Applying Distributional Compositional Categorical Models of Meaning to Language Translation

Brian Tyrrell
Department of Mathematics
University of Oxford
brian.tyrrell@maths.ox.ac.uk

The aim of this paper is twofold: first we will use vector space distributional compositional categorical models of meaning to compare the meaning of sentences in Irish and in English (and thus ascertain when a sentence is the translation of another sentence) using the cosine similarity score. Then we shall outline a procedure which translates nouns by understanding their context, using a conceptual space model of cognition. We shall use metrics on the category **ConvexRel** to determine the distance between concepts (and determine when a noun is the translation of another noun). This paper will focus on applications to Irish, a member of the Gaelic family of languages.

1 Introduction

The raison d'être of Distributional Compositional Categorical (henceforth referred to as DisCoCat) Models of Meaning originates in the oft quoted mantra of the field:

> *"You shall know a word by the company it keeps."*
> -John R. Firth, *A synopsis of linguistic theory 1930-1955*, (1957).

The broad idea of such models in natural language processing is to marry the semantic information of words with the syntactic structure of a sentence using category theory to produce the whole meaning of the sentence. The semantic information of a word is captured (in early models, cf. [7, 14, 13]) by a vector in a tensor product of vector spaces using a corpus of text to represent a given word in terms of a fixed basis of other words; i.e. by distributing the meaning of the word across the corpus. In later models [3] convex spaces are used instead of vector spaces in an effort to better capture the representation of words in the human mind.

It is the focus of this paper to exploit the existing DisCoCat structure for language translation. We shall use a vector space model of meaning, defined by Coecke et al. [7] and introduced in *Section 3*, to assign meaning to sentences in English and then in Irish. These meanings are then compared via an inner product on the shared sentence space of English and Irish vector space models of meaning in *Section 4*. We discuss the results of this on example sentences in *Section 4* & *Appendix C*. The next three sections of the paper focus on using conceptual spaces (or *concepts*) in place of vector spaces to understand the meaning of nouns. In *Section 5* we work on a system of adjective & noun classification which, in *Section 6*, leads to the generation of convex spaces representing noun concepts from a given corpus. While other authors (cf. Derrac & Schockaert [8]) have also induced conceptual spaces algorithmically from corpora, the treatment we propose is tailored towards the authors use of metrics in *Section 7* to determine the 'distance' between concepts coming from different languages. The system we will propose cannot capture every type of adjective, however it is sufficiently complex and complete to allow us to start analysing text in a meaningful way.

B. Coecke, J. Hedges, D. Kartsaklis, M. Lewis, D. Marsden (Eds.):
2018 Workshop on Compositional Approaches
for Physics, NLP, and Social Sciences (CAPNS)
EPTCS 283, 2018, pp. 28–49, doi:10.4204/EPTCS.283.3

Before this, we must determine the Lambek pregroup grammar structure for Irish (which does not exist in the current literature) and, as we shall see in *Section 2*, is nontrivial in some aspects. We will only deal with an elementary fragment of Irish, in much the same way an elementary fragment of English is used in [7]. The ideas presented in this paper can be applied to many other languages, however the author has chosen Irish due to its relative rareness in literature and its high regularity in verb structure. For instance, across all of Irish there exist exactly eleven irregular verbs; with the exception of these eleven, every other verb can be conjugated in an extremely efficient and easy manner.

2 Lambek Pregroup Grammar Structure for Irish

Coecke et. al. use the Lambek pregroup grammar structure to determine maps which assign meanings to sentences [7, §3.5]. This is all built exclusively through English, but there are no barriers to moving to a different language; Lambek et al. [1, 2, 5, 18] detail a pregroup structure for French, Arabic, Latin and German, respectively. However the author cannot find evidence of the same treatment in Irish. Thus, in order to create Irish DisCoCat models, we must create a 'Lambek Pregroup Grammar' for the language.

2.1 Irish Grammatical Structure

For our purposes, we do not need a structure as complicated as Lambek's work [17], rather we shall mirror the English approach: four basic types - nouns (n), declarative statements (s), infinitives of verbs (j) and gluing types (σ). We hand construct the following compound types:

(1) **Transitive verbs** are assigned the type $sn^l n^l$. This is because Irish follows the rule *Verb Subject Object*. The only exception to this is the copula *is*, which we also assign the type $sn^l n^l$ but note the sentence structure must be *Verb Object Subject*. This verb-like word is used in sentences that state equivalences between, or crucial attributes of, the subject and object.

For example, even though the Irish for the verb "to be" is *bí*, which in the present tense is *tá*, one would say

Is dochtúir mé	for	"I am a doctor" and
Tá scamaill sa spéir	for	"There are clouds in the sky",

(2) **Adjectives** are assigned the type $n^r n$. This is because Irish follows the rule *Noun Adjective*.

(3) **Adverbs** are assigned the type $s^r s$; they appear at the end of sentences.

(4) **Prepositions** as *whole phrases* are assigned the type $n^r n$. This is because Irish follows the rule *Preposition Noun*, as in English, so we give the same type assignment as in [14]. Note that prepositions in Irish always come before the noun, and adjectives after, so we cannot confuse them.

It should be noted that Irish (sometimes) modifies the noun after a preposition directly by inserting an *urú* or *séimhiú* into the noun - additional letters to change the pronunciation of the word. So, for example, whilst *table* in Irish is *bord*, *on the table* becomes *ar an* ***m****bord*. This is a sign that correlates with the change in type assignment of the affected noun.

Consider the following demonstration:

Example 2.1. English sentence: "I broke the vase under the large bridge yesterday" In Irish: "*Bhris mé an vása faoin droichead mór inné*". This has the type assignment

Bhris	mé	an vása	faoin droichead	mór	inné
$sn^l n^l$	n	n	$n^r n$	$n^r n$	$s^r s$

◊

Finally, Sadrzadeh et al. [23] consider subject relative pronouns (such as *who(m)* and *which*) and object relative pronouns (such as *that*). They assign the pregroup types as follows:

$$n^r n s^l n \text{ (subject relative pronoun)} \qquad n^r n n^{ll} s^l \text{ (object relative pronoun)}$$

However, in Irish these particular words (*who(m), which, that*) are simply represented by one word: *a*. Moreover, the grammatical structure of a sentence containing these relative pronouns is the same regardless of whether the relative pronouns are object or subject modifying; the sentence structure is always "noun relative-pronoun verb noun". So for Irish we can define:

(5) **Relative Pronouns.** Let $n^r n n^{ll} s^l$ be the pregroup type of *a*, the Irish relative pronoun *who(m), which,* and *that*.

This concludes the work required to use a pregroup grammar structure in Irish.

3 A Vector Space based Model of Meaning

We will now create a vector space model of meaning from *Corpus A.1*, located in *Appendix A*. The section after this will create another vector space model of meaning, this time in Irish, from the translation of *Corpus A.1*. The underlying principal is that once we have the meaning of a sentence in an abstract vector space S, it does not matter what the language of the sentence is, as it can be compared via an inner product on S. An application of this idea is to measure the accuracy of translation tools such as *Google Translate*, and also to potentially train software (off large corpora) to accept input commands in any language.

The corpus of text chosen by the author is a modified copy of the plot of *Star Wars: Episode III - Revenge of the Sith* obtained from Wikipedia. The full corpus of text is presented in *Appendix A*. We shall closely follow the exposition presented by Grefenstette and Sadrzadeh [12, 13] throughout.

As we are primarily interested in the vector space N of nouns, we shall begin there. We define the basis to consist of the five most commonly occurring words against which we shall measure all other nouns in the corpus:

$$\text{Basis of } N = \{\text{Anakin, Palpatine, Jedi, Obi-Wan, arg-evil}\},$$

where 'arg-evil' denotes the argument of the adjective 'evil' (cf. [14, §3]). The coordinates of a noun K follow from counting the number of times each basis word has appeared in an m word window around K; in particular, K is given a coordinate of k for 'arg-evil' if K has appeared within m words of a noun described as 'evil' in the same sentence, k times in the corpus. For this paper, set $m = 3$. In this basis

Anakin $= [1,0,0,0,0]$,	Padmé $= [5,0,0,2,0]$,	Mace Windu $= [1,1,2,0,0]$,
Palpatine $= [0,1,0,0,0]$,	Yoda $= [0,2,1,3,1]$,	Sith Lord $= [1,1,0,0,1]$,
Jedi $= [0,0,1,0,0]$,	Emperor $= [1,5,0,1,1]$,	General Grievous $= [0,1,4,1,1]$,
Obi-Wan $= [0,0,0,1,0]$,	mastermind $= [2,2,0,0,1]$,	dark side of the Force $= [4,2,1,1,1]$,
arg-evil $= [0,0,0,0,1]$,	evil (noun) $= [0,2,1,0,0]$	

where we treat 'dark side of the Force', as one noun. It has appeared within 3 words of 'Anakin' 4 times, 'Palpatine' 2 times, 'Jedi' once, 'Obi-Wan' once, and the argument of 'evil' once (see *Corpus A.1*).

As described by Grefenstette and Sadrzadeh [12] there exists an exact procedure for learning the weights for matrices of words. If we assume $S = N \otimes N$ (so the basis of S is of the form $\overrightarrow{n}_i \otimes \overrightarrow{n}_k$) then the meaning vector of a transitive verb in a sentence is

$$\overrightarrow{\text{subject verb object}} = \overrightarrow{verb} \odot (\overrightarrow{subject} \otimes \overrightarrow{object}),$$

where $\odot$ is the Kronecker product. A transitive verb is described by a two dimensional matrix; using the data of *Corpus A.1*,

$$C^{\text{turn}} = \begin{bmatrix} 8 & 7 & 4 & 2 & 2 \\ 1 & 0 & 0 & 0 & 0 \\ 4 & 2 & 1 & 1 & 1 \\ 0 & 2 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad C^{\text{is}} = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 5 & 4 & 1 & 1 & 3 \\ 4 & 0 & 1 & 4 & 0 \\ 1 & 0 & 3 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

We only use sentences from *Corpus A.1* which have a transitive use of "is", e.g. "Anakin is a powerful Jedi" as opposed to "he is too powerful" in our calculation of C^{is}. An adjective A can be determined in the same fashion, so an adjective vector is computed to be the sum of the vectors of its arguments, e.g. $C^{\text{powerful}} = [1,2,2,1,1]$, or $C^{\text{brave}} = [7,1,2,5,0]$.

