Completely Positive Maps in Natural Language Processing

Andreas Mair

Supervisor

Univ.-Prof. Dr. Tim Netzer

29th August 2021

Contents

Chapter 1

Introduction

In his seminal paper *Computing Machinery and Intelligence* Alan Turing proposed the famous *Turing test*, also called *the imitation game*, a test that supposedly assesses a machine's level of intelligence. Basically, it is an A/B test in which an interrogator communicates with both a human being and a human-imitating machine not knowing which one is which. According to Alan Turing a machine displays human levels of intelligence if the interrogator is not able to discern between the two. Even though, this type of test is considered flawed in many ways one can nevertheless state how important the role of language is for perceived intelligence. Thus, linguists and computer scientists alike study the interface of human language and machines. This area of research is commonly referred to as *natural language processing* and involves the parsing, modelling and processing of human language in a machine readable format.

A very important part of natural language processing is the modelling of semantics, i.e. the meaning that is baked into the words and the structure of sentences. Many modern language models like the all-famous GPT-3 rely on distributional semantics which is a modelling approach that uses vast amounts of text to deduce meaning from co-occurrence. This way of semantic analysis also lends itself to deep learning algorithms and may explain its relevance in many present-day implementations.

This bachelor thesis is all about the mathematical framework for a certain problem in distributional semantics, namely, the composition of semantic entities like words and phrases in a matrix-based model of meaning. Imagine every word has its meaning represented by a matrix. What is the natural way to combine these? Mappings, one might intuitively say. At this point we can bring mathematics to the table. A very natural way of looking at matrices is in terms of C^*-algebras, where mappings satisfy several properties we will come back to at a later time.

For now, let us continue asking questions. Given some mapping to combine words in to sentences with what are the semantic implications? We clearly want to lose as little information as possible but there is a wide range of linguistic phenomena that one could focus on like synonymy, antonymy, meronymy, hyponymy and many more. The one this thesis focuses on is hyponymy because mappings between matrix algebras can be completely positive which ensures that the hyponym-hypernym relationship between two semantic units is preserved under composition.

In the beginning we will kick things off with a 101 in linguistics. Not many readers (apart from linguists) are familiar with terms like *semantic* or *hyponymy*, therefore a short introduction will cover the most basic terminology and give some examples. Then there will be a chapter on the topic of C^*-algebras, operator spaces and completely positive maps. This will culminate in a proof of *Choi's theorem on completely positive maps* which forms the basis of a class of compositional maps presented in [1]. Therefore, this thesis finishes by deriving and analysing these maps and bringing together linguistics and mathematics.

Chapter 2

Linguistics – a very short introduction

According to the Oxford dictionary linguistics is "the scientific study of language and its structure, including the study of grammar, syntax, and phonetics". The field of study we know today encompasses a vast amount of subdisciplines like sociolinguistics, the study of society's influence on language and vice versa, psycholinguistics, which is concerned with the brain's faculty to produce and understand language or dialectology, the study of dialects, only to name a few. An area of utmost topicality that developed in the 1950s is computational linguistics.

2.1 Computational linguistics and semantics

The *Association for Computational Linguistics* describes computational linguistics as the subfield of linguistics that is concerned with language from the point of view of a computer. They give a very nice summary of a computational linguist's area of interest.

> "Computational linguists are interested in providing computational models of various kinds of linguistic phenomena. These models may be 'knowledge-based' ('handcrafted') or 'data-driven' ('statistical' or 'empirical'). Work in computational linguistics is in some cases motivated from a scientific perspective in that one is trying to provide a computational explanation for a particular linguistic or psycholinguistic phenomenon; and in other cases the motivation may be more purely technological in that one wants to provide a working component of a speech or natural language system. Indeed, the work of computational linguists is incorporated into many working systems today, including speech recognition systems, text-to-speech synthesizers, automated voice response systems, web search engines, text editors, language instruction materials, to name just a few." [1]

Computational linguists are thus concerned with developing theories and models that allow for a computer-based representation of natural, i.e. human, language.

A popular comic series by Bill Watterson involving Calvin, an imaginative six-year-old boy, and Hobbes, an energetic and sardonic tiger, is known for their funny dialogues. One of them

[1] `https://www.aclweb.org/portal/`

goes:

Calvin:	I like to verb words.
Hobbes:	What?
Calvin:	I take nouns and adjectives and use them as verbs.
	Remember when "access" was a thing? Now it's something you do. It got verbed.
Calvin:	Verbing weirds language.
Hobbes:	Maybe we can eventually make language a complete impediment to understanding.

This goes to show that language is messy and a model that properly understands "weird" as an adjective might have a hard time picking up on the *conversion*, which is the technical term for creating a word from an existing one and is exactly what Calvin does when he "verbs a word". The computer modelling of the meaning of words is commonly referred to as *computational semantics*. We now want to take a closer look at a certain type of model, one that is based on statistical properties of text corpora.

2.2 Distributional semantics

As early as the 1950s prominent linguists were turning their attention towards distributional models that is analysing speech based on descriptive qualities. In 1954 Zellig S. Harris, an American linguist and mathematical syntactician, published *Distributional Structure* [2] a paper arguing that language can be completely understood in terms of distributional facts. He writes:

> "At various times it has been thought that one could only state the normative rules of grammar (e.g. because colloquial departures from these were irregular) or the rules for a standard dialect but not for 'sub-standard' speech or slang; or that distributional statements had to be amplified by historical derivation (e.g. because the earlier form of the language was somehow more regular). However, in all dialects studied it has been possible to find elements having regularities of occurrence; and while historical derivation can be studied both independently and in relation to the distribution of elements, it is always also possible to state the relative occurrence of elements without reference to their history (i.e. 'descriptively')."

Then he goes on to explain how regularities in text might translate to meaning representations. For example, not every adjective might pair with a certain noun, so thereby one can infer a difference in meaning. Having two words A and B that are more different in meaning than A and another word C often entails that A and B are more different in the way that they are distributed in some text than A and C. Put differently, Harris points out a correlation between difference of meaning and difference of distribution. As an example, he considers the words *oculist* and *eye-doctor*. Only in very few cases is it not possible to switch out these two words in a given sentence because they mean the same for the most part (they are what linguists call synonyms). The possibility of *oculist* and *eye-doctor* occurring in the same "environements" is much more likely than that of *oculist* and *lawyer*. This clearly illustrates how environments of meaning and environments of distribution might correlate. In his publication Zellig Harris also proposes a conceptual method of analysis breaking speech down into irreducible elements first and analysing them based on similarity, dependence and substitutability. In essence, he recommends analysing co-occurrence and this is why we will now take a closer look at a vector-based model of meaning that tries to do just that.

2.3 Vector-based models of meaning

Many models in distributional semantics use high dimensional vector spaces where distributional and thus semantic similarities are expressed in terms of similarities between vectors. A well-known vector-based model of meaning is called `word2vec`. The program classifies as a neural network that learns words from a large corpus of text. It autonomously chooses vector representations such that the angle between two vectors indicates the level of semantic similarity between the words represented by them. This is called the cosine similarity of the vectors and is defined by

$$\text{similarity}(v, w) := \frac{v \cdot w}{\|v\|\|w\|} = \cos(\theta)$$

where v, w are two vectors in a high dimensional vector space and θ is the angle between them. To conclude this section on vector-based models let us take a look at a toy example from [3] that will illustrate in very broad terms what a vector representation looks like and to which we will come back to in chapter 4.

Imagine the noun *pet* and suppose that we have three types of pets: a *pug*, a *goldfish* and a *tabby cat*. These nouns might co-occur in some text with a bunch of adjectives, say *furry, domestic, working* and *aquatic*. For example, *furry* could be an adjective that is either directly before or not far from the noun *pug* in some naturally produced sentence. As a simple model lets imagine a count of all co-occurrences, something like in table 2.1 where we take the columns of this matrix

	pug	goldfish	tabby cat
furry	3	0	5
domestic	4	5	5
working	0	0	0
aquatic	0	6	0

Table 2.1: A simple toy model

as the vector representations of the three nouns *pug, goldfish* and *tabby cat*.

