Mr Darcy and Mr Toad, gentlemen: distributional names and their kinds

Aurélie Herbelot
University of Cambridge
aurelie.herbelot@cantab.net

Abstract

This paper investigates the representation of proper names in distributional semantics. We define three properties we expect names to display: uniqueness (being a unique entity), instantiation (being an instance of a relevant kind) and individuality (being separable from the subspace of concepts). We show that taking a standard distribution as the representation of a name does not satisfy those properties particularly well. We propose an alternative method to compute a name vector, which relies on re-weighting the distribution of the appropriate named entity type – in effect, producing an individual out of a kind. We illustrate the behaviour of such representations over some characters from two English novels.

1 Introduction

Distributional semantics (DS) rests on the idea that the meaning of words is given by their linguistic context (Harris, 1954). While DS has been very successful in modelling a range of phenomena – in particular those linked to similarity – the scope of the theory remains unclear. Some have called it a semantics of the lexicon, i.e. not a fully formed theory of meaning but an aspect of lexical knowledge, with influence over other parts of semantics (Baroni et al., 2014; Erk, 2014). This view is supported by the fact that by nature, distributional information is an aggregate of all contexts in which a word appears, and therefore some kind of average representation tending towards generic, conceptual meaning. In turn, it has been suggested that, as conceptual representations, they might be linked to a traditional notion of intension (Erk, 2013).

While a lot of work in DS has focused on noun phrases, one particular type has been mostly neglected: those phrases classically referred to as **proper names** in philosophy of language. Mill (1843) defines a proper name as 'a word that answers the purpose of showing what thing it is that we are talking about but not of telling anything about it'. Under that category fall proper nouns (*Alice, Carroll*), noun phrases containing proper nouns (*Lewis Carroll*) and even noun phrases denoting unique objects but consisting solely of (in English, usually capitalised) common nouns (*the Tower Bridge*).

Proper names are an interesting test-bed for distributional semantics for several reasons. First, although they are fully part of the lexicon, they refer to uniquely referenced objects: to our knowledge, this type of linguistic entity has never been studied in DS. Second, their semantics is contested and some have claimed that they do not, in fact, have intension and are purely extensional – while others support various types of intensional notions, ranging from minimalist accounts where e.g. *Smith* means 'the individual named Smith' to complex clusters of definite descriptions. Thirdly, they are good examples of lexical items which, often, are previously unknown to the hearer (i.e. we may have no preconception of the meaning of *Anna Pavlovna* until we encounter the corresponding character in *War and Peace*).

This paper is an exploration of proper names in distributional settings. We introduce three properties that are expected from names: uniqueness (being a unique entity), instantiation (being an instance of a relevant kind) and individuality (being separable from the subspace of concepts). We argue that the distributional representation of individuals is badly modelled by a standard co-occurrence vector in a static corpus: not only does it fail to satisfy the above properties, but it amounts to an unrealistic view of the lexicon where all items are *a priori* present. While this view is defensible with regard to a general representation of an adult's lexical knowledge, it does not hold for proper names, neologisms, etc.

In the following, we contrast 'standard' vectorial representations of names with a model based on the contextualisation of a more general distribution. For instance, we assume that upon encountering the name *Mr Darcy* for the first time in the novel *Pride and prejudice*, a reader will attribute it the representation of the concept *man* and subsequently specialise it as per the linguistic contexts in which the name appears. We show that this approach performs better in modelling the expected properties.

The resulting model can be described in terms of kinds and individuals. Distributions obtained in the standard way from large corpora represent generic, conceptual information about a kind. The contextualisation process produces an individual from the relevant kind. Although the focus of this paper is on names, it will be clear that the proposed method is applicable to common nouns, so that the individual referred to as *my cat* can be generated from the concept *cat* across relevant contexts in a discourse. To finish, we tentatively show that the created individuals can be added up to form pluralities which stand between concepts and their instances in the semantic space.

2 Related work

2.1 Proper names

There is an extensive philosophical literature on proper names, which we will not attempt to summarise here (see Cumming, 2013 for an overview). We will instead focus on the particular approach which we can best relate to distributional theories.

Mill (1843) started the debate on the semantics of proper names with a purely extensional view, stating that the meaning of a name is its referent in a world – and only its referent. This view came under attack from proponents of both sense theories and so-called 'descriptivism'. The sense theory (Frege, 1892) relies on a notion of 'sense' or 'intension' to describe the cognitive content of a name. This notion allows us to explain the semantic difference between *Evening Star* and *Morning Star*, while acknowledging that their referents in the real world are one same object, the planet Venus. Similarly, descriptivists took the stance that a name has the semantics of a definite description which, by proxy, provides its meaning (be it through its intension or any other device). For instance, the meaning of *Aristotle* might be equated with the meaning of *the teacher of Alexander the Great* (Russell, 1911).