3.1 Representing a sentence as a vector.

Now consider the sentence at the start of *Corpus A.1*:

Palpatine is a mastermind who turns Anakin to the dark side of the Force.

Let us calculate a meaning vector for this sentence. The prepositional phrase "to the dark side of the Force" (abbreviated as "to DSOF") is represented as a vector and is given by the sum of the vectors of its arguments; $C^{\text{to DSOF}} = [3,0,1,0,0]$. The sentence "*Palpatine is a mastermind who turns Anakin to the dark side of the Force*" has the type assignment $n \quad n^r s n^l \quad n \quad n^r n s^l n \quad n^r s n^l \quad n \quad n^r n$, using the convention from [12] that the prepositional phrase "to the dark side of the Force" as a whole has the assignment $n^r n$. The sentence has a meaning vector of

$$\begin{aligned} &\overrightarrow{\text{Palpatine is a mastermind who turns Anakin to the dark side of the Force}} \\ &= \sum_{k,p,r,v,x} c_k^{\text{Palp}} c_{kpr}^{\text{is}} c_r^{\text{mm}} c_{rx}^{\text{turns}} c_v^{\text{Anakin}} c_{vx}^{\text{to DSOF}} \overrightarrow{s}_p \quad = \quad \sum_p (64 c_{2p1}^{\text{is}} + 8 c_{2p2}^{\text{is}}) \overrightarrow{s}_p. \end{aligned} \tag{$\clubsuit$}$$

This does not have much meaning, as S is an arbitrary vector space. If we set $S = N \otimes N$, then $\overrightarrow{s}_p = \overrightarrow{n}_i \otimes \overrightarrow{n}_j$. The verb matrix C^{is} is given above, thus $(\clubsuit) = 320 \overrightarrow{n}_2 \otimes \overrightarrow{n}_1 + 32 \overrightarrow{n}_2 \otimes \overrightarrow{n}_2$. This sum of tensor products only becomes meaningful when we are comparing sentences via an inner product on S, as we do next.

Consider the sentence "*Mace Windu is a mastermind who turns Anakin to the dark side of the Force*". When we compare this sentence to "*Palpatine is a mastermind who turns Anakin to the dark side of the Force*" we obtain a similarity score of $\frac{103424}{\sqrt{41066496 \cdot 103424}} = 0.53$. This is calculated by taking the inner product of the two sentences (103424) and dividing by the square root of the product of their lengths.

However, if we compare "*The Emperor* is a mastermind who turns Anakin to the dark side of the Force" to "Palpatine is a mastermind who turns Anakin to the dark side of the Force" we obtain a much higher similarity score of 0.99. Of course, as Palpatine is the Emperor, the similarity should be very high! A similar inner product with the sentence "*Padmé* is a mastermind who turns Anakin to the dark side of the Force" gives a similarity score of 0; as expected these sentences are not similar at all, as "Padmé" is very different to "Palpatine".

The vector space model of meaning has managed to extract these key themes from the corpus. Now our goal is to extract the same key ideas from an Irish corpus.

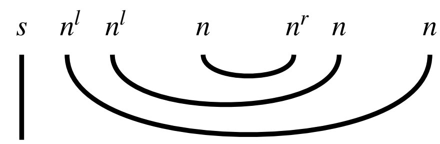
4 Bilingual Sentence Comparison via the Vector Space Model of Meaning

We shall now compare sentences between corpora in different languages. Our Irish vector space model of meaning shall be created from *Corpus B.1*, using the methods detailed in the previous section.

The calculations in *Section 3* & *Appendix C* require $S = N \otimes N$; to that end let the basis of N', the Irish noun space, be $\{$*Anakin, Palpatine, Jedi, Obi-Wan, arg-olc*$\}$, where "*arg-olc*" corresponds to the argument for the adjective *olc* - in English, 'evil'. This is also the collection of the five most commonly occurring nouns in *Corpus B.1* exactly (which might not really be a surprise as *Corpus B.1* is a translation of *Corpus A.1*, and nouns in English typically have one translation to Irish).

Take for example the sentence "*Palpatine is an evil Emperor*". In Irish, this is "*Is Impire olc é Palpatine*". The sentence has the type assignment and reduction diagram

Is	Impire	olc	é Palpatine
$sn^l n^l$	n	$n^r n$	n



corresponding to a map

$$f = (1_S \otimes \varepsilon_N) \circ (1_S \otimes 1_N \otimes \varepsilon_N \otimes 1_N) \circ (1_S \otimes 1_N \otimes 1_N \otimes \varepsilon_N \otimes 1_N \otimes 1_N).$$

Therefore the sentence "*Is Impire olc é Palpatine*" is assigned the following meaning vector:

$$\overrightarrow{\text{Is Impire olc é Palpatine}} = \sum_{ijk,p} c^{\text{Is}}_{ijk} c^{\text{Imp}}_{p} c^{\text{olc}}_{pk} c^{\text{Palp}}_{j} \overrightarrow{s}_i.$$

In order to evaluate this sentence, we need values for $c^{\text{Is}}_{ijk}, c^{\text{Imp}}_{p}, c^{\text{olc}}_{pk}$, and c^{Palp}_{j}. The adjectives and adverbs are calculated in the same way as in §3. We must be a little more careful with verbs. The copula *is* follows the rule *Is Object Subject*, and all other transitive verbs follow the rule *Verb Subject Object*, so we define

$$\overrightarrow{is} := \sum_{ij} \overrightarrow{subject}_i \otimes \overrightarrow{object}_j, \qquad \overrightarrow{verb} := \sum_{ij} \overrightarrow{object}_i \otimes \overrightarrow{subject}_j,$$

and also that

$$\overrightarrow{\text{is object subject}} := \overrightarrow{is} \odot (\overrightarrow{subject} \otimes \overrightarrow{object}),$$
$$\overrightarrow{\text{verb subject object}} := \overrightarrow{verb} \odot (\overrightarrow{object} \otimes \overrightarrow{subject}).$$

Note the copula is treated the same as in English[1]. Now, let us compare sentences between Irish and English. We obtain the following data from *Corpus B.1*. (Note "*Impire*" is "Emperor", "*Tiarna Sith*" is "Sith Lord", "*taobh dorcha na Fórsa*" is "dark side of the Force", and "*cróga*" is "brave".)

$$\begin{aligned} &\text{Impire} = [1,5,0,1,1], && \text{Yoda} = [0,1,2,3,0], \\ &\text{Palpatine} = [0,1,0,0,0] && \text{Tiarna Sith} = [1,0,0,0,2], \\ &\text{taobh dorcha na Fórsa} = [5,4,1,1,1], && \text{Padmé} = [5,0,0,1,0], \\ &C^{\text{olc}} = [2,5,2,1,3], && C^{\text{cróga}} = [7,1,2,4,0]. \end{aligned} \qquad C^{\text{Is}} = \begin{bmatrix} 6 & 0 & 1 & 0 & 1 \\ 6 & 4 & 1 & 0 & 2 \\ 2 & 0 & 2 & 0 & 2 \\ 0 & 0 & 3 & 0 & 0 \\ 3 & 0 & 0 & 0 & 2 \end{bmatrix}.$$

It is quite welcome that C^{Is} is different to the English C^{is}, as in Irish the verb "to be" is sometimes used in conjunction with another verb, which becomes the main transitive verb of the sentence. Thus, there are fewer occurrences of "*tá*" or "*is*" in *Corpus B.1* than "is" in *Corpus A.1*.

The result of our two assumptions (that $S = N \otimes N$ and the basis of N' is the exact translation of the basis of N) is we can meaningfully compare the following sentences. In the first instance,

$$\langle \text{Palpatine is an evil Emperor} \mid \textit{Is Impire olc é Palpatine} \rangle = 10174.$$

The length of the former is 10182 and the length of the latter is 10180, meaning the similarity score between the sentence "Palpatine is an evil Emperor" and its Irish translation "*Is Impire olc é Palpatine*" is $\frac{10174}{\sqrt{10182 \cdot 10180}} = 0.99$; very high. On the other hand, if we try to compare sentences that are not translates of one another, say "Yoda is a powerful Jedi" to "*Is Jedi cróga é Palpatine*" (in English, "Palpatine is a brave Jedi"), we receive low scores[2]:

$$\langle \text{Yoda is a powerful Jedi} \mid \textit{Is Jedi cróga é Palpatine} \rangle = 8.$$

The length of the former is 348 and the length of the latter is 4, so the similarity score between the sentence "Yoda is a powerful Jedi" and "*Is Jedi cróga é Palpatine*" is $\frac{8}{\sqrt{348 \cdot 4}} = 0.21$; relatively low.

The following table houses a collection of sentence pairs and their similarity scores, calculated according to this method. The reader is invited to see *Appendix C* for a more complicated sentence comparison.

English Sentence	Irish Translation	Similarity Score
Yoda is a powerful Jedi	*Is Jedi cumhachtach é Yoda*	0.94
Palpatine is an evil Emperor	*Is Impire olc é Palpatine*	0.99
A brave Padmé turns to Anakin	*Casann Padmé cróga chuig Anakin*	1
Obi Wan turns to the powerful Yoda	*Casann Obi-Wan go Yoda cumhachtach*	0.87
Padmé is a brave Jedi	*Is Jedi cróga é Padmé*	0.94
Anakin is a Sith Lord	*Is Tiarna Sith é Anakin*	0.32
The Jedi turn to the brave Mace Windu	*Casann na Jedi go Mace Windu cróga*	0.99

[1] One small caveat: "*is*" is translated to have the same meaning as the verb "to be" in English, which in Irish corresponds to the verb "*bí*", which in *Corpus B.1* is conjugated as "*tá*". The result? $C^{\text{Is}} = C^{\text{tá}}$ is calculated by including sentences with use of either "*Tá...*" or "*Is...*".

[2] Technically when calculating $C_{ij}^{\text{cróga}}$ - the matrix for the adjective 'brave' - we are also counting the various different translations of "brave" occurring in *Corpus B.1*, such as "*go crua*" or "*go láidir*".