Later on, we will return to this example as it is simple yet very illustrative. For now, we know all the historical and linguistic basics we need and we shall continue with the mathematical centrepiece.

Chapter 3

Operator theory and completely positive maps

This chapter builds in part upon Vern Paulsen's book *Completely bounded maps and operator algebras* [5] which relies on *Banach algebra techniques in operator theory* [6] for several definitions and focuses on the groundwork relevant to certain compositional models of meaning. The theory of operator algebras and their mappings is an interesting area of research in its own right and this chapter can certainly be read independently of computational semantics.

3.1 Introduction

Let us start by revisiting and defining basic concepts such as algebras and Banach spaces.

Definition 3.1 (Algebra over a field)**.** Let K be a field, A a ring and $Z(A)$ its centre. Then A together with a ring homomorphism $K \to Z(A)$ constitutes an algebra over the field K. △

Remark/Example 3.2. (i) An algebra can thus be thought of as a kind of vector space over the field K where in addition to the scalar multiplication there exists a multiplication between the elements of A themselves. A typical example of an algebra over a field is the matrix algebra. Square matrices of size $n \times n$ with entries in K form a ring and the scalar multiplication is defined by scaling every entry. In terms of the ring homomorphism

$$\begin{aligned} \mu : K &\to Z\left(\mathrm{Mat}_n(K)\right) \\ k &\mapsto \mathrm{diag}(\underbrace{k, \ldots, k}_{n \text{ times}}) \end{aligned}$$

one can think of the scalar multiplication as $k \cdot A := \mu(k)\, A = A\, \mu(k)$ since $\mu(k)$ being diagonal obviously commutes with every element in $\mathrm{Mat}_n(K)$.
(ii) Note especially that A in the definition above does not have to be a commutative ring. Still, the ring elements that represent scalars of K need to commute with every other element in A.
(iii) Furthermore, the ring homomorphism is necessarily injective as its domain is a field and thereby its kernel always $\{0\}$.
(iv) We will abbreviate the name *algebra over a field* by *algebra* or K*-algebra* if we want to explicitly state the field over which the algebra is constructed. △

Definition 3.3 (Algebra homomorphism)**.** Given two K-algebras A and B, we call $\varphi : A \to B$ a (unital) algebra homomorphism iff φ is a ring homomorphism for the corresponding rings, i.e. for all $a, b \in A$

$$\begin{aligned}\varphi(a+b) &= \varphi(a) + \varphi(b)\\ \varphi(a \cdot b) &= \varphi(a) \cdot \varphi(b)\\ \varphi(1_A) &= 1_B\end{aligned}$$

and furthermore for all $a \in A$ and $k \in K$

$$\varphi(k\,a) = k\,\varphi(a)$$

where $k\,a$ and $k\,\varphi(a)$ are scalar multiplications in the respective algebras as we have seen earlier. △

Remark/Example 3.4. Observe that *unital* merely states that the ring units are mapped to each other. Contrary to the zero elements which are linked by the first property of ring homomorphisms

$$\varphi(0_A) = \varphi(0_A + 0_A) = \varphi(0_A) + \varphi(0_A) \implies \varphi(0_A) = 0_B$$

we need to axiomatically state that $\varphi(1) = 1$. This property cannot be deduced from the other axioms. △

Let us take the concept of algebras a little further. For the application in computational semantics we need to introduce the notion of C^*-algebras which in someway resemble and build upon the K-algebras over a field we have seen so far.

Definition 3.5 (Banach space)**.** A Banach space is a complex linear space $\mathcal{H}$ with a norm $\|\cdot\|$ satisfying

(i) $\|f\| = 0$ if and only if $f = 0$,

(ii) $\|\lambda f\| = |\lambda|\|f\|$ for λ in $\mathbb{C}$ and $f \in \mathcal{H}$ and

(iii) $\|f + g\| \leq \|f\| + \|g\|$ for f and $g \in \mathcal{H}$,

such that $\mathcal{H}$ is complete in the metric given by this norm. △

Definition 3.6 (Banach algebra)**.** A Banach algebra $\mathcal{B}$ is an algebra over the field $\mathbb{C}$ with identity 1 which has a norm making it into a Banach space, satisfying $\|1\| = 1$ and the inequality $\|fg\| \leq \|f\|\|g\|$ for f and g in $\mathcal{B}$. △

Remark/Example 3.7. Even though this definition might seem abstract there are several examples of Banach algebras that might feel familiar.
(i) The field of complex numbers $\mathbb{C}$ equipped with the usual addition and multiplication and the absolute value as the norm is a Banach algebra.
(ii) If K is a compact Hausdorff space, then $C(K)$ is a Banach algebra where the addition and multiplication are defined pointwise and the norm is given by $\|f\|_\infty := \sup_{t \in K} |f(t)|$.

(iii) An interesting example, especially for our purposes in computational semantics, is the set of $n \times n$ real or complex matrices equipped with a submultiplicative matrix norm. $\triangle$

Definition 3.8 (Involution)**.** If $\mathcal{B}$ is a Banach algebra, then an involution on $\mathcal{B}$ is a mapping $a \mapsto a^*$ which satisfies:

(i) $a^{**} = a$ for $a \in \mathcal{B}$,

(ii) $(\lambda a + \mu b)^* = \bar{\lambda} a^* + \bar{\mu} b^*$ for $a, b \in \mathcal{B}$ and $\lambda, \mu \in \mathbb{C}$,

(iii) $(ab)^* = b^* a^*$ for $a, b \in \mathcal{B}$; $\triangle$

Definition 3.9 (C^*-algebra)**.** Let $\mathcal{B}$ be a Banach algebra with an involution $a \mapsto a^*$. If

$$\|a^* a\| = \|a\|^2$$

for all $a \in \mathcal{B}$, then $\mathcal{B}$ is called a C^*-algebra. $\triangle$

Remark/Example 3.10. (i) First, observe that for any element a of a C^*-algebra the inequality

$$\|a\|^2 = \|a^* a\| \leq \|a^*\| \|a\|$$

implies that $\|a\| \leq \|a^*\|$ and hence $\|a\| = \|a^*\|$ because the same argument can be made for a^* since $a^{**} = a$. Thus, the involution on a C^*-algebra is an isometry.
(ii) There is a canonical way to define homomorphisms between C^*-algebras. In essence, we want a mapping to be compatible with the involution, i.e. for two C^*-algebras $\mathcal{A}, \mathcal{B}$ an algebra homomorphism $\varphi : \mathcal{A} \to \mathcal{B}$ is called a $*$**-homomorphism** if

$$\varphi(a^*) = \varphi(a)^*$$

for each $a \in \mathcal{A}$. A bijective $*$-homomorphism is called a $*$**-isomorphism**.
(iii) Importantly, the Banach algebra $\mathrm{Mat}_n(\mathbb{C})$ equipped with the operator norm $\|\cdot\|$ induced by the euclidean norm on $\mathbb{C}^n$ turns into a C^*-algebra if we let the involution $A \mapsto A^*$ be defined by the conjugate transpose. $\triangle$

There is one final definition left before we can dive into completely positive maps. Bounded operators are a matter of peculiar interest since we want to look at elements of C^*-algebras as bounded operators on Hilbert spaces.

Definition 3.11 (Bounded operator)**.** Let $\mathcal{H}$ and $\mathcal{K}$ be two Hilbert spaces and $T : \mathcal{H} \to \mathcal{K}$ an operator. Then T is called bounded if there exists some $C \geq 0$ such that for all $x \in \mathcal{H}$

$$\|Tx\|_{\mathcal{K}} \leq C\|x\|_{\mathcal{H}}.$$

We denote the set of all bounded linear operators on a Hilbert space $\mathcal{H}$ by $B(\mathcal{H})$. $\triangle$

3.2 Completely positive maps

Having now laid all the groundwork we turn our attention to completely positive maps between C^*-algebras. There exists a theorem due to Isreal Gelfand and Mark Naimark (1943) which states that every C^*-algebra is isometrically $*$-isomorphic to some subalgebra of bounded operators on a Hilbert space.