When trying to relate a distributional description of proper names to standard philosophical theories, the most natural correspondence is probably 'cluster' descriptivism (Searle, 1958), which states that the semantics of a name is a complex definite expression potentially including several predicates, i.e. the semantics of *Aristotle* might be that of *the teacher of Alexander the Great and most famous pupil of Plato*. Distributionally, we might say that the meaning of *Aristotle* is a distribution where contexts are complex predications and the predicates *teacher of Alexander* and *pupil of Plato* are highly weighted. Note that the distribution does not, at first glance, encapsulate the definiteness which makes the individual unique. In §3.2, we will discuss how to assess the uniqueness and individuality of a name.

2.2 Distributional meaning in context

Various distributional methods have been proposed to compute the meaning of a word in context (this work is summarised in Dinu et al., 2012). Proposals can be classified in two groups: those which consist, roughly, in 'reweighting' a target vector using its close syntactic context (e.g. Erk and Padó, 2008; Thater et al., 2011), and those which build the contextualised vector by selecting corpus occurrences of the target word that are 'similar' to the context under consideration (e.g. Rapp, 2004).

Contextualisation methods have been mainly studied from the point of view of disambiguation and selectional preference, showing they could solve a range of lexical problems. We build on this work by applying the technique further in order to separate, not various senses of a word, but various instances of a concept (§5).

3 Building distributional names: corpora and experimental design

3.1 Corpora

In this work, we use three different corpora:

- the British National Corpus (BNC), a balanced corpus of British English totalling 100M words, which we use for the purpose of obtaining a 'general' semantic space for English. This gives us distributional representations for the most frequent words in the language.
- *Pride and prejudice* by Jane Austen (1813), a novel of around 13,000 tokens (henceforth *P&P*).
- *The wind in the willows* by Kenneth Grahame (1908), a children's novel of around 6000 tokens (henceforth *WitW*).

We will be analysing the behaviour of a range of names occurring in the above novels. One natural question that arises is whether their distributions should be computed from the relevant novel only, or from its concatenation with the BNC. Insofar as a novel creates a different world from the real one – potentially creating or altering word meanings – it seems more reasonable to create a semantic space for the book in isolation. This has however the disadvantage of extracting vectors from very sparse data: the counts obtained for the names' contexts may not reflect their general distributions.

Following preliminary experimentation, we decided to create semantic spaces for each book separately. Indeed, extracting name distributions from the concatenated corpora resulted in vectors heavily biased towards book-specific contexts (with e.g. Mr Darcy showing high weights against other *P&P* names and rarer words like *parsonage* or *to-morrow*). In §5, we will show how to make use of the lexical information in the BNC as a complement to the novel-specific distributions.

We build a distributional space for each corpus, using the DISSECT toolkit (Dinu et al., 2012). We select Positive Pointwise Mutual Information (PPMI) as weighting measure and word windows of size 10 as context. We apply sum normalisation to the vectors. We tune the size of each space by evaluating it against a standard similarity task: we take the MEN dataset[1] (Bruni et al., 2012), which consists of word pairs annotated with human similarity judgements, and calculate the Spearman (ρ) correlation between the cosine distances for those pairs in the semantic space and the human annotations. When a pair contains a word that does not occur in the corpus, it is discarded.

For the BNC, we tune both the number of dimensions and vocabulary size in the range 1000-10000. We obtain a correlation of 0.689 for the BNC with 4000 dimensions and a vocabulary consisting of the 5000 most frequent words in the corpus. This is in line with state-of-the-art results for the MEN dataset. We note, however, that this space does not contain the words *toad, badger* and *mole*, which are crucial to our analysis. We add them to our best space, with no significant reduction to the correlation ($\rho = 0.688$).

For the novels, we tune the number of dimensions over a range from 500 to 5500 and the vocabulary size from 500 to 2000. *P&P* gets its highest correlations with 500 dimensions and a vocabulary made of the 1000 most frequent words in the corpus. At $\rho = 0.376$, the correlation is much lower than that obtained on the BNC, but this is expected given the size of the corpus. *WitW* performs best with 1500 dimensions and a smaller vocabulary of the top 500 items in the novel ($\rho = 0.358$).