The BLEU score, introduced in [21], provides a measure of how accurate a machine translation system is; in the below table we demonstrate how the BLEU score of two sentences with minor variations compares to the similarity score of the sentences. This score is calculated as follows: let R be a reference sentence and C be a candidate sentence. The goal is to measure how the *n-grams* (blocks of n words in succession) appear and compare in R and C. To that end, define

$$p_n = \frac{\sum_{\text{n-gram}\in C} \min\{\#\text{times n-gram appears in } C, \#\text{times n-gram appears in } R\}}{\sum_{\text{n-gram}\in C} \#\text{times n-gram appears in } C}.$$

We can also introduce a *brevity penalty* (BP) if the candidate is too short: this is $\exp(1-\frac{|R|}{|C|})$ if $|C| \leq |R|$ and 1 otherwise. Finally, the BLEU score $:= \text{BP}\cdot\exp\left(\sum_{n=1}^{4}\frac{1}{4}\log p_n\right)$. If there is no n-gram overlap between C and R for any order of n-grams ($1 \leq n \leq 4$), BLEU returns the value 0. The author used smoothing function 7 of [6] to avoid this harsh behaviour.

Reference Sentence	Candidate Sentence	Similarity Score	BLEU Score
Yoda is a powerful Jedi	Yoda turns to the powerful Jedi	0.95	0.32
Anakin is a Sith Lord	Obi-Wan is a Sith Lord	0	0.7
Is Impire olc é Palpatine	*Is Impire olc é Mace Windu*	0.98	0.7
Casann na Jedi go Mace Windu cumhachtach	*Casann Ginearál Grievous go Mace Windu cróga*	0.76	0.27

In [3] the authors work with conceptual spaces: instead of nouns being labelled relative to nouns they appear often with, instead nouns are represented by other words that describe them. The hope would be this removes instances of problematic translation, like in the sixth example of the first table. However, building on the ideas of Bolt et al. [3] and Gärdenfors [10] much work would need to be done to capture the intricacies presented here. In the next sections we will instead tackle a simpler example involving planets and fruit.

5 Word Classification

According to Dixon and Aikhenvald [9], "three word classes are . . . implicit in the structure of each human language: nouns, verbs and adjectives." It is the goal of this section to specify a treatment of nouns and adjectives for use in conceptual space creation. Once we have some sort of classification system for each of these, we can proceed with creating a conceptual space from a given corpus. For example, in the case of adjectives we wish to classify words such as 'heavy', 'red' or 'hot', and to each assign a numeric value that transcends language and thus can be compared across (say) Irish and English.

5.1 Adjectives

In their landmark work, Dixon and Aikhenvald [9] give a complete treatment of adjective classes as they arise in various languages across the globe, such as Japanese, Korean, Jarawara, Mam and Russian. In particular, they name seven core types of adjectives that consistently and naturally arise:

(1) **Dimension.** (big, small, short, tall, etc.)
(2) **Age.** (new, old, etc.)
(3) **Value.** (good, bad, necessary, expensive, etc.)
(4) **Colour.** (green, white, orange, etc.)
(5) **Physical Property.** (hard, hot, wet, etc.)
(6) **Human Propensity.** (happy, greedy, etc.)
(7) **Speed.** (fast, slow, etc.)

In this paper our focus will be representing nouns other than human beings in conceptual spaces, therefore we will not consider adjectives from item (6). Also, for the purposes of language translation, our focus will be on recreating a human's process of concept construction so the author proposes reframing some of these seven core adjective types from the perspective of our five senses; *sight, smell, sound, sensation and savour*. This paper does not produce examples involving *smell* or *sound*, so these categories have been omitted.

(1) **Dimension.**
(2) **Age.**
(3) **Value.**
(4) **Speed.**
(5) **Physical Property.** Further classified as:
   (a) Colour, Intensity (Sight)
   (b) Savour
   (c) Temperature, Density, Mass, Texture (Sensation)

How do we represent this data numerically? Fortunately most aspects of the five categories lend themselves to a linear interpretation. For example, in **Dimension** we can manually order adjectives in this class from 'small' to 'large' and represent **Dimension** as an interval $[0,1]$, assigning adjectives representing 'small' sizes values close to 0, and adjectives representing 'large' sizes values close to 1. The allocation of these values depends on the preferences of the programmer. This will not be extremely precise - nor, in fact, do we want it to be - by our very nature spaces visualised by humans are fuzzy, and our use of adjectives reflects this. Therefore while there are fuzzy boundaries between sizes within the class, overall there will be a distinction between 'small' and 'large', though how big that distinction is depends on the programmer inputting these values.

Similarly we allow **Age** to be represented by $[0,1]$ (where adjectives such as *young, new, baby* are valued closer to 0, and *old, mature, antiquated* are closer to 1), and **Value** and **Speed** to be represented by $[0,1]$ as well. We will take these spaces as given and assume one can preload a list of common adjectives with assigned $[0,1]$ values, in much the same way it is assumed one can preload a list of colours with assigned $[0,1]^3$ values in the common RGB colour cube. For **Physical Properties**,

(a) *Colour* will be represented numerically by the RGB colour cube, and Intensity by the interval $[0,1]$.

(b) *Savour* by Gärdenfors' taste tetrahedron (*see* [3, Fig. 2]). Embed this into $\mathbb{R}^3$ in the usual way: Salt $=[1,0,0]$, Sour $=[-\frac{1}{2},-\frac{\sqrt{3}}{2},0]$, Bitter $=[-\frac{1}{2},\frac{\sqrt{3}}{2},0]$, and Sweet $=[0,0,\sqrt{2}]$.

(c) *Sensation* the author suggests representing by a hypercube $[0,1]^4$ with the first dimension temperature (from low to high), the second dimension density (from low - e.g. *gaseous, wispy, fine*, to high - e.g. *solid, dense, hard*, with items like *soft, mushy, wet, gloopy, sticky, brittle, crumbly* in between), the third dimension mass (from light to heavy) and the fourth dimension texture (from smooth to rough).

This system cannot capture every type of adjective. At present this view is not sophisticated enough to capture 'dry', 'clear', 'sunny' etc. - however, this system does allow us to start analysing text in a meaningful way. Going forward, we shall assign numerical values to adjectives based on our intuition and assume a complex set of adjectives has been hard coded into our algorithm a priori. This may seem a little ad hoc, but it is how we learn adjectives in the early years of our lives: by repeated exposure and memorisation. Of course, our mental picture of objects comes not just from adjectives but also other nouns.

5.2 Nouns

The advantage to allowing nouns to classify other nouns is twofold; first, nouns can identify structure that adjectives might have missed. For example, describing apples and cars as "red, smooth and fresh smelling" might be accurate, but paints the wrong conceptual picture. The picture is corrected once we include the sentences "an apple is a fruit" and "a car is a vehicle". Such classifying words as 'fruit' or 'vehicle' are known as *hypernyms*; a word A is a hypernym of a word B if the sentence "B is a (kind of) A" is acceptable to English speakers. The converse, a *hyponym*, is defined as a word B such that the sentence "B is a (kind of) A" is acceptable. This brings us to the second advantage of allowing nouns into our classification system; like adjectives, they can be ordered (this time in a tree[3]) by the hypernym-hyponym relationship.

There is already a substantial amount of work done on classifying nouns by the hypernym-hyponym relationship, and there exist algorithms which extract this sort of structure from a given corpus [4, 15, 16, 22]. Hearst [16] in 1992 revolutionarily algorithmitised hypernym-hyponym relationships according to a certain set of English rules (which, incidentally, can be recreated for Irish). Caraballo [4] took this work further and produced a working example with the 'Wall Street Journal' Penn Treebank corpus [19].

As well as this, there already exists the knowledge base WordNet [26] and its Irish counterpart LSG (*Líonra Séimeantach na Gaeilge*) [24], both of which have organised thousands of nouns into this hierarchical relationship. Therefore we shall assume a hierarchy such as *food* $\to$ *fruit* $\to$ *berry* can already be extracted from text. Using these tools, we have the following options when making use of the hypernym-hyponym tree in concept creation:

(1) If we are interested solely in concept creation (i.e. are only concerned with concepts for *one* language) we can remove the dependency of the tree on the corpus being analysed by using WordNet to create a hypernym-hyponym tree. Relabelling the vertices gives us a convex space associated to each noun in the text via their path from root to leaf.

(2) If we are interested in using concepts for language translation the matter becomes trickier - the trees generated by WordNet and LSG might not have the same structure. However, if we assume we are given two copies of the same corpus, one in English and the other in Irish, then we can assume the *same* (up to synonyms[4]) hierarchy of nouns is produced in the corresponding languages, using the extraction algorithms created by Hearst and Caraballo ([16], [4], resp.).

The key point: given a corpus of text in English producing the hierarchy *food* $\to$ *fruit* $\to$ *berry*, we will assume the hierarchy *bia* $\to$ *torthaí* $\to$ *caora* produced by the Irish corpus is *directly comparable* to the English hierarchy, meaning we can instead label the hierarchy as $v_0 \to v_1 \to v_2$ and refer to berry (and *caora*) by its path in the hierarchy: $\{v_0, v_1, v_2\}$. In §6.1 there is an example of this proposal working successfully.

[3]It might not be technically correct to refer to the structure as a tree, as each word might have several hypernyms. Nevertheless, the terminology has stuck.

[4]Each node in the hypernym-hyponym tree is a 'synset' (a class of synonyms).

6 Conceptual Space Creation from a Corpus

In 2004 Gärdenfors [10] introduced *conceptual spaces* as a means of representing information in a 'human' way; the founding idea being if two objects belong to the same concept, then every object somehow 'in between' these objects also belongs to the same concept. We can mathematically describe the property of 'in between' via *convex algebras*, an introduction to which is given by Bolt et al. [3, §4]. The work in this section and the next is carried out in the category **ConvexRel**; the category with convex algebras as objects and convex relations as morphisms. The two convex algebras of interest to us are *Examples 1 & 6* of [3]:

(1) The closed real interval $[0,1]$ has a convex algebra structure induced by the vector space $\mathbb{R}$. The formal sums $\sum_i p_i|x_i\rangle$ are sums of elements in $[0,1]$ with addition and multiplication from $\mathbb{R}$. The mixing operation is the identity map.

(2) A finite tree can be a convex algebra - in particular, the hypernym-hyponym trees we are interested in are affine semilattices, hence the formal sums

$$\sum_i p_i|a_i\rangle := \bigvee_i \{a_i : p_i > 0\}$$

are well defined. (So, for example the formal sum $p_1|x_1\rangle + p_2|x_2\rangle + p_3|x_3\rangle$ is the lowest level in the tree containing x_1, x_2, x_3; their *join*.)