From now on we will often focus solely on the special case where the C^*-algebra is the matrix algebra $\mathrm{Mat}_n(\mathbb{C})$ because the connection to operators is readily understandable. Every finite dimensional Hilbert space has a basis and the inner product induces a norm, thus we can look at these maps through linear algebraic spectacles associating elements of $\mathrm{Mat}_n(\mathbb{C})$ with maps from a n-dimensional Hilbert space onto itself.

Definition 3.12 (Positivity)**.** Let $\mathcal{A}$ be a C^*-algebra. An element A is called positive if it satisfies

(i) $A = A^*$ and

(ii) $\{\lambda \in \mathbb{C} : A - \lambda \text{ is not invertible in } \mathcal{A}\} \subseteq [0, \infty)$.

The set in (ii) is called the spectrum of A. $\triangle$

Remark/Example 3.13. (i) If $\mathcal{A} = \mathrm{Mat}_n(\mathbb{C})$, it is clear that the positive elements consist of all Hermitian matrices with nonnegative eigenvalues, i.e. positive semidefinite matrices, since the spectral theorem from linear algebra tells us that eigenvalues of Hermitian matrices are always real.
(ii) It can also be shown that the positivity of an element $A \in \mathcal{A}$ can be characterised by

$$\langle Ax, x \rangle \geq 0 \quad \text{for all} \quad x \in \mathcal{H}$$

where A generally needs to be interpreted in terms of operators via the Gelfand-Naimark theorem and $\mathcal{H}$ is the corresponding Hilbert space. Yet in the case of $\mathcal{A} = \mathrm{Mat}_n(\mathbb{C})$ this is an alternative characterisation of positive semidefinite matrices most readers will be familiar with.
(iii) Furthermore, an element A is positive iff $A = S^*S$ for some $S \in \mathcal{A}$. If $A = S^*S$, then

$$\langle S^*Sx, x \rangle = \|Sx\|^2 \geq 0.$$

Conversely, it can be shown that for every positive operator A there exists a unique, positive operator Q with $Q^2 = A$. Therefore, one can define $S := \sqrt{A} = Q$. Again, for the finite dimensional case this is clear by the spectral theorem for Hermitian matrices. $\triangle$

Definition 3.14 (Positive and completely positive maps)**.** (i) Let $\mathcal{A}$ and $\mathcal{B}$ be C^*-algebras. A linear map $\varphi : \mathcal{A} \to \mathcal{B}$ is said to be **positive** if for all $A \in \mathcal{A}$

$$A \geq 0 \implies \varphi(A) \geq 0.$$

(ii) Again let $\mathcal{A}$ and $\mathcal{B}$ be C^*-algebras and $\mathrm{Mat}_n(\mathcal{A})$ and $\mathrm{Mat}_n(\mathcal{B})$ the sets of $n \times n$ matrices with components in $\mathcal{A}$ and $\mathcal{B}$ respectively. Then, $\mathrm{Mat}_n(\mathcal{A})$ and $\mathrm{Mat}_n(\mathcal{B})$ are also C^*-algebras defining the involution by the conjugate transpose and the norm via the Gelfand-Naimark theorem as follows. Since $\mathcal{A}$ is isometrically $*$-isomorphic to some subalgebra of bounded operators on a Hilbert space, we can embed $\mathcal{A} \subseteq B(\mathcal{H})$. Now $\mathrm{Mat}_n(\mathcal{A}) \subseteq \mathrm{Mat}_n(B(\mathcal{H})) \cong B(\mathcal{H}^n)$ and thus we can define the norm of any element in $\mathrm{Mat}_n(\mathcal{A})$ by the norm of the corresponding element in $B(\mathcal{H}^n)$. The same is true for $\mathrm{Mat}_n(\mathcal{B})$. Then, every linear map $\varphi : \mathcal{A} \to \mathcal{B}$ induces another map $\varphi_n : \mathrm{Mat}_n(\mathcal{A}) \to \mathrm{Mat}_n(\mathcal{B})$ via

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \mapsto \begin{pmatrix} \varphi(a_{11}) & \cdots & \varphi(a_{1n}) \\ \vdots & \ddots & \vdots \\ \varphi(a_{n1}) & \cdots & \varphi(a_{nn}) \end{pmatrix}.$$

The map φ is called **n-positive** if φ_n is a positive map.
(iii) For two C^*-algebras $\mathcal{A}$ and $\mathcal{B}$ a map $\varphi : \mathcal{A} \to \mathcal{B}$ is called completely positive if φ is n-positive for all $n \in \mathbb{N}$. $\triangle$

Remark/Example 3.15. (i) It is easy to see that every $*$-homomorphism is positive since for any positive element A

$$\varphi(A) = \varphi(S^*S) = \varphi(S^*)\varphi(S) = \varphi(S)^*\varphi(S) = \|\varphi(S)\|^2 \geq 0$$

for some $S \in \mathcal{A}$.
(ii) $*$-homomorphisms are even completely positive. For such a map $\varphi : \mathcal{A} \to \mathcal{B}$ we need to take a look at $\varphi_n : \mathrm{Mat}_n(\mathcal{A}) \to \mathrm{Mat}_n(\mathcal{B})$ for any $n \in \mathbb{N}$. Let $A, B \in \mathrm{Mat}_n(\mathcal{A})$ and $\lambda \in \mathcal{A}$. Then

(a) φ_n is an algebraic homomorphism because

$$(\varphi_n(A+B))_{ij} = \varphi(A_{ij} + B_{ij}) = \varphi(A_{ij}) + \varphi(B_{ij}) = (\varphi_n(A))_{ij} + (\varphi_n(B))_{ij},$$

$$(\varphi_n(AB))_{ij} = \varphi\left(\sum_{k=1}^{n} A_{ik}B_{kj}\right) = \sum_{k=1}^{n} \varphi(A_{ik})\varphi(B_{kj}) = \sum_{k=1}^{n} (\varphi_n(A))_{ik}(\varphi_n(B))_{kj}$$

and

$$(\varphi_n(\lambda A))_{ij} = \varphi(\lambda A_{ij}) = \lambda\varphi(A_{ij}) = \lambda(\varphi_n(A))_{ij}.$$

(b) In addition, φ_n is compatible with the involution. We get that

$$(\varphi_n(A^*))_{ij} = \varphi(A^*_{ji}) = \varphi(A_{ji})^* = (\varphi_n(A)^*)_{ij}$$

Therefore, φ_n is a $*$-homomorphism itself and thus a positive map for every $n \in \mathbb{N}$ meaning φ is completely positive.
(iii) Not every positive map is completely positive though. Let $\varphi : \mathrm{Mat}_n(\mathbb{C}) \to \mathrm{Mat}_n(\mathbb{C})$ be defined by $\varphi(A) = A^T$. Then it is true that

$$\varphi(A^*A) = (A^*A)^T = A^T(A^*)^T = A^T(A^T)^* = \|A^T\| \geq 0$$

since the involution on $\mathrm{Mat}_n(\mathbb{C})$ is the conjugate transpose. Thus, φ is positive but not completely positive as the following example will show. Consider the case $n = 2$ and

$$\varphi_2 : \mathrm{Mat}_2(\mathrm{Mat}_2(\mathbb{C})) \cong \mathrm{Mat}_4(\mathbb{C}) \to \mathrm{Mat}_2(\mathrm{Mat}_2(\mathbb{C})) \cong \mathrm{Mat}_4(\mathbb{C}).$$

Let

$$A := \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

which is positive since

$$\langle Ax, x\rangle = (x_1 + x_4)x_1 + (x_1 + x_4)x_4 = (x_1 + x_4)^2 \geq 0$$

