=com_content&view=article&id=46&Itemid=58&lang=en%2%A0%7D.
- Vandekerckhove, R. (2005). "Belgian versus Netherlandic Dutch: New patterns of divergence? On pronouns of address and diminutives". In: *Multilingua* 24.4, pp. 379–397.
- Verhoeven, J. (2005). "Belgian Standard Dutch". In: *Journal of the International Phonetic Association* 35.2, pp. 243–247.
- Versteegh, K. (2014). *Arabic Language*. Edinburgh University Press.

- Von Humboldt, W. (1988 [1836]). *On language: the diversity of human language structure and its influence on the mental development of mankind* (trans. P. Heath). Cambridge University Press.
- Watson, O. (n.d.). *Ygyde Language - Comparison of philosophical languages*. Retrieved: April 8, 2023. URL: %5Curl%7Bhttp://www.orenwatson.be/saved/ygyde.htm%C2%A0%7D.
- White, J.M. (2002). *Everyday Life in Ancient Egypt*. Courier Dover, p. 175.
- Whitehead, A.N. (2010 [1929]). *Process and Reality*. Free Press Riverside.
- Wichmann et al., (2009). *Where prosody meets pragmatics*. Brill, pp. 1–20.
- Wikimedia Foundation (2022). *Chromatic Circle*. Retrieved: April 9, 2023. URL: %5Curl%7Bhttps://en.wikipedia.org/wiki/Chromatic_circle%C2%A0%7D.
- Wilkins, J. (1968 [1668]). *Essay Towards a Real Character and a Philosophical Language*. Scholar Press.
- Wittgenstein, L. (2010 [1953]). *Philosophical Investigations*. Wiley-Blackwell.
- Yallop, C. and J. Fletcher (2007). *An Introduction to Phonetics and Phonology*. Wiley-Blackwell.

BIBLIOGRAPHY

Yip, M. (2002). *Tone*. Cambridge University Press.

Young, T. (2009 [1823]). *An Account of Some Recent Discoveries in Hieroglyphical Literature, and Egyptian Antiquities*. John Murray.