ConvexRel is compact closed [3, Theorem 1], hence (by Coecke et al. [7]) combines perfectly with the Lambek grammar category allowing us to create a functor to interpret meanings in the **ConvexRel** category via the type reductions in the Lambek grammar category.

The first hurdle we must overcome if we wish to use the **ConvexRel**-DisCoCat machinery is taking words in our foreign language (here Irish) and systematically representing them as convex spaces. The method we propose is reminiscent of how language is learnt in humans. For example, if a friend tells you an *úll* is a red, round, smooth, bitter or sweet fruit, you will (eventually, with enough information) come to understand they are describing an *apple*. It is in this vein of thought we present the following definition:

Definition 6.1. *A* descriptor *D of a noun N is an adjective or noun which aids in the description of N; if D is an adjective it describes physical properties of N (e.g. red, bitter, smooth) and if D is a noun it classifies N according to nouns in an already-known hierarchical structure (e.g* fruit*, belonging to food $\rightarrow$ fruit $\rightarrow$ berry).*

The basic idea of conceptual space creation we propose is as follows: given a corpus of text involving heavy use of a noun N, parse the text identifying descriptors of N. The example corpus, *Corpus D.1*, is quite simple so this parsing can be achieved by forming a collection of all words occurring in the same sentence as N, then sorting this collection into adjectives (which are represented as vectors according to §5.1) and nouns (which are represented by a tree according to §5.2). Taking the convex hull of the points in each adjective type, then the tensor product of the convex hulls, we represent the adjective descriptors of N as a convex set. Noun descriptors are represented as a convex set a là §5.2. Combining these convex subsets under a tensor product gives us a conceptual space, as required.

6.1 Example: Planets, the Sun and Fruit.

Consider *Corpus D.1* from *Appendix D*. Let us examine five main nouns from this corpus;

$$N^1 = \text{Venus},\ N^2 = \text{Jupiter},\ N^3 = \text{Mars},\ N^4 = \text{apple},\ N^5 = \text{The Sun}.$$

Organising this into a table we obtain:

Venus	**Adjectives**	solid, rocky, same size as Earth, hot, high pressure, bright.
	Nouns	planet, Earth's sister, ball.
Jupiter	**Adjectives**	very large, gassy, orange, brown, red, far away, windy, freezing, very bright.
	Nouns	planet, ball, outer space.
Mars	**Adjectives**	very red, brown, orange, cold, smaller than Earth, rocky.
	Nouns	planet, outer space.
Apple	**Adjectives**	round, soft, red, green, bitter, sweet.
	Nouns	fruit, ball.
The Sun	**Adjectives**	brightest, huge, very hot, round, very dense.
	Nouns	star, ball.

We first deal with the adjectives. These can be organised according to *Section 5.1*:
(1) **Venus.**

$$\begin{aligned}
&N_{\text{dimension}} = \mathit{Conv}(\text{same size as Earth}) = \{0.5\},\\
&N_{\text{intensity}} = \mathit{Conv}(\text{bright}) = \{0.7\},\\
&N_{\text{temperature}} = \mathit{Conv}(\text{hot}) = \{0.75\},\\
&N_{\text{density}} = \mathit{Conv}(\text{solid}) = \{0.9\},\\
&N_{\text{texture}} = \mathit{Conv}(\text{rocky}) = \{0.9\}.
\end{aligned}$$

(These values were assigned according to the author's own preference, however they can be assigned different values according to each readers' wishes.) D^1_{adj} is the tensor product of these spaces, where if an adjective class is not mentioned, its corresponding noun space (or *property*) is set to $[0,1]$ (e.g. $N_{\text{age}} = [0,1]$). Note that we were required to drop some adjectives, such as *high pressure*, as our adjective classification from *Section 5.1* is not specific enough to capture all details. Also note that it is also unusual that these spaces are singleton sets; in a larger, more complicated corpus these properties would be intervals.

(2) **Jupiter.**

$$\begin{aligned}
&N_{\text{dimension}} = \mathit{Conv}(\text{very large}) = \{0.7\},\\
&N_{\text{colour}} = \mathit{Conv}(\text{orange} \cup \text{brown} \cup \text{red})^5,\\
&N_{\text{intensity}} = \mathit{Conv}(\text{very bright}) = \{0.8\},\\
&N_{\text{temperature}} = \mathit{Conv}(\text{freezing}) = \{0\},\\
&N_{\text{density}} = \mathit{Conv}(\text{gassy}) = \{0.1\}.
\end{aligned}$$

(3) **Mars.**

$$\begin{aligned}
&N_{\text{dimension}} = \mathit{Conv}(\text{smaller than Earth}) = \{0.25\},\\
&N_{\text{colour}} = \mathit{Conv}(\text{red} \cup \text{brown} \cup \text{orange}),\\
&N_{\text{temperature}} = \mathit{Conv}(\text{cold}) = \{0.4\},\\
&N_{\text{texture}} = \mathit{Conv}(\text{rocky}) = \{0.9\}.
\end{aligned}$$

(4) **Apple.**

$$\begin{aligned}
&N_{\text{colour}} = \mathit{Conv}(\text{red} \cup \text{green}),\\
&N_{\text{taste}} = \mathit{Conv}(\text{bitter} \cup \text{sweet}),\\
&N_{\text{texture}} = \mathit{Conv}(\text{soft}) = \{0.4\}.
\end{aligned}$$

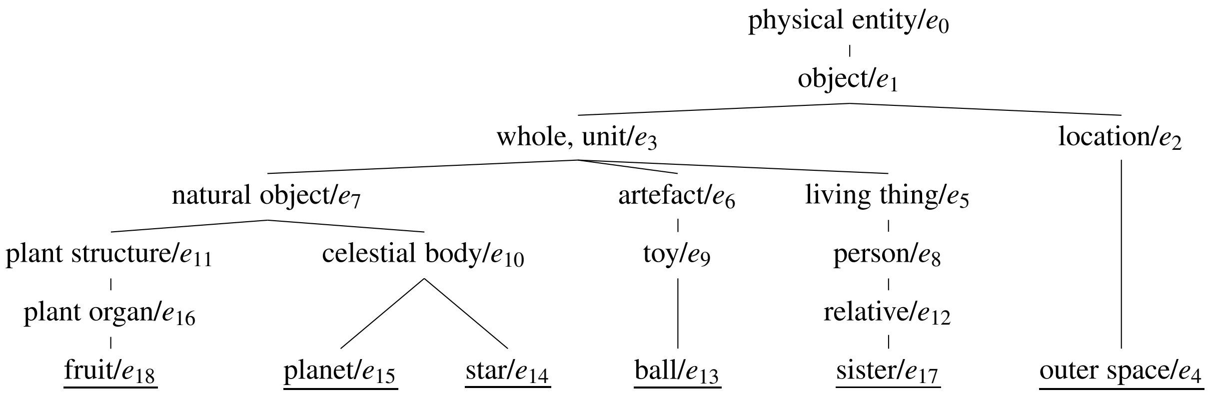
(5) **The Sun.**

$$\begin{aligned}
&N_{\text{dimension}} = \mathit{Conv}(\text{huge}) = \{1\},\\
&N_{\text{intensity}} = \mathit{Conv}(\text{brightest}) = \{1\},\\
&N_{\text{temperature}} = \mathit{Conv}(\text{very hot}) = \{1\},\\
&N_{\text{density}} = \mathit{Conv}(\text{very dense}) = \{1\}.
\end{aligned}$$

[5] The RGB values we use for *orange* and *brown* are (255,165,0) and (153,76,0) respectively.

For $K = 2,3,4,5$, D^K_{adj} is the tensor product of the $N_{(-)}$ spaces of item (K). The additional information of the noun interdependence is added by the following tree, generated by WordNet [26]: (see the tree on the following page).

The underlined leaves are those nouns appearing as descriptors in *Corpus D.1*. Referring to each node by its e_i label, we can define the sets D^i_{noun}.



$$\begin{aligned}
D^1_{\text{noun}} &:= \{e_0,e_1,e_3,e_5,e_6,e_7,e_8,e_9,e_{10},e_{12},e_{13},e_{15},e_{17}\}, && \text{Venus} := D^1_{\text{adj}} \otimes D^1_{\text{noun}},\\
D^2_{\text{noun}} &:= \{e_0,e_1,e_2,e_3,e_4,e_6,e_7,e_9,e_{10},e_{13},e_{15}\}, && \text{Jupiter} := D^2_{\text{adj}} \otimes D^2_{\text{noun}},\\
D^3_{\text{noun}} &:= \{e_0,e_1,e_2,e_3,e_4,e_7,e_{10},e_{15}\}, && \text{Mars} := D^3_{\text{adj}} \otimes D^3_{\text{noun}},\\
D^4_{\text{noun}} &:= \{e_0,e_1,e_3,e_6,e_7,e_9,e_{11},e_{13},e_{16},e_{18}\}, && \text{Apple} := D^4_{\text{adj}} \otimes D^4_{\text{noun}},\\
D^5_{\text{noun}} &:= \{e_0,e_1,e_3,e_6,e_7,e_9,e_{10},e_{13},e_{14}\}, && \text{The Sun} := D^5_{\text{adj}} \otimes D^5_{\text{noun}}.
\end{aligned}$$

…and finally we obtain the conceptual spaces for the 5 nouns, on the right.

In Irish, the same corpus is *Corpus D.2*, located in *Appendix D*. The five main nouns of this corpus are (in no particular order) $M^1 = $ *Véineas*, $M^2 = $ *Iúpatar*, $M^3 = $ *Mars*, $M^4 = $ *Úll*, $M^5 = $ *Grian*.