On the other hand

$$\varphi_2(A) = \begin{pmatrix} \varphi\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & \varphi\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \\ \varphi\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} & \varphi\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

which is not positive as

$$\langle \varphi_2(A)(0, 1, -1, 0)^T, (0, 1, -1, 0)^T\rangle = -2$$

shows.
(iv) Another example of a completely positive map we will get to see later in more detail is the following. Let $\mathcal{A}$ be a C^*-algebras and $B \in \mathcal{A}$ arbitrary. Define $\varphi : \mathcal{A} \to \mathcal{A}$ by $\varphi(A) = B^*AB$. Then

$$\varphi(A^*A) = B^*A^*AB = (AB)^*AB = \|AB\|^2 \geq 0.$$

Thus, φ is positive. For $\varphi_n : \mathrm{Mat}_n(\mathcal{A}) \to \mathrm{Mat}_n(\mathcal{A})$

$$(\varphi_n(A))_{ij} = \varphi(A_{ij}) = B^*A_{ij}B$$

and therefore

$$\varphi_n(A) = \mathrm{diag}_n(B)^* A \, \mathrm{diag}_n(B)$$

where $\mathrm{diag}_n(B)$ is the diagonal $n \times n$ matrix with B in each diagonal entry and zero elsewhere. If $A \in \mathrm{Mat}_n(\mathcal{A})$ is positive, then $A = N^*N$ for some $N \in \mathrm{Mat}_n(\mathcal{A})$ and thereby

$$\mathrm{diag}_n(B)^* A \, \mathrm{diag}_n(B) = (N \, \mathrm{diag}_n(B))^* \, (N \, \mathrm{diag}_n(B)) \geq 0.$$

This shows the complete positivity of the so called conjugation. △

3.3 Choi's theorem on completely positive maps

In 1975 Man-Duen Choi published a theorem characterising completely positive maps between complex matrix algebras [7].

Theorem 3.16 (Choi's theorem, 1975)**.** *Let* $\Phi : \mathrm{Mat}_n(\mathbb{C}) \to \mathrm{Mat}_m(\mathbb{C})$ *be a linear map. Then* Φ *is completely positive iff* Φ *is of the form* $\Phi(A) = \sum_{i=1}^{nm} V_i^* A V_i$ *for all* $A \in \mathrm{Mat}_n(\mathbb{C})$ *where* V_i *are* $n \times m$ *matrices.*

Proof. As we have seen in section 3.2 every map of the form $\sum_{i=1}^{nm} V_i^* A V_i$ is completely positive, thus we only need to prove the converse. Each $1 \times nm$ matrix v can be regarded as a $1 \times n$ block matrix $(x_1, ..., x_n)$ with $1 \times m$ matrices as entries x_j. Therefore, we associate with it the $n \times m$ matrix V which has x_j as the j-th row. According to this association

$$(V^* E_{jk} V)_{1 \leq j,k \leq n} = (x_j^* x_k)_{1 \leq j,k \leq n} = v^* v$$

since E_{jk} is 1 at the component (j, k) and 0 elsewhere. Suppose $\Phi : \mathrm{Mat}_n(\mathbb{C}) \to \mathrm{Mat}_m(\mathbb{C})$ is completely positive. The matrix $(E_{jk})_{1 \leq j,k \leq n}$ is positive by the association above for $V = I_n$, so $(\Phi(E_{jk}))_{1 \leq j,k \leq n} \in \mathrm{Mat}_n(\mathrm{Mat}_m(\mathbb{C}))$ must be positive by the complete positivity of Φ. By decomposing into eigenvectors we get

$$(\Phi(E_{jk}))_{jk} = \sum_{i=1}^{nm} v_i^* v_i$$

and because of the calculation above it holds true that

$$(\Phi(E_{jk}))_{jk} = \sum_{i=1}^{nm} (V_i^* E_{jk} V_i)_{jk}.$$

Extending the result by linearity we get that

$$\Phi(A) = \sum_{i=1}^{nm} V_i^* A V_i$$

for all A. □

The preceding proof came down to realising that $(E_{jk})_{1 \leq j,k \leq n}$ is a positive element thus $(\Phi(E_{jk}))_{1 \leq j,k \leq n}$ is positive too. This is why we can now easily formulate another characterisation of completely positive maps.

Corollary 3.17. *Let* $\Phi : \mathrm{Mat}_n(\mathbb{C}) \to \mathrm{Mat}_m(\mathbb{C})$ *be a linear map from. Then* Φ *is completely positive iff* $(\Phi(E_{jk}))_{1 \leq j,k \leq n}$ *is positive.*

Proof. If Φ is a completely positive map, then $(\Phi(E_{jk}))_{1 \leq j,k \leq n}$ is positive by the argument in theorem 3.16. Conversely, if $(\Phi(E_{jk}))_{1 \leq j,k \leq n}$ is positive and we take a look at said theorem, then Φ allows for a decomposition of the form

$$\Phi(A) = \sum_{i=1}^{nm} V_i^* A V_i$$

and is thereby completely positive. □

We have seen that the original proof of Choi's theorem merely uses methods from linear algebra which is remarkable in itself. Yet, there is another way of looking at the result above. In 1955 William Forrest Stinespring proved a dilation theorem from which Choi's results can be deduced.

Theorem 3.18 (Stinespring's dilation theorem, 1955)**.** *Let $\mathcal{A}$ be a unital C^*-algebra, $\mathcal{H}$ a Hilbert space and $B(\mathcal{H})$ the bounded operators on $\mathcal{H}$. Let $\Phi : \mathcal{A} \to B(\mathcal{H})$ be a completely positive map. Then there exists a Hilbert space $\mathcal{K}$, a unital $*$-homomorphism $\pi : \mathcal{A} \to B(\mathcal{K})$ and a bounded operator $V : \mathcal{H} \to \mathcal{K}$ with $\|\Phi(1)\| = \|V\|^2$, such that*

$$\Phi(a) = V^*\pi(a)V \quad \textit{for all} \quad a \in \mathcal{A}.$$

Proof. Let us consider the vector space $\mathcal{A} \otimes \mathcal{H}$. Define a symmetric bilinear function $\langle \cdot, \cdot \rangle$ on $\mathcal{A} \otimes \mathcal{H}$ by

$$\langle a \otimes x, b \otimes y \rangle := \langle \Phi(b^*a)x, y \rangle_{\mathcal{H}} \quad \text{with} \quad a,\, b \in \mathcal{A}, \quad x,\, y \in \mathcal{H}$$

and extend it bilinearly, where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner product on $\mathcal{H}$.

Since Φ is completely positive it follows that $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is positive semidefinite. Indeed, for any $n \geq 1, a_1, \ldots, a_n \in \mathcal{A}$ and $x_1, \ldots, x_n \in \mathcal{H}$ we have

$$\begin{aligned} \left\langle \sum_{j=1}^n a_j \otimes x_j, \sum_{i=1}^n a_i \otimes x_i \right\rangle &= \sum_{i,j=1}^n \langle \Phi(a_i^* a_j)x_j, x_i \rangle_{\mathcal{H}} \\ &= \left\langle \Phi_n((a_i^* a_j)) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\rangle_{\mathcal{H}^{(n)}} \geq 0, \end{aligned}$$

where $\langle \cdot, \cdot \rangle_{\mathcal{H}^{(n)}}$ denotes the inner product on the direct sum $\mathcal{H}^{(n)}$ of n copies of $\mathcal{H}$, given by

$$\left\langle \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\rangle_{\mathcal{H}^{(n)}} := \langle x_1, y_1 \rangle_{\mathcal{H}} + \ldots + \langle x_n, y_n \rangle_{\mathcal{H}}.$$

Positive semidefinite bilinear forms satisfy the Cauchy-Schwarz inequality,

$$|\langle u, v \rangle|^2 \leq \langle u, u \rangle \cdot \langle v, v \rangle,$$

hence

$$\mathcal{N} := \{u \in \mathcal{A} \otimes \mathcal{H} \mid \langle u, u \rangle = 0\} = \{u \in \mathcal{A} \otimes \mathcal{H} \mid \forall v \in \mathcal{A} \otimes \mathcal{H} : \langle u, v \rangle = 0\}$$

is a subspace of $\mathcal{A} \otimes \mathcal{H}$. This means that

$$\langle u + \mathcal{N}, v + \mathcal{N} \rangle := \langle u, v \rangle$$

is an inner product on the quotient space $\mathcal{A} \otimes \mathcal{H}/\mathcal{N}$. Let $\mathcal{K}$ be the completion of this space to a Hilbert space. For an element $a \in \mathcal{A}$, define $\pi(a) : \mathcal{A} \otimes \mathcal{H} \to \mathcal{A} \otimes \mathcal{H}$ by

$$\pi(a) \left(\sum a_i \otimes x_i \right) := \sum (aa_i) \otimes x_i.$$

Because of the properties of the tensor product $\pi(a)$ is clearly linear. $\pi(a)$ also satisfies the following inequality

$$\langle \pi(a)u, \pi(a)u \rangle \leq \|a\|^2 \langle u, u \rangle \quad \text{for all} \quad u \in \mathcal{A} \otimes \mathcal{H}. \tag{3.1}$$

To see this, observe that $a^*b^*ba \leq \|b\|^2 a^*a$ in any C^*-algebra. It follows that

$$(a_i^* a^* a a_j) \leq \|a\|^2 (a_i^* a_j)$$

is satisfied in $\mathrm{Mat}_n(\mathcal{A})^+$, i.e. the set of positive semidefinite $n \times n$ matrices over $\mathcal{A}$. Therefore,

$$\begin{aligned}
\left\langle \pi(a) \left(\sum_j a_j \otimes x_j \right), \pi(a) \left(\sum_i a_i \otimes x_i \right) \right\rangle &= \sum_{i,j} \langle \Phi(a_i^* a^* a a_j) x_j, x_i \rangle_{\mathcal{H}} \\
&\leq \|a\|^2 \sum_{i,j} \langle \Phi(a_i^* a_j) x_j, x_i \rangle_{\mathcal{H}} \\
&= \|a\|^2 \left\langle \sum_i a_j \otimes x_j, \sum_j a_i \otimes x_j \right\rangle.
\end{aligned}$$

Inequality 3.1 shows that $\pi(a)$ leaves $\mathcal{N}$ invariant since

$$u \in \mathcal{N} \iff \langle u, u \rangle = 0.$$

By said inequality for any $u \in \mathcal{N}$

$$\langle \pi(a)u, \pi(a)u \rangle = 0$$

and thus $\pi(a)u \in \mathcal{N}$. This is why we can view $\pi(a)$ as a linear operator on $\mathcal{A} \otimes \mathcal{H}/\mathcal{N}$ which we will still denote by $\pi(a)$. Again by 3.1 we can see that $\pi(a)$ is bounded, i.e. $\|\pi(a)\| \leq \|a\|$. Therefore, it extends to a bounded linear operator on $\mathcal{K}$, which we still denote by $\pi(a)$. Furthermore, $\pi : \mathcal{A} \to B(\mathcal{K})$ is a unital $*$-homomorphism and we end by defining $V : \mathcal{H} \to \mathcal{K}$ via

$$Vx = 1 \otimes x + \mathcal{N}.$$

Clearly, V is linear and we have that V is bounded since

$$\begin{aligned}
\|Vx\|^2 &= \langle 1 \otimes x, 1 \otimes x \rangle \\
&= \langle \Phi(1)x, x \rangle_{\mathcal{H}} \\
&\leq \|\Phi(1)\| \|x\|^2
\end{aligned}$$

for $x \in \mathcal{H}$ by the Cauchy-Schwarz inequality. Thus,

$$\begin{aligned}
\|V\|^2 &= \sup\{\|Vx\|^2 : \|x\| \leq 1\} \\
&= \sup\{\langle \Phi(1)x, x \rangle_{\mathcal{H}} : \|x\| \leq 1\} = \|\Phi(1)\|
\end{aligned}$$

because $\Phi(1)$ is a bounded operator on $\mathcal{H}$. Finally,

$$\langle V^* \pi(a) V x, y \rangle_{\mathcal{H}} = \langle \pi(a) 1 \otimes x, 1 \otimes y \rangle_{\mathcal{K}} = \langle \Phi(a) x, y \rangle_{\mathcal{H}}$$

for all x and y in $\mathcal{H}$ hence $V^* \pi(a) V = \Phi(a)$ which completes the proof. $\square$

Remark/Example 3.19. The triple $(\pi, V, \mathcal{K})$ is called a **Stinespring representation** of Φ. If we let

$$\mathcal{K}_1 := \overline{\operatorname{span}}_{\mathcal{K}}(\pi(\mathcal{A})V\mathcal{H}) = \overline{\operatorname{span}}_{\mathcal{K}}(\{\pi(a)Vh : a \in \mathcal{A} \text{ and } h \in \mathcal{H}\}),$$

then $V\mathcal{H}$ lies in $\mathcal{K}_1$ since π is unital. Therefore, we can assume that $V : \mathcal{H} \to \mathcal{K}_1$. In addition $\pi(a)(\mathcal{K}_1)$ lies in $\mathcal{K}_1$ for all $a \in \mathcal{A}$ since π is multiplicative and continuous. So $(\pi_1, V, \mathcal{K}_1)$ is also a Stinespring representation of Φ with the additional property that $\mathcal{K}_1$ is the closed linear span of $\pi(\mathcal{A})V\mathcal{H}$. Such a representation is called a **minimal Stinespring representation** of Φ. A minimal representation is unique up to some unitary operator. △

Having now proven Stinespring's dilation theorem we want go on proving Choi's theorem as a corollary. In order to do so, we need some more information about unital $*$-homomorphisms.

Lemma 3.20. *Let $\pi : \mathrm{Mat}_n(\mathbb{C}) \to B(\mathcal{K})$ be a unital $*$-homomorphism. Then there exists a Hilbert space $\mathcal{H}$ such that*

$$\mathcal{K} \cong \underbrace{\mathcal{H} \oplus \ldots \oplus \mathcal{H}}_{n \text{ times}}.$$

and $\pi : \mathrm{Mat}_n(\mathbb{C}) \to B(\mathcal{K}) \cong \mathrm{Mat}_n(B(\mathcal{H}))$ satisfies $\pi(E_{ij}) = \widetilde{E}_{ij}$ for all $i, j = 1, \ldots, n$ where E_{ij} and $\widetilde{E}_{ij}$ are the canonical basis elements in $\mathrm{Mat}_n(\mathbb{C})$ and $\mathrm{Mat}_n(B(\mathcal{H}))$ respectively.