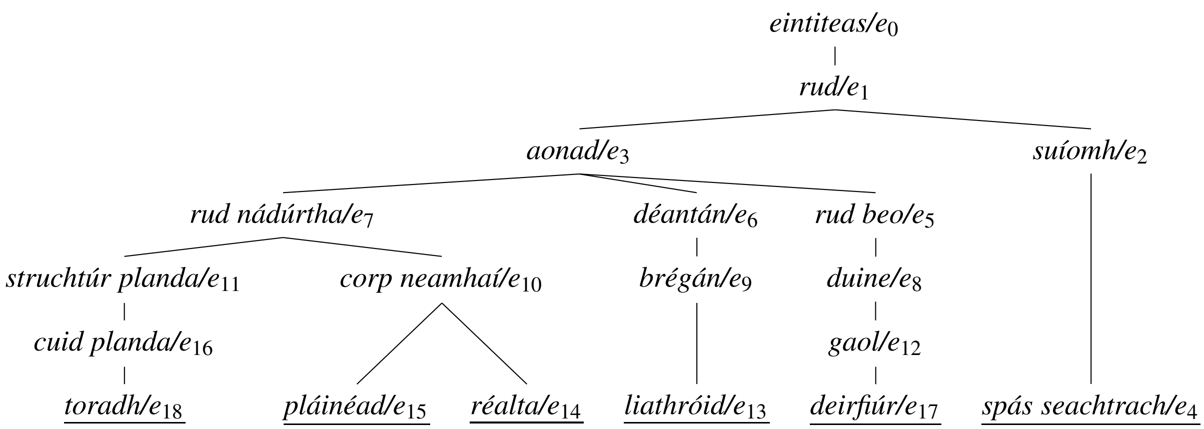
We can once again organise the information of *Corpus D.2* into a table (see *Appendix E*) and determine noun spaces $N_{(-)}$ from adjectives as was explained in §5.1. For example, in the case of *Véineas:*

$$\begin{aligned}
N_{\text{dimension}} &= Conv(\textit{méid céanna leis an Domhan}) = \{0.5\},\\
N_{\text{intensity}} &= Conv(\textit{geal}) = \{0.6\},\\
N_{\text{temperature}} &= Conv(\textit{an-te}) = \{0.85\},\\
N_{\text{density}} &= Conv(\textit{tathagach}) = \{0.9\},\\
N_{\text{texture}} &= Conv(\textit{carraigeach}) = \{0.9\}.
\end{aligned}$$

Note that the values here are different than the corresponding values in English for *geal* (bright), *an-te* (hot), etc. The reasoning here is as follows: in Irish there is no word for "hot" - to describe high temperatures there is just "warm" and "very warm". So "*an-te*" ("very warm") suffices for "hot", therefore since "*an-te*" is the hottest the weather can be described, it is assigned a value of 0.85 in Irish (because in English, "very hot" would need to correspond to a higher value than "hot", which is 0.75).

$\overline{D}^1_{\text{adj}}$ we define to be the tensor product of the $N_{(-)}$ spaces of *Véineas*. Note that we were required to drop some adjectives, such as *brú… ard* (*high pressure*), as our adjective classification from §5.1 is not specific enough to capture all details.

The additional linguistic information from the descriptor nouns is obtained by referencing a hypernym-hyponym tree, which for example WordNet (in Irish) organises as:



If we label the tree counterpart to the English tree, we can finally define the conceptual spaces for the 5 main nouns.

$$\overline{D}^1_{\text{noun}} := \{e_0, e_1, e_3, e_5, e_6, e_7, e_8, e_9, e_{10}, e_{12}, e_{13}, e_{15}, e_{17}\}, \qquad \textit{Véineas} := \overline{D}^1_{\text{adj}} \otimes \overline{D}^1_{\text{noun}},$$
$$\overline{D}^2_{\text{noun}} := \{e_0, e_1, e_2, e_3, e_4, e_6, e_7, e_9, e_{10}, e_{13}, e_{15}\}, \qquad \textit{Iúpatar} := \overline{D}^2_{\text{adj}} \otimes \overline{D}^2_{\text{noun}},$$
$$\overline{D}^3_{\text{noun}} := \{e_0, e_1, e_2, e_3, e_4, e_7, e_{10}, e_{15}\}, \qquad \textit{Mars} := \overline{D}^3_{\text{adj}} \otimes \overline{D}^3_{\text{noun}},$$
$$\overline{D}^4_{\text{noun}} := \{e_0, e_1, e_3, e_6, e_7, e_9, e_{11}, e_{13}, e_{16}, e_{18}\}, \qquad \textit{Úll} := \overline{D}^4_{\text{adj}} \otimes \overline{D}^4_{\text{noun}},$$
$$\overline{D}^5_{\text{noun}} := \{e_0, e_1, e_3, e_6, e_7, e_9, e_{10}, e_{13}, e_{14}\}. \qquad \textit{Grian} := \overline{D}^5_{\text{adj}} \otimes \overline{D}^5_{\text{noun}}.$$

7 Metrics for concepts

Our final goal is to compare the concepts created in §6.1 in Irish and English. To do this we require some measure of distance between concepts; we require a metric on **ConvexRel**. We will define the metric d then leave the technical details concerning the combination of convex structures and metrics to [20] (cf. *Example 28*, ibid.).

Definition 7.1. ***(Hausdorff Metric).*** *Let X, Y be two nonempty subsets of a metric space (M, f). Define their* Hausdorff distance *to be*

$$d(X,Y) := \max\{\sup_{x\in X}\inf_{y\in Y} f(x,y),\ \sup_{y\in Y}\inf_{x\in X} f(x,y)\}.$$

In the case of **ConvexRel**, all the concepts we define are subsets of $\mathbb{R}^{20} \otimes E$ where $E = \{e_0, \ldots, e_n\}$ represents a tree, according to §5.2. On $\mathbb{R}^{20}$ there is the taxicab metric and for E there is the metric f on $\mathscr{P}(E)$, the power set of E, defined by

$$\text{for } A, B \subseteq E, \quad f(A,B) := \max\{|A \setminus B|,\ |B \setminus A|\}.$$

Example 7.2. Consider the distance between "Apple" and "Jupiter", whose conceptual spaces were calculated in §6.1.

$$
\begin{aligned}
d(\text{“Apple”}, \text{“Jupiter”}) \quad &= \quad \max\{ \sup_{x\in\text{Apple}} \inf_{y\in\text{Jupiter}} f(x,y), \ \sup_{y\in\text{Jupiter}} \inf_{x\in\text{Apple}} f(x,y)\} \\
&= \max\{ \sup_{x\in\text{Apple}} \inf_{y\in\text{Jupiter}} (f(N^{\text{Apple}}_{\text{dimension}}, N^{\text{Jupiter}}_{\text{dimension}}) + f(N^{\text{Apple}}_{\text{colour}}, N^{\text{Jupiter}}_{\text{colour}}) + \cdots + f(D^{\text{Apple}}_{\text{noun}}, \overline{D}^{\text{Jupiter}}_{\text{noun}})), \ \inf_{x\in\text{Apple}} \sup_{y\in\text{Jupiter}} \ldots \} \\
&= \max\{ \sup_{x\in\text{Apple}} \inf_{y\in\text{Jupiter}} (f([0,1],\{0.7\}) + f(\mathit{Conv}(\text{red}\cup\text{green}), \mathit{Conv}(\text{orange}\cup\text{brown}\cup\text{red}) + \ldots \\
&\qquad + f([0,1],\{0.1\}) + f(\{e_0,e_1,e_3,e_6,e_7,e_9,e_{11},e_{13},e_{16},e_{18}\}, \{e_0,e_1,e_2,e_3,e_4,e_6,e_7,e_9,e_{10},e_{13},e_{15}\}), \ldots \} \\
&= 8.7.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
d(\text{“Mars”}, \text{“Jupiter”}) &= 5.55, \\
d(\text{“Jupiter”}, \text{“Sun”}) &= 7.7, \\
d(\text{“Apple”}, \text{“Sun”}) &= 7.97.
\end{aligned}
$$

This seems to capture the rough picture we desire: relatively speaking[6], the planets Mars and Jupiter are close, while nouns like “Apple” and “Jupiter” or “Apple” and “Sun” are distant. “Sun” is also technically closer to “Jupiter” than to “Apple”, though not by much. One might expect “Jupiter” to be closer to “Sun”, however this is not the picture *Corpus D.1* paints; in it, the Sun is not a planet and is described as “very hot” or “very dense”. Perhaps if the corpus noted the Sun is ‘planet-like’ or ‘round like a planet’, and described the colour of the sun as ‘yellow and orange’, the distance between these two concepts might be smaller.

Finally, let us return to translation between Irish and English. Using the same metric, the distance between “Apple” and its Irish translation, “*Úll*”, is given by $d(\text{“Apple”}, \text{“}\textit{Úll}\text{”}) = 0$, which is to say as concepts, “Apple” and “*Úll*” are equal. On the other hand, the distance between “Apple” and “*Grian*” (English: “Sun”) is $d(\text{“Apple”}, \text{“}\textit{Grian}\text{”}) = 7.97$.

If we were to attempt to translate “*Iúpatar*” using the metric on **ConvexRel**, we see

$$
\begin{aligned}
d(\text{“Venus”}, \text{“}\textit{Iúpatar}\text{”}) &= 8.75, \\
d(\text{“Jupiter”}, \text{“}\textit{Iúpatar}\text{”}) &= 0.3,^{7} \\
d(\text{“Mars”}, \text{“}\textit{Iúpatar}\text{”}) &= 5.45.
\end{aligned}
\qquad
\begin{aligned}
d(\text{“Apple”}, \text{“}\textit{Iúpatar}\text{”}) &= 8.6, \\
d(\text{“Sun”}, \text{“}\textit{Iúpatar}\text{”}) &= 7.6,
\end{aligned}
$$

Hence choosing the concept closest to “*Iúpatar*”, which is “Jupiter”, we deduce we have indeed successfully translated this word.

Remark 7.3. It is the opinion of the author that the exercise of manually inputting values for the seven core adjective types is an important, maybe even necessary, one. This method is how we first master colours and smells and sizes; by hearing about them and memorising terms, ordered relative to each other. In the words of Gärdenfors [11], “we are not born with our concepts; they must be learned”. The author believes it is also necessary to preform this exercise separately for Irish, as adjectives in this language can have different emphases and occasionally different meanings. ◊

[6] Relative to the other distances calculated in this example.

[7] That this is non-zero stems from the fact that adjectives can have different meanings with different intensities in different languages.

8 Conclusion

This paper has outlined two methods of translating from Irish to English using the distributional compositional categorical model of meaning; via vector spaces and the category **FVect** and via concepts and the category **ConvexRel**. The former allowed us to compare the meanings of sentences between languages by calculating similarity scores, and the latter allowed us to focus more on the meaning behind nouns and calculate distances between concepts across languages.

The work of this paper can be extended in many ways. In *Section 2*, the Lambek pregroup grammar structure for Irish can be further embellished and more grammatical features of Irish captured as Lambek does for English [17]. The ideas behind *Section 5* can also be expanded to address adjectives not captured by the system presented in *Section 5.1* and *Section 6*, titled 'Conceptual Space Creation from a Corpus', could also be computationally tested with larger corpora. In particular, a more precise explanation and demonstration of how the descriptors of a noun N are identified and sorted for any large corpus could be further addressed.