Proof. Define $\mathcal{H}_i := \pi(E_{ii})\mathcal{K}$ for all $i = 1, \ldots, n$. Then $\mathcal{K} = \mathcal{H}_1 \oplus \ldots \oplus \mathcal{H}_n$ which is easy to see as π is a unital $*$-homomorphism and therefore

$$\begin{aligned}
\mathcal{H}_1 + \ldots + \mathcal{H}_n &= \pi(E_{11})\mathcal{K} + \ldots + \pi(E_{nn})\mathcal{K} \\
&= (\pi(E_{11}) + \ldots + \pi(E_{nn}))\mathcal{K} \\
&= \pi(I_n)\mathcal{K} = \mathcal{K}
\end{aligned}$$

and

$$\begin{aligned}
\langle H_i, H_j \rangle &= \langle \pi(E_{ii})\mathcal{K}, \pi(E_{jj})\mathcal{K} \rangle \\
&= \langle \mathcal{K}, \pi(E_{ii})^* \pi(E_{jj})\mathcal{K} \rangle \\
&= \langle \mathcal{K}, \pi(E_{ii}^*) \pi(E_{jj})\mathcal{K} \rangle \\
&= \langle \mathcal{K}, \pi(E_{ii}E_{jj})\mathcal{K} \rangle = 0
\end{aligned}$$

which means $\mathcal{H}_i \perp \mathcal{H}_j$ for $i \neq j$. We claim that $\mathcal{H}_1, \ldots, \mathcal{H}_n$ are isometric isomorphic. Since the range of $\pi(E_{ji})$ lies in $\mathcal{H}_j$, $U_{ji} := \pi(E_{ji})|_{\mathcal{H}_i}$ is well-defined as an operator from $\mathcal{H}_i$ to $\mathcal{H}_j$. Clearly, U_{ji} is linear and since $\pi(E_{ij})\mathcal{K}$ lies in $\mathcal{H}_i$ and $U_{ji}\pi(E_{ij})x = \pi(E_{jj})x$, U_{ji} is surjective. Again for an element $\pi(E_{ii})x$ of $\mathcal{H}_i$

$$\begin{aligned}
\langle U_{ji}\pi(E_{ii})x, U_{ji}\pi(E_{ii})y \rangle &= \langle \pi(E_{ji})x, \pi(E_{ji})y \rangle \\
&= \langle \pi(E_{ji})x, \pi(E_{ij})^* y \rangle \\
&= \langle \pi(E_{ii})x, y \rangle = \langle \pi(E_{ii})x, \pi(E_{ii})y \rangle.
\end{aligned}$$

This means that U_{ji} is surjective and preserves the inner product and is thus one-to-one. Finally,

$$\begin{aligned}
\langle \pi(E_{ii})x, \pi(E_{ii})y \rangle &= \langle \pi(E_{ii})x, y \rangle \\
&= \langle \pi(E_{ij}E_{ji})x, y \rangle \\
&= \langle \pi(E_{ji})x, \pi(E_{ji})y \rangle
\end{aligned}$$

thus $U_{ji}^{-1} = U_{ij}$. Every operator on $\mathcal{K} = \mathcal{H}_1 \oplus \ldots \oplus \mathcal{H}_n$ can be represented by a $n \times n$ matrix with operator entries. We see that $\pi(E_{ij})$ corresponds to the matrix which is U_{ij} in the i, j-th entry and 0 elsewhere since $U_{ij} = \pi(E_{ij})|_{\mathcal{H}_j}$ by definition. We have shown above that the operator U_{ij} maps $\mathcal{H}_i$ and $\mathcal{H}_j$ isometrically and bijectively onto each other and now for $h \in \mathcal{K}$

$$\pi(E_{ij}) \underbrace{\pi(E_{jj})h}_{\in \mathcal{H}_j} = \pi(E_{ij}E_{jj})h = \pi(E_{ij})h \cong h$$

meaning $\pi(E_{ij}) = \widetilde{E}_{ij}$. □

In order to prove the corollary we want to take a closer look at the finite dimensional case.

Lemma 3.21. *Let $\mathcal{K}$ be a finite dimensional Hilbert space and $\pi : \mathrm{Mat}_n(\mathbb{C}) \to B(\mathcal{K})$ be a unital $*$-homomorphism. Then*

$$\mathcal{K} \cong \underbrace{\mathbb{C}^n \oplus \ldots \oplus \mathbb{C}^n}_{r\ times}$$

where $r := \dim \mathcal{K}/n$. Furthermore, $\pi : \mathrm{Mat}_n(\mathbb{C}) \to B(\mathcal{K}) \cong B(\mathbb{C}^n \oplus \ldots \oplus \mathbb{C}^n) \cong \mathrm{Mat}_r(B(\mathbb{C}^n)) \cong \mathrm{Mat}_r(\mathrm{Mat}_n(\mathbb{C}))$ satisfies

$$\pi(A) = \begin{pmatrix} A & 0 & \cdots & 0 \\ 0 & A & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & A \end{pmatrix}$$

Proof. By the lemma above $\mathcal{K} \cong \underbrace{\mathbb{C}^r \oplus \ldots \oplus \mathbb{C}^r}_{n \text{ times}}$ such that

$$\pi : \mathrm{Mat}_n(\mathbb{C}) \to B(\mathcal{K}) \cong B(\mathbb{C}^r \oplus \ldots \oplus \mathbb{C}^r) \cong \mathrm{Mat}_n(B(\mathbb{C}^r)) \cong \mathrm{Mat}_n(\mathrm{Mat}_r(\mathbb{C}))$$

satisfies

$$\pi(A) = \begin{pmatrix} a_{11}I_r & \cdots & a_{1n}I_r \\ \vdots & \ddots & \vdots \\ a_{n1}I_r & \cdots & a_{nn}I_r \end{pmatrix}$$

where I_r is the identity in $\mathbb{C}^r$. Be reshuffling rows and columns we obtain the desired result. □

Corollary 3.22 (Choi's theorem). *Let $\Phi : \mathrm{Mat}_n(\mathbb{C}) \to \mathrm{Mat}_k(\mathbb{C})$ be completely positive. Then there exist at most nk linear maps $V_i : \mathbb{C}^k \to \mathbb{C}^n$ such that $\Phi(A) = \sum_i V_i^* A V_i$ for all $A \in \mathrm{Mat}_n(\mathbb{C})$.*

Proof. Let $(\Phi, V, \mathcal{K})$ be a Stinespring representation for Φ. Having a look at the proof for Stinespring's dilation theorem we know that $\dim K \leq \dim(\mathrm{Mat}_n(\mathbb{C}) \otimes \mathbb{C}^k) = n^2 k$. Since π is a unital $*$-homomorphism, we can write

$$\mathcal{K} \cong \underbrace{\mathbb{C}^n \oplus \ldots \oplus \mathbb{C}^n}_{r \text{ times}}$$

with $r \leq nk$ by the lemma above such that $\pi : \mathrm{Mat}_n(\mathbb{C}) \to B(\mathcal{K}) \cong B(\mathbb{C}^n \oplus \ldots \oplus \mathbb{C}^n) \cong \mathrm{Mat}_r(B(\mathbb{C}^n)) \cong \mathrm{Mat}_r(\mathrm{Mat}_n(\mathbb{C}))$ satisfies

$$\Phi(A) = \begin{pmatrix} A & 0 & \cdots & 0 \\ 0 & A & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & A \end{pmatrix}.$$

$V : \mathbb{C}^k \to \mathcal{K} \cong \mathbb{C}^n \oplus \ldots \oplus \mathbb{C}^n$ can be represented as a column operator matrix

$$V = \begin{pmatrix} V_1 \\ \vdots \\ V_r \end{pmatrix}$$

for some $V_i : \mathbb{C}^k \to \mathbb{C}^n$ and so $V^* = (V_1^*, \ldots, V_r^*)$. Therefore,

$$\Phi(A) = V^*\pi(A)V = (V_1^*, \ldots, V_r^*) \begin{pmatrix} A & 0 & \cdots & 0 \\ 0 & A & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & A \end{pmatrix} \begin{pmatrix} V_1 \\ \vdots \\ V_r \end{pmatrix} = \sum_{i=1}^{r} V_i^* A V_i$$

giving us the desired result. $\square$

Chapter 4

Hyponymy and compositional models of meaning

After an introduction to computational linguistics and vector-based models of meaning followed by some theory on operators and completely positive maps culminating in Choi's theorem we want to amalgamate the two. Our goal is to combine word representations by some map and get representations for combinations of words, e.g. for sentences or phrases.

4.1 From vectors to matrices

First, we turn our vector representations into matrices by computing the outer product which means for a word v we compute

$$v\, v^T.$$

Words are now generally represented by rank-one matrices and furthermore representations are symmetric and positive semidefinite since

$$(v\, v^T)^T = (v^T)^T\, v^T = v\, v^T$$

and

$$h^T(v\, v^T)h = (h^T\, v)\,(v^T\, h) = \langle h, v\rangle^2 \geq 0.$$

Even though straight forward, this is a smart way to associate words with matrices for two reasons. On the one hand taking v to be an element of $\mathbb{R}^n$ for some, possibly large, $n \in \mathbb{N}$ we get that the representations satisfy definition 3.12 and are thus positive elements of the C^*-algebra $\mathrm{Mat}_n(\mathbb{C})$. On the other hand, the set of positive semidefinite matrices forms a convex cone which hints at summation for representing higher level concepts, i.e. some word ρ might be defined by

$$\rho = \sum_i v_i\, v_i^T$$

for some words v_i. In section 2.3 we have introduced a toy example from [3] and with its help let us contextualise the approach above. First, we need to turn the columns of table 2.1 into matrices by use of the outer product

$$\mathsf{pug} = \begin{pmatrix} 9 & 12 & 0 & 0 \\ 12 & 16 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \mathsf{goldfish} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 25 & 0 & 30 \\ 0 & 0 & 0 & 0 \\ 0 & 30 & 0 & 36 \end{pmatrix} \quad \mathsf{tabby} = \begin{pmatrix} 25 & 25 & 0 & 0 \\ 25 & 25 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and then the concept of *pet* can be expressed by the sum like so

$$\mathsf{pet} = \mathsf{pug} + \mathsf{goldfish} + \mathsf{tabby} = \begin{pmatrix} 34 & 37 & 0 & 0 \\ 37 & 66 & 0 & 30 \\ 0 & 0 & 0 & 0 \\ 0 & 30 & 0 & 36 \end{pmatrix}.$$

One can already guess with this set-up that the sum of words reflects a semantic phenomenon. In the next section we therefore want to introduce the notion of hyponymy.

4.2 Hyponymy

Second, we want to model the semantic relation of hyponymy. So, what is hyponymy? The Greek roots ὑπό and ὄνυμα – meaning "under" or "beneath" and "name" respectively – already give away the core concept. Hyponymy denotes the relation between a *hyponym* which is a subtype and a *hypernym* which is a supertype. For example, *banana* is a hyponym of the hypernym *fruit* and at least for nouns hyponymy can be characterised by a *type-of-relationship*, e.g. *a bus is a type of vehicle*. Still, the idea extends to various word classes since for example *to walk* is a hyponym of *to move*.

In summary, hyponymy is a relation on the set of all words in our vector space. From the description above we can state that the relation is

(i) **reflexive**: This might give reason for philosophical debate but arguably "A banana is a type of banana." is, though tautological, a true statement.

(ii) **antisymmetric**: A banana is a type of fruit yet a fruit is not a type of banana.

(iii) **transitive**: The Cavendish is a type of banana and a banana is a type of fruit; thus the Cavendish is a type of fruit.

The mathematically inclined might spot this as the very definition of a partial ordering and since the positive semidefinite matrices form a convex cone and every convex cone induces a partial ordering on its elements we can define a "natural" ordering.

Lemma 4.1. *If $C \subseteq \mathbb{R}^n$ is a salient convex cone, i.e. $C \cap (-C) = \{0\}$, there exists a partial ordering $\leqslant$ of the vector space defined by*

$$x \leqslant y :\Longleftrightarrow \; y - x \in C.$$

Proof. (i) Since by the definition of a convex cone $0 \in C$, it is true that $x - x = 0 \in C$ for any $x \in \mathbb{R}^n$. So $x \leqslant x$, proving the reflexivity.

(ii) Since C is salient $C \cap (-C) = \{0\}$ which implies $x = y$ for any $x, y \in \mathbb{R}^n$ with $x \leqslant y$ and $y \leqslant x$ since

$$y - x \in C \quad \text{and} \quad x - y \in C \quad \Longrightarrow \quad y - x \in C \cap (-C) = \{0\}.$$

(iii) For $x, y, z \in \mathbb{R}^n$ with $x \leqslant y$ and $y \leqslant z$

$$y - x \in C \quad \text{and} \quad z - y \in C.$$

Since C is convex

$$(y - x) + (z - y) \in C$$

implying $z - x \in C$ thus $x \leqslant z$. □

The preceding lemma now motivates the Löwner order on the convex cone of positive semidefinite matrices.

Definition 4.2 (Löwner order)**.** Let A and B be two Hermitian matrices. Then

$$A \leqslant B :\Longleftrightarrow B - A \text{ is positive semidefinite.}$$ △

By the construction of supertypes in section 4.1 it is clear how the Löwner order can be used for modelling hyponymy. Let us again take a look at the *pet* example. Intuitively, *pug* is a hyponym of the hypernym *pet* and by construction

$$\mathsf{pet} - \mathsf{pug} = \begin{pmatrix} 34 & 37 & 0 & 0 \\ 37 & 66 & 0 & 30 \\ 0 & 0 & 0 & 0 \\ 0 & 30 & 0 & 36 \end{pmatrix} - \begin{pmatrix} 9 & 12 & 0 & 0 \\ 12 & 16 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 25 & 25 & 0 & 0 \\ 25 & 50 & 0 & 30 \\ 0 & 0 & 0 & 0 \\ 0 & 30 & 0 & 36 \end{pmatrix}$$

is a positive semidefinite matrix having the eigenvalues $\frac{1}{2}\left(111 \pm \sqrt{2621}\right)$ and 0. This means $\mathsf{pug} \leqslant \mathsf{pet}$.

4.3 Composition

It is now we start to wonder how our model should reflect the composition of words into sentences and phrases. If words are represented by positive semidefinite matrices and we assume, for simplicity's sake, that every matrix has the same dimensions, say $n \times n$, then we would like to define a map like the following

$$\varphi : \mathrm{Mat}_n(\mathbb{R}) \times \mathrm{Mat}_n(\mathbb{R}) \to \mathrm{Mat}_n(\mathbb{R}).$$

Here, we explicitly want the composition to be of the same dimensions as the arguments which eases the use of φ in iterative composition, i.e. $\varphi(s, \varphi(o, v))$ makes sense for any subject s, verb v and object o representing the sentence $s\ v\ o$.

The next few pages are taken from [1] where a framework for compositional maps that relies on *Choi's theorem* from chapter 3 is proposed. Henceforth, let $\mathrm{Mat}_n(\mathbb{R})^+$ denote the set of real, positive semidefinite matrices. We start by stating the minimal requirements for our map φ.

(i) *Preserves positivity*
If n, v are positive semidefinite then $\varphi(n, v)$ should be too, i.e.

$$\varphi : \mathrm{Mat}_n(\mathbb{R})^+ \times \mathrm{Mat}_n(\mathbb{R})^+ \to \mathrm{Mat}_n(\mathbb{R})^+.$$

(ii) *Preserves hyponymy*
Using the Löwner order as introduced above to model hyponymy we require for all words $n_1, n_2, v_1, v_2 \in \mathrm{Mat}_n(\mathbb{R})$ satisfying

$$n_1 \leqslant n_2 \text{ and } v_1 \leqslant v_2$$

that

$$\varphi(n_1, v_1) \leqslant \varphi(n_2, v_2).$$

(iii) *Bilinearity*
For $\alpha \in \mathbb{R}$ and $n, n', v, v' \in \mathrm{Mat}_n(\mathbb{R})$:

(a) $\varphi(\alpha n, v) = \alpha\varphi(n, v)$
(b) $\varphi(n, \alpha v) = \alpha\varphi(n, v)$
(c) $\varphi(n + n', v) = \varphi(n, v) + \varphi(n', v)$
(d) $\varphi(n, v + v') = \varphi(n, v) + \varphi(n, v')$

Actually, properties (i) and (iii) already imply (ii). Take $n_2 \geqslant n_1$ and $v_2 \geqslant 0$. Then by (i) it is true that $\varphi(n_2 - n_1, v_2) \geqslant 0$ and by the bilinearity $\varphi(n_1, v_2) \leqslant \varphi(n_2, v_2)$. By the same argument $\varphi(n_1, v_1) \leqslant \varphi(n_1, v_2)$ and using the transitivity of $\leqslant$ we get that

$$\varphi(n_1, v_1) \leqslant \varphi(n_2, v_2).$$

Additionally, we might want to exclude maps where $\varphi(n, v) = \varphi(v, n)$ since these lead to nonsensical representations, e.g. "Calvin likes Hobbes" and "Hobbes likes Calvin" would have the same meaning representation.