Ongoing work includes determining a treatment for quantification and negation, often important for language translation. Also, perhaps most importantly, the setting proposed in this paper must be implemented, evaluated, and experimented with to produce a useful tool for understanding Irish.

Acknowledgements

The author would like to extend his thanks to Professor Coecke and Dr. Marsden for their help in this topic throughout the year. The author's thanks to Kevin Scannell for his communication and for providing the source files to the LSG. The author would also like to thank Abby Breathnach and Soinbhe Nic Dhonncha for their help in translating key examples in this paper to Irish. Finally, the author thanks Martha Lewis and the CAPNS referees for their helpful comments and suggestions in their review of this paper.

References

[1] D. Bargelli & J. Lambek (2001): *An Algebraic Approach To French Sentence Structure*. In P. de Groote, G. Morrill & C. Retoré, editors: *Lecture Notes in Computer Science*, 2099, International Conference on Logical Aspects of Computational Linguistics, Springer, Berlin, Heidelberg, pp. 62–78, doi:10.1007/3-540-48199-0_4.

[2] D. Bargelli & J. Lambek (2003): *An Algebraic Approach To Arabic Sentence Structure*. *Linguistic Analysis* 31(3-4), pp. 301–315.

[3] J. Bolt, B. Coecke, F. Genovese, M. Lewis, D. Marsden & R. Piedeleu (2016): *Interacting Conceptual Spaces*. In D. Kartsaklis et. al., editor: *Proceedings of the 2016 Workshop on Semantic Spaces at the Intersection of NLP, Physics and Cognitive Science*, pp. 11–19, doi:10.4204/EPTCS.221.

[4] S. Caraballo (1999): *Automatic construction of a hypernym-labeled noun hierarchy from text*. In R. Dale & K. Church, editors: *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, Association for Computational Linguistics, pp. 120–126, doi:10.3115/1034678.1034705.

[5] C. Casadio & J. Lambek (2005): *A Computational Algebraic Approach to Latin grammar*. *Research on Language and Computation* 3(1), doi:10.1007/s11168-005-1286-0.

[6] B. Chen & C. Cherry (2014): *A Systematic Comparison of Smoothing Techniques for Sentence-Level BLEU*. In O. Bojar et. al., editor: *Proceedings of the Ninth Workshop on Statistical Machine Translation*, Association for Computational Linguistics, pp. 362–367, doi:10.3115/v1/w14-3346.

[7] B. Coecke, M. Sadrzadeh & S. Clark (2010): *Mathematical foundations for a compositional distributional model of meaning*. Available at `https://arxiv.org/abs/1003.4394`.

[8] J. Derrac & S. Schockaert (2015): *Inducing Semantic Relations from Conceptual Spaces: A Data-Driven Approach to Plausible Reasoning*. *Artificial Intelligence* 228, doi:10.1016/j.artint.2015.07.002.

[9] R. M. Dixon & A. Y. Aikhenvald, editors (2004): *Adjective Classes: A Cross-Linguistic Typology*, first edition. Oxford University Press. ISBN: 9780199270934.

[10] P. Gärdenfors (2004): *Conceptual Spaces: The Geometry of Thought*, paperback edition. The MIT Press. ISBN: 9780262572194.

[11] P. Gärdenfors (2014): *The Geometry of Meaning: Semantics based on Conceptual Spaces*, hardcover edition. The MIT Press. ISBN: 9780262026789.

[12] E. Grefenstette & M. Sadrzadeh (2011): *Experimental support for a categorical compositional distributional model of meaning*. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1394–1404. ISBN: 978-1-937284-11-4.

[13] E. Grefenstette & M. Sadrzadeh (2015): *Concrete Models and Empirical Evaluations for the Categorical Compositional Distributional Model of Meaning*. *Computational Linguistics* 41(1), doi:10.1162/COLI_a_00209.

[14] E. Grefenstette, M. Sadrzadeh, S. Clark, B. Coecke & S. Pulman (2014): *Concrete Sentence Spaces for Compositional Distributional Models of Meaning*. In H. Bunt et. al., editor: *Computing Meaning*, Text, Speech and Language Technology, vol 47, pp. 71–86, doi:10.1007/978-94-007-7284-7_5. Available at `https://arxiv.org/abs/1101.0309`.

[15] A. Gruenstein (2001): *Learning Hypernyms from Corpora*. Published online. Available at `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.13.4605`.

[16] M. Hearst (1992): *Automatic Acquisition of Hyponyms from Large Text Corpora*. In A. Zampolli, editor: *Proceedings of the Fourteenth International Conference on Computational Linguistics*, 2, Association for Computational Linguistics, pp. 539–545, doi:10.3115/992133.992154.

[17] J. Lambek (2008): *From Word to Sentence: a Computational Algebraic Approach to Grammar*, first edition. Polimetrica. ISBN: 9788876991172.

[18] L. Lambek & A. Preller (2004): *An Algebraic Approach to the German Sentence*. *Linguistic Analysis* 31(3).

[19] M. P. Marcus, M. A. Marcinkiewicz & B. Santorini (1993): *Building a large annotated corpus of English: The Penn Treebank*. *Computational Linguistics* 19(2), doi:10.21236/ada273556.

[20] D. Marsden & F. Genovese (2017): *Custom Hypergraph Categories via Generalized Relations*. In F. Bonchi & B. König, editors: *7th Conference on Algebra and Coalgebra in Computer Science (CALCO 2017)*, 72, pp. 17:1–17:16, doi:10.4230/LIPIcs.CALCO.2017.17.

[21] K. Papineni, S. Roukos, T. Ward & W. Zhu (2002): *BLEU: a Method for Automatic Evaluation of Machine Translation*. In P. Isabelle, editor: *Proceedings of the 40th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, Association for Computational Linguistics, pp. 311–318, doi:10.3115/1073083.1073135. Available at `http://www.aclweb.org/anthology/P02-1040.pdf`.

[22] E. Riloff & J. Shepherd (1997): *A Corpus-Based Approach for Building Semantic Lexicons*. In: *Second Conference on Empirical Methods in Natural Language Processing, EMNLP 1997*. Available at `https://arxiv.org/abs/cmp-lg/9706013`.

[23] M. Sadrzadeh, S. Clark & B. Coecke (2013): *The Frobenius anatomy of word meanings I: subject and object relative pronouns*. *Journal of Logic and Computation* 23(6), doi:10.1093/logcom/ext044. Available at `http://arxiv.org/abs/1404.5278v1`.

[24] K. Scannell (2006): *Líonra Séimeantach na Gaeilge*. Available at `http://borel.slu.edu/lsg/`.

[25] Wikipedia (2018): *Star Wars: Episode III - Revenge of the Sith*. Available at `https://bit.ly/2a7ZSYJ`. Accessed 20/03/2018.

[26] Wordnet (2018): *Princeton University "About WordNet"*. Available at `http://wordnet.princeton.edu`.

Appendices

A Corpus for Vector Space Model of Meaning (English)

The following is a summary of *Star Wars: Episode III - Revenge of the Sith*, obtained from Wikipedia [25] and edited by the author. Note that we are making some assumptions in using this corpus. The author is assuming the model of meaning can understand third-person sentences as if they were first-person sentences; i.e. "she is pregnant" is understood to be "Padmé is pregnant". We are also assuming the model can understand sentences with conjunction; e.g. "Anakin and Obi-Wan are known for their bravery" is "Anakin is known for his bravery" and "Obi-Wan is known for his bravery". We assume the model can understand the use of the present participle, i.e. "After infiltrating General Grevious' flagship" is understood to be "After Anakin and Obi-Wan infiltrate General Grevious' flagship". Finally we also assume the corpus has been lemmatised for *Sections 3 & 4*.

It is true that some of these assumptions might be difficult to work into the vector space model of meaning, however the author feels the use of this corpus gives good examples in *Sections 3 & 4* while still being interesting for humans to parse. *Corpora A.1 & B.1* can be rewritten such that the above assumptions are no longer necessary, however the story becomes tedious to read.

Corpus A.1. *Palpatine is a mastermind who turns Anakin to the dark side of the Force.*

The galaxy is in a state of civil war. Jedi Knights Obi-Wan Kenobi and Anakin Skywalker lead a mission to rescue the kidnapped Supreme Chancellor Palpatine from the evil General Grievous, who is a Seperatist commander. Anakin and Obi-Wan are known for their bravery and skill. After infiltrating General Grievous's flagship, the Jedi duel Dooku, whom Anakin eventually executes at Palpatine's urging. General Grievous escapes the battle-torn cruiser, in which the Jedi crash-land on Coruscant. There Anakin reunites with his beautiful wife, Padmé Amidala, who reveals that she is pregnant. While initially excited, the prophetic visions that Anakin has cause him to worry. Anakin believes Padmé will die in childbirth.

Palpatine appoints Anakin to the Jedi Council as his representative. The Jedi do not trust Palpatine as they believe he is too powerful. The Council orders Anakin to spy on Palpatine, his friend. Anakin begins to turn away from the Jedi because of this. Meanwhile the Jedi are searching for a Sith Lord. A Sith Lord is evil person who uses the dark side of the Force. The Jedi try prevent anyone from turning to the dark side of the Force and to evil. Palpatine tempts Anakin with secret knowledge of the dark side of the Force, including the power to save his loved ones from dying. Meanwhile, the powerful Obi-Wan travels to confront General Grievous. The Jedi and General Grievous duel and Obi-Wan fights bravely. Obi-Wan wins his duel against General Grievous. The Jedi Yoda travels to Kashyyyk to defend the planet from invasion. The mastermind Palpatine eventually reveals that he is a powerful Sith Lord to Anakin. Palpatine claims only he has the knowledge to save Padmé from death. Anakin turns away from Palpatine and reports Palpatine's evil to the Jedi Mace Windu. Mace Windu then bravely confronts Palpatine, severely disfiguring him in the process. Fearing that he will lose Padmé, Anakin intervenes. Anakin is a powerful Jedi and he severs Mace Windu's hand. This distraction allows Palpatine to throw Mace Windu *out of a window to his death. Anakin turns himself to the dark side of the Force and to Palpatine, who dubs him Darth Vader. Palpatine issues Order 66 for the clone troopers to kill the remaining Jedi, then dispatches Anakin with a band of clones to kill everyone in the Jedi Temple. Anakin ventures to Mustafar and massacres the remaining Separatist leaders hiding on the volcanic planet, while Palpatine addresses the Galactic Senate. He transforms the Republic into the Galactic Empire and declares himself Emperor Palpatine.*

Obi-Wan and Yoda return to Coruscant and learn of Anakin's betrayal against them. Obi-Wan leaves to talk to Padmé. Obi-Wan tries to convince her that Anakin has turned to the dark side of the Force; that Anakin has turned to evil. A brave Padmé travels to Mustafar and implores Anakin to abandon the dark side of the Force. Anakin refuses to stop using the dark side of the Force and sees Obi-Wan hiding on Padmés ship. Anakin angrily chokes Padmé into unconsciousness. Obi-Wan duels and defeats Anakin. Obi-Wan severs both of his legs and leaves him at the bank of a lava river where he is horribly burned. Yoda duels Emperor Palpatine on Coruscant until their battle reaches a stalemate. Yoda is a powerful Jedi, but he cannot defeat the evil Emperor Palpatine. Yoda then flees with Bail Organa while Palpatine travels to Mustafar. Evil Emperor Palpatine uses the dark side of the Force to sense Anakin is in danger.