Initially, imagine that we have already found a map φ with the properties (i)–(iii). Then, by the bilinearity we can reformulate the map somewhat

$$\begin{aligned} \mathrm{Mat}_n(\mathbb{R}) &\to \mathrm{Lin}(\mathrm{Mat}_n(\mathbb{R}), \mathrm{Mat}_n(\mathbb{R})) \\ v &\mapsto (\varphi(\cdot, v) : n \mapsto \varphi(n, v)). \end{aligned}$$

The linearity in the first component ensures that the image is a linear map whereas the second component turns the map itself into a linear one. Let us denote this map by

$$\Phi : \mathrm{Mat}_n(\mathbb{R}) \to \mathrm{Lin}(\mathrm{Mat}_n(\mathbb{R}), \mathrm{Mat}_n(\mathbb{R})).$$

It is clear that if we know Φ we also know φ and since $\mathrm{Mat}_n(\mathbb{R}) \cong \mathbb{R}^{n^2}$ we see that

$$\mathrm{Lin}(\mathrm{Mat}_n(\mathbb{R}), \mathrm{Mat}_n(\mathbb{R})) \cong \mathrm{Mat}_{n^2}(\mathbb{R})$$

To be more precise, we take a look at Choi's isomorphism

$$\begin{aligned}\operatorname{Lin}(\operatorname{Mat}_n(\mathbb{R}), \operatorname{Mat}_n(\mathbb{R})) &\to \operatorname{Mat}_n(\mathbb{R}) \otimes \operatorname{Mat}_n(\mathbb{R}) \\ \varphi &\mapsto \sum_{i,j} \varphi(e_i\, e_j^T) \otimes e_i\, e_j^T\end{aligned}$$

where $\{e_1, \ldots, e_n\}$ are the standard basis vectors in $\mathbb{R}^n$. Since φ fulfils property (i) we have that Φ maps positive semidefinite matrices in $\operatorname{Mat}_n(\mathbb{R})$ to positivity preserving linear maps in $\operatorname{Lin}(\operatorname{Mat}_n(\mathbb{R}), \operatorname{Mat}_n(\mathbb{R}))$. From the definition of the isomorphism we see that these positivity preserving maps correspond to the tensor elements $\sum_i A_i \otimes B_i$ that fulfil the following property of positivity

$$\sum_i v^* A_i v \cdot w^* B_i w \geq 0$$

for any two vectors $v, w \in \mathbb{C}^n$. These are called block positive matrices.

Still, we have no explicit description of Φ. This is where *Choi's theorem on completely positive maps* comes in. Remember that in the proof of said theorem we have looked at the matrix $(E_{ij})_{1 \leq i,j \leq n}$ using the complete positivity to show that the map allows for a decomposition. Since $e_i e_j^T = E_{ij}$, we see that Choi's isomorphism maps completely positive maps onto positive semidefinite matrices in $\operatorname{Mat}_n(\mathbb{R}) \otimes \operatorname{Mat}_n(\mathbb{R}) \cong \operatorname{Mat}_{n^2}(\mathbb{R})$ and is exactly what we have used in chapter 3. So additionally, we require Φ to be completely positive itself, i.e.

$$\Phi : \operatorname{Mat}_n(\mathbb{R})^+ \to \operatorname{Mat}_{n^2}(\mathbb{R})^+$$

allowing us to finally write the map as

$$\begin{aligned}\Phi : \operatorname{Mat}_n(\mathbb{R}) &\to \operatorname{Mat}_n(\mathbb{R}) \otimes \operatorname{Mat}_n(\mathbb{R}) \cong \operatorname{Mat}_{n^2}(\mathbb{R}) \\ v &\mapsto \sum_{i=1}^{n^2} V_i^* v V_i\end{aligned}$$

by Choi's theorem.

Now, one might ask in what way this description is any better. Admittedly, $\sum_{i=1}^{n^2} V_i^* v V_i$ is still rather abstract but in [1] the authors also make use of a diagrammatic calculus to systematically categorise these maps and obtain explicit descriptions which can then be tested against other compositional maps from the literature on whether the preservation of hyponymy is better or worse using real world test data. This part of the paper will not be discussed in this thesis and neither will the diagrammatic calculus. The interested reader is thus referred to [1] and [9].

Chapter 5

Summary

It is time for a recap. The problem of creating meaning representations for semantic objects that are more complex than single words raises the question whether hyponymy can be preserved under such circumstances. The title of [1] explains the goal fittingly: *Cats climb entails mammals move* – where cat being a hyponym of mammal and climb a hyponym of move the combination should satisfy the hyponym-hypernym relation. In a sense, we have put the cart before the horse because we initially assumed to have already found a map that by definition satisfies the preservation of hyponymy.

The succeeding analysis is at its core not very difficult to understand. Every bilinear map φ can be turned into Φ by the manner presented in chapter 4 and identifying linear maps between two finite n^2-dimensional vector spaces with $n^2 \times n^2$ matrices is relatively straightforward. Still, the class of maps that satisfies the constraining properties cannot be described explicitly – save a subset of mappings satisfying one more condition: *complete positivity*.

Equipped with *Choi's theorem on completely positive maps* (or Stinespring's dilation theorem for that matter) we could come to grips with these unruly and abstract definitions of mappings between matrix algebras by requiring one more thing, namely the complete positivity of the map Φ.

From the perspective of the 1950s and 1960s when distributional semantics was still in its infancy it is surprising that operator theory and completely positive maps show up in linguistics, especially since these mathematical concepts are also tied to quantum physics creating an interesting link between the natural sciences and the study of human language. This just goes to show how unpredictable the application of mathematics can be and that a theorem might find its use in many different settings unbeknownst to the author coming up with the initial proof.

Bibliography

[1] Gemma De las Cuevas, Andreas Klingler, Martha Lewis and Tim Netzer. *Cats climb entails mammals move: preserving hyponymy in compositional distributional semantics*. 2020. arXiv: 2005.14134.

[2] Zellig S Harris. 'Distributional structure'. In: *Word* 10.2-3 (1954), pp. 146–162.

[3] Martha Lewis. 'Compositional hyponymy with positive operators'. In: *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*. 2019, pp. 638–647.

[4] Madeleine Bates. 'Models of natural language understanding'. In: *Proceedings of the National Academy of Sciences* 92.22 (1995), pp. 9977–9982.

[5] Vern Paulsen. *Completely bounded maps and operator algebras*. 78. Cambridge University Press, 2002.

[6] Ronald G Douglas. *Banach algebra techniques in operator theory*. Vol. 179. Springer Science & Business Media, 2012.

[7] Man-Duen Choi. 'Completely positive linear maps on complex matrices'. In: *Linear algebra and its applications* 10.3 (1975), pp. 285–290.

[8] Ali Samil Kavruk. *Complete Positivity in operator algebras*. 2006.

[9] Bob Coecke, Ross Duncan, Aleks Kissinger and Quanlong Wang. *Generalised Compositional Theories and Diagrammatic Reasoning*. 2015. arXiv: 1506.03632.

[10] Tim Netzer. *Lecture notes in Linear Algebra*. 2019.

[11] Tim Netzer. *Lecture notes in Algebra*. 2020.

universität innsbruck

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Stellen, die wörtlich oder inhaltlich den angegebenen Quellen entnommen wurden, sind als solche kenntlich gemacht.

Ich erkläre mich mit der Archivierung der vorliegenden Bachelorarbeit einverstanden.

29.08.2021	Mair Andreas
Datum	Unterschrift