Obi-Wan turns to Yoda to regroup. Padmé gives birth to a twin son and daughter whom she names Luke and Leia. Padmé dies of sadness shortly after. Palpatine finds a horribly burnt Anakin still alive on Mustafar. After returning to Coruscant, Anakins mutilated body is treated and covered in a black armored suit. Palpatine lies to Anakin that he killed Padmé in his rage. Palpatine is an evil mastermind and leaves Anakin feeling devastated. Palpatine has won; the dark side of the Force now flows through Anakin. Meanwhile, Obi-Wan and Yoda work to conceal the twins from the dark side of the Force, because the twins are the galaxy's only hope for freedom. Yoda exiles himself to the planet Dagobah, while Anakin and the Emperor Palpatine oversee the construction of the Death Star. Bail Organa adopts Leia and takes her to Alderaan. Obi-Wan travels with Luke to Tatooine. There Obi-Wan intends to bravely watch over Luke and his step-family until the time is right to challenge the Empire. □

B Corpus for Vector Space Model of Meaning (Irish)

For the sake of completeness we give the full Irish corpus whose translated meaning replicates *Corpus A.1*.

Corpus B.1. *Is ceannmáistir a casann Anakin go taobh dorcha na Fórsa é Palpatine.*

Tá an réaltra i stát cogaidh shibhialta. Rinne Ridirí Jedi Obi-Wan Kenobi agus Anakin Skywalker misean chun an Seansailéir Uachtarach Palpatine a shábháil ón Ginearál Grievous olc, ceannasaí Seperatist é. Aithnítear Anakin agus Obi-Wan dá a gcrógacht agus dá scileanna. Tar éis longcheannais Ginearál Grievous a ionsíothláit, troid na Jedi le Dooku, a mhoraíonn Anakin ar deireadh thiar ar mholadh Palpatine. Éalaíonn an Ginearál Grievous ón t-éadromaire caithe, ina dturlingíonn na Jedi chun talamh Coruscant. Ansin, tagann Anakin le chéile lena bhean álainn, Padmé Amidala, a léiríonn go bhfuil sí ag iompar clainne. Cé go bhfuil Anakin ar bís ar dtús, tugann a fhíseanna fáidhiúla cúis imní dó. Creideann Anakin go gheobhaidh Padmé bás i mbreithe clainne.

Ceapann Palpatine Anakin chuig Chomhairle na Jedi mar ionadaí. Níl muinín ag na Jedi i Palpatine mar a chreideann siad go bhfuil sé ró-chumhachtach. D'ordaíonn an Chomhairle Anakin a dhéanann spiaireacht ar Palpatine, a chara. Casann Anakin as an Jedi as seo. Idir an dá linn tá na Jedi ag cuardach do Tiarna Sith. Is duine olc é Tiarna Sith a úsáideann an taobh dorcha den Fhórsa. Déanann na Jedi iarracht a chur ar dhuine ar bith a bheith ag casadh go taobh dorcha na Fórsa agus go holc. Tacaíonn

Palpatine Anakin le heolas rúnda ar thaobh dorcha na Fórsa, lena n-áirítear an chumhacht chun a mhuintir a shábháil ó bhás. Idir an dá linn, téann Obi-Wan cumhachtach chun dul i ngleic leis an Ginearál Grievous. Troideann an Jedi agus Ginearál Grievous agus tá Obi-Wan ag troid go crua. Buaileann Obi-Wan a chath i gcoinne Ginearál Grievous. Téann Jedi Yoda go Kashyyyk chun an phláinéid a chosaint ó ionradh. Léiríonn an ceannmáistir Palpatine sa deireadh gurb é Tiarna cumhachtach Sith é go Anakin. Éilíonn Palpatine ach go bhfuil eolas air amháin Padmé a shábháil ón mbás. Casann Anakin i gcoinne Palpatine agus tuairiscíonn sé olc Palpatine chuig an Jedi Mace Windu. Tabhair Mace Windu aghaidh cróga ar Palpatine, agus é a dhíshealbhú go mór sa phróiseas. Ag eagla go gcaillfidh sé Padmé, idirghabhann Anakin. Is Jedi cumhachtach é Anakin agus sealaíonn sé lámh Mace Windu. Tugann an t-imréiteach seo do Palpatine Mace Windu a chaitheamh as fuinneog go dtí a bhás. Casann Anakin féin go taobh dhorcha na Fórsa agus chuig Palpatine, a ainm Darth Vader dó. Eisíonn Palpatine Ordú 66 do na trúpaí clón chun na Jedi atá fágtha a mharú, agus ansin cuireann sé Anakin le banna cluainé chuig an Teampaill Jedi a chuir bás ar gach duine. Taistilíonn Anakin go Mustafar agus maisíonn na ceannairí Separatist atá fágtha i bhfolach ar an phláinéid bólcanach, agus tugann Palpatine aitheasc don Seanad Réaltrach. Athraíonn sé an Poblacht isteach sa Impireacht Réaltrach agus dearbhaíonn sé féin an t-Impire Palpatine.

Fágann Obi-Wan agus Yoda go Coruscant agus foghlaimíonn siad bradú Anakin i gcoinne iad. Fágann Obi-Wan chun labhairt le Padmé. Déanann Obi-Wan iarracht a chur ina luí di go bhfuil Anakin tar éis casadh go taobh dorcha na Fórsa; go bhfuil Anakin tar éis casadh go holc. Taistealaíonn Padmé cróga go Mustafar agus cuireann sí ar Anakin an taobh dorcha den Fhórsa a thréigean. Diúltaíonn Anakin gan stop a bhaint as an taobh dorcha den Fhórsa agus feiceann sé Obi-Wan i bhfolach ar long Padmé. Tachtaíonn Anakin Padmé feargach go neamhfhiosach. Troideann Obi-Wan Anakin agus buann sé. Freastalaíonn Obi-Wan dá chuid cosa agus fágann sé é i mbruach abhainn láibhe ina dhóitear go mór. Troideann Yoda an t-Impire Palpatine ar Coruscant go dtí go dtarlaíonn an cath mar gheall air. Is Jedi cumhachtach é Yoda, ach ní féidir leis an olc Impire Palpatine a chosc. Téann Yoda ansin le Bail Organa agus téann Palpatine chuig Mustafar. Úsáideann an t-Impire Palpatine olc taobh dorcha na Fórsa le tuiscint go bhfuil Anakin i mbaol.

Casann Obi-Wan go Yoda chun athghrúthú. Tugann Padmé dá mhac agus d'iníon dúbailte a n-ainmníonn sí Luke agus Leia. Braitheann Padmé brón go gairid ina dhiaidh. Faigheann Palpatine Anakin dóite go fóill fós beo ar Mustafar. Tar éis dó dul ar ais chuig Coruscant, déileálfar le comhlacht máinliachta Anakin agus clúdaítear é in oireann armúrtha dubh. Bíonn Palpatine ag Anakin go maraíodh Padmé ina chlog. Is ceannmáistir olc é Palpatine agus fágann mothú Anakin ar a chéile. Bhuaigh Palpatine; tá taobh dorcha na Fórsa anois ag Anakin. Idir an dá linn, oibríonn Obi-Wan agus Yoda chun an cúpla a cheilt ó thaobh dorcha na Fórsa, toisc gurb é an cúpla is dóchas ach amháin le haghaidh saoirse. Téann Yoda féin chuig an bplainéad Dagobah, agus maoiríonn Anakin agus an t-Impire Palpatine an déantús an Réalt Bás. Uchtaíonn Bail Organa Leia agus tógann sí í chuig Alderaan. Taistealaíonn Obi-Wan le Luke go Tatooine. Tá sé i gceist ag Obi-Wan féachaint go láidir ar Luke agus ar a theaghlach go dtí go mbeidh an t-am ceart dúshlán a thabhairt don Impireacht. □

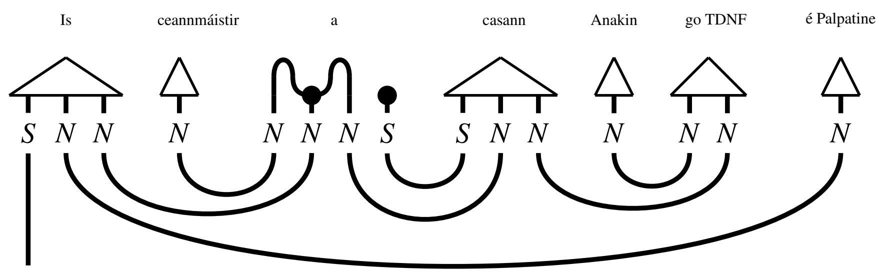
C A More Complicated Example

We shall compute the similarity of meaning between "Palpatine is a mastermind who turns Anakin to the dark side of the Force" and its Irish equivalent, "*Is ceannmáistir a casann Anakin go taobh dorcha na Fórsa é Palpatine*". The Irish sentence is assigned the following type:

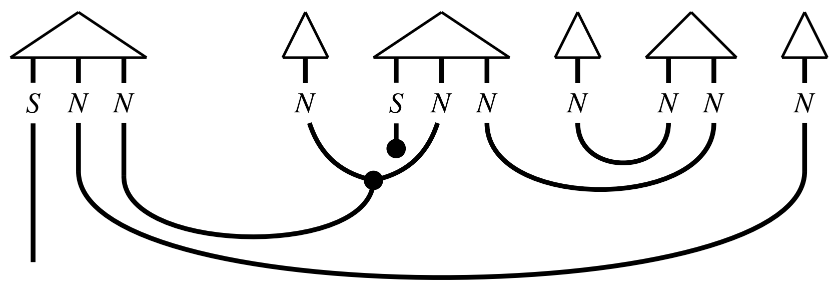
Is ceannmáistir a casann Anakin go taobh dorcha na Fórsa é Palpatine

$$sn^l n^l \quad n \quad n^r n n^{ll} s^l \quad sn^l n^l \quad n \quad n^r n \quad n$$

Abbreviating “taobh dorcha na Fórsa” as “TDNF”, the reduction diagram is[8]:



and is simplified to:



This diagram corresponds to the map

$$\begin{aligned} f = {} & (1_S \otimes \varepsilon_N) \circ (1_S \otimes 1_N \otimes \varepsilon_N \otimes 1_N) \circ (1_S \otimes 1_N \otimes 1_N \otimes \mu_N \otimes \varepsilon_N \otimes 1_N) \\ & \circ (1_S \otimes 1_N \otimes 1_N \otimes 1_N \otimes i_S \otimes 1_N \otimes 1_N \otimes \varepsilon_N \otimes 1_N \otimes 1_N). \end{aligned}$$

Note in this sentence the verb *casann* (“turns”) has been modified to follow the rule *Verb Object Subject*. Therefore we will use the transpose of the matrix C^{cas} to accommodate this change. The meaning vector of the sentence “*Is ceannmaáistir a cassan Anakin go taobh dorcha na Fórsa é Palpatine*” is:

$$\begin{aligned} & \overrightarrow{\textit{Is ceannmáistir a casann Anakin go taobh dorcha na Fórsa é Palpatine}} \\ & = 330\vec{n}_2 \otimes \vec{n}_1 + 40\vec{n}_2 \otimes \vec{n}_2. \end{aligned}$$

The author has excluded the calculations of the matrices $(C^{\text{cas}})^T$ and $C^{\text{go TDNF}}$ for brevity, but these are calculated as per §4. Taking the inner product,

$$\begin{aligned} & \langle \overrightarrow{\text{Palpatine is a mastermind who turns Anakin to the dark side of the Force}} \mid \\ & \quad \overrightarrow{\textit{Is ceannmáistir a casann Anakin go taobh dorcha na Fórsa é Palpatine}} \rangle = 106880. \end{aligned}$$

The length of the former is 103424, and the length of the latter is 110500. Therefore their similarity score is $\frac{106880}{\sqrt{103424 \cdot 110500}} = 0.999$; very high.

[8]Taking cues from the English “who” [23] regarding the depiction of “*a*” in the diagram.

Suppose we thought the translation of "*Is ceannmáistir a casann Anakin go taobh dorcha na Fórsa é Palpatine*" was "**The Emperor** is a mastermind who turns Anakin to the dark side of the Force". The calculations tell us:

$$\langle \overrightarrow{\text{The Emperor is a mastermind who turns Anakin to the dark side of the Force}} \mid \overrightarrow{\textit{Is ceannmáistir a casann Anakin go taobh dorcha na Fórsa é Palpatine}} \rangle = 534400.$$

The length of the former is 2593792, and the length of the latter is 110500. Therefore the similarity score of the sentences is $\frac{534400}{\sqrt{2593792 \cdot 110500}} = 0.998$; nearly as high as the actual translation. Of course, as Palpatine *is* the Emperor, one could argue this is still a valid translation.

D Corpora for a Conceptual Space Model of Meaning

We shall make the same assumptions regarding this corpus as in *Appendix A*.

Corpus D.1. *Venus is a planet in the solar system. Venus has a solid and rocky surface. Venus is called Earth's sister because it is nearly the same size as Earth. Venus is very hot and the pressure on its surface is high. Venus is bright in the night sky and looks like a ball.*

Jupiter, another planet in the solar system, also looks like a ball. Jupiter sits in outer space. The size of Jupiter is very large; it is the largest planet in the solar system. Jupiter is large and gassy. Jupiter is primarily orange and brown and red in colour. Jupiter is far away from Earth. It is very windy on Jupiter and also freezing cold. Jupiter is very bright in the night sky.

Mars is a planet next to Earth. Mars is coloured very red, and brown and orange. Mars is cold, but not very cold. Mars is smaller than Earth. Mars is rocky like Venus and Earth. Mars sits in outer space.

Apples are fruits. Apples are round and soft. Apples can be red or green, and some apples taste bitter and other apples taste sweet. An apple looks like a ball.

The Sun is a star, not a planet. It sits in the centre of the solar system. The Sun is the brightest thing in the sky. The Sun is huge and very hot. The Sun is round and also looks like a ball. The gravity on the Sun is very strong, meaning it is very dense. □

Here is the same corpus, but in Irish:

Corpus D.2. *Is í Véineas plánéad sa ghrianchóras. Tá dromchla tathagach agus carraigeach ag Véineas. Glaotar Véineas deirfiúr an Domhan í mar tá sí beagnach an méid céanna leis an Domhan. Tá sé an-te ar Véineas agus tá an brú ar a dromchla ard. Tá Véineas geal i spéir na hoíche agus breathnaíonn sí cosúil le liathróid.*

Breathnaíonn Iúpatar, plánéad eile sa ghrianchóras, cosúil le liathróid freisin. Suíonn Iúpatar i spás seachtrach. Tá Iúpatar an-mhór; tá sé an plánéad is mó sa ghrianchóras. Tá Iúpatar mór agus déanta sé as gáis. Tá Iúpatar go príomha oráiste agus donn agus dearg i ndath. Tá Iúpatar i bhfad gcéin ó an Domhan. Tá sé an-ghaothmhar ar Iúpatar agus an-fhuar freisin. Tá Iúpatar an-gheal i spéir na hoíche.

Is é Mars pláinéid in aice leis an Domhan. Tá Mars daite an-dearg, agus donn agus oráiste. Tá sé fuar ar Mars, ach níl sé an-fhuar. Tá Mars níos lú ná an Domhan. Tá Mars carraigeach cosúil le Véineas agus an Domhan. Suíonn Mars i spás seachtrach.

Is torthaí iad úlla. Tá úlla liathróideach agus bog. Féadfaidh úlla a bheith dearg nó glas, agus tá blas searbh ar roinnt úill agus blas milis ar úlla eile. Breathnaíonn úll cosúil le liathróid.

Is réalta í an Grian, ní phláinéid. Tá sí suite i lár an chórais ghréine. Is í an grian an rud is gile sa spéir. Tá an grian ollmhór agus an-te. Tá an grian liathróideach agus breathnaíonn sí ar liathróid fresin. Tá an imtharraingt ar an ghrian an-láidir, rud a chiallaíonn go bhfuil sí an-dlúth. □

E Additional assignment of noun spaces arising from *Corpus D.2*.

Organising the information of *Corpus D.2* into a table we obtain:

Véineas	**Adjectives**	*tathagach, carraigeach, beagnach an méid céanna leis an Domhan, an-te, brú . . . ard, geal.*	solid, rocky, same size as Earth, hot, high pressure, bright.
	Nouns	*plánéad, deirfiúr an Domhan, liathróid.*	planet, Earth's sister, ball.
Iúpatar	**Adjectives**	*an-mhór, déanta as gáis, oráiste, donn, dearg, i bhfad i gcéin, an-ghaothmhar, an-fhuar, an-gheal.*	very large, gassy, orange, brown, red, far away, windy, freezing, very bright.
	Nouns	*plánéad, spás seachtrach, liathróid.*	planet, outer space, ball.
Mars	**Adjectives**	*an-dearg, oráiste, donn, fuar, níos lú ná an Domhan, carraigeach.*	very red, brown, orange, cold, smaller than Earth, rocky.
	Nouns	*plánéad, spás seachtrach.*	planet, outer space.
Úll	**Adjectives**	*liathróideach, bog, dearg, glás, searbh, milis.*	round, soft, red, green, bitter, sweet.
	Nouns	*torthaí, liathróid.*	fruit, ball.
Grian	**Adjectives**	*an rud is gile, ollmhór, an-te, liathróideach, an-dlúth.*	brightest, huge, very hot, round, very dense.
	Nouns	*réalta, liathróid.*	star, ball.

Recall that, if an adjective class is not mentioned, its corresponding noun space is set to $[0,1]$.

(2) ***Iúpatar.***

$N_{\text{dimension}} = Conv(an\text{-}mhór) = \{0.8\}$,
$N_{\text{colour}} = Conv(oráiste \cup donn \cup dearg)^9$,
$N_{\text{intensity}} = Conv(an\text{-}geal) = \{0.7\}$,
$N_{\text{temperature}} = Conv(an\text{-}fuar) = \{0.1\}$,
$N_{\text{density}} = Conv(déanta\ as\ gáis) = \{0.1\}$.

(3) ***Mars.***

$N_{\text{dimension}} = Conv(níos\ lú\ ná\ an\ Domhan) = \{0.25\}$,
$N_{\text{colour}} = Conv(dearg \cup donn \cup oráiste)$,
$N_{\text{temperature}} = Conv(fuar) = \{0.4\}$,
$N_{\text{texture}} = Conv(carraigeach) = \{0.9\}$.

(4) ***Úll.***

$N_{\text{colour}} = Conv(dearg \cup glás)$,
$N_{\text{taste}} = Conv(searbh \cup \text{milis})$,
$N_{\text{texture}} = Conv(bog) = \{0.4\}$.

(5) ***Grian.***

$N_{\text{dimension}} = Conv(ollmhór) = \{0.9\}$,
$N_{\text{intensity}} = Conv(an\ rud\ is\ gile) = \{1\}$,
$N_{\text{temperature}} = Conv(an\text{-}te) = \{0.85\}$,
$N_{\text{density}} = Conv(an\text{-}dlúth) = \{1\}$.

For $K = 1, \ldots, 5$, D^K_{adj} is the tensor product of the $N_{(-)}$ spaces of item (K). These values were assigned according to the author's own preference, however they can be assigned different values according to each readers' wishes.

[9] The RGB values we use for *oráiste*, *donn* and *dearg* are (255,165,0), (153,76,0), and (255,0,0) respectively.