3.2 Design

For clarity reasons, we adopt a somewhat arbitrary terminology in the following, where we use the word 'concept' to refer solely to the (distributionally modelled) intensions of *non*-proper names (e.g. *aardvark, blue, think*) (see Erk, 2013; McNally, 2014 for accounts of distributions as concepts). Further, we will also abide by the idea that concept vectors express some sort of 'generic' information. It is worth considering why this is a reasonable assumption for common nouns. We recall that a noun's distribution is a statistical representation of its usages, which normally gives more weight to contexts

[1] http://clic.cimec.unitn.it/ elia.bruni/MEN.

that are characteristic for it. It is calculated over a large number of tokens which themselves partake in references to a) individuals (*my* dog), b) pluralities (*the neighbour's* dogs) and c) kinds (*the* dog *is a mammal*). There is perhaps little intuition as to how those referent types contribute to the distribution. We attempted to gain some understanding of this by consulting previous work on the annotation of generic noun phrases. Herbelot and Copestake (2011) produced a small dataset[2] consisting of 300 randomly selected noun phrases, annotated (amongst other things) with kinds.[3] From this data, we extrapolate that around 10% of noun phrases in text refer to kinds, with the overwhelming majority denoting individuals or (existential) plurals. This indicates that in the distribution of a common noun, a context with high weight is one which is characteristic of individuals or groups, rather than of the concept itself.

We can then say that a distribution is a representation of the type of things *generally* said of the *instances* of a particular concept. In other words, the distribution is generic (a kind) not by virtue of collecting contexts associated with kind references but by virtue of generating a model of the concept's 'supremum', i.e. of all its instances, as given by a closed corpus (for an account of kinds as supremums, see e.g. Chierchia, 1998). In what follows, we will assume that 'kind' and 'concept' can be used interchangeably (McNally, 2014 follows a similar argument).

Having clarified our notion of concept, we can turn to individuals. In an ideal semantic space, name distributions would have the following properties:

1. **Uniqueness:** the intension of a proper name should let us capture its unique extension in a given world. This implies that intensions themselves should be separable within the distributional space. In other words, two Smiths referring to separate individuals should also have separate intensions, i.e. occupy different points in the semantic space.

2. **Instantiation:** names should stand in a learnable relationship to the concept they are instantiating. For instance, Mr Darcy should clearly be an instance of *man, person*, etc.

3. **Individuality:** proper names should be distinguishable from concepts. This is related to the uniqueness property but not identical to it. Let's for instance assume a world with exactly one dodo named Dolly. If the intension of *Dolly* were the same as the intension of *dodo*, this would make Dolly unique (because there is only one dodo in that world) but wouldn't stress her individuality, i.e. that she is *a* dodo and not the kind *dodo*.

Assuming that different individuals do not occur exactly in the same contexts in corpora, the **uniqueness** property is satisfied by selecting only those occurrences of a name which refer to the same entity.[4] In other words, the 'relevant' contexts for building a name distribution are the ones surrounding mentions in a co-reference chain. We note that this context selection is similar to the method used by e.g. Rapp (2004) for dealing with polysemy. In the following, for simplicity reasons and to avoid noise, we do not run co-reference resolution over our corpora. Rather, we only consider unambiguous proper names and build a distribution over their occurrences (this is in keeping with the standard way to build distributions, where e.g. pronominal anaphors are ignored).

Instantiation is testable by borrowing distributional measures which have been shown to perform well in the task of hyponymy detection. That is, given a distributional name, we can attempt to extract the concept(s) it most likely instantiates by assuming they partake in the same kind of relation as nouns to their hypernyms. In our experiments, we use the invCL measure (Lenci and Benotto, 2012), which takes into account how much a hyponym occurs in a subset of the contexts in which its hypernym appears, and how much the hypernym occurs in a superset of the contexts associated with the hyponym. For each name, we calculate its invCL score with respect to its 50 nearest neighbours (specifically, the 50

[2] Available at `http://www.cl.cam.ac.uk/~ah433/underquantification.kind.annot`.

[3] In the related paper, a kind is defined as a noun phrase which can be paraphrased in context as either a bare plural or a singular: *(A/The) Scottish fiddler(s) emulating 18th century playing styles sometimes use a replica of this type of bow.*

[4] We can imagine the extreme case of two Smiths described in exactly the same contexts but referring to two different individuals. But we would argue that, given the linguistic contexts only, a human would not be able to capture their extensional difference.

common nouns which are closest to it in the semantic space). We output the highest scores as potentially instantiated concepts and manually verify the results.

The **individuality** property is not trivial to capture. One clear-cut test, from the generics literature, would be to inspect how a name's distribution interacts with those predicates which are only applicable to kinds (*extinct, widespread, common* – see Carlson and Pelletier, 1995). We would expect that the composition of an individual with, say, *widespread* would result in a semantically anomalous sentence: **Mr Toad is widespread*. Unfortunately, there are very few such 'kind predicates', making them inadequate for a quantitative study. More promisingly perhaps, some predicates – typically those with so-called 'positive alternatives' – are known to be unsuitable for kinds but appropriate for individuals: **Badgers are male/Mr Badger is male* (the generic sentence is blocked by the positive alternative **Badgers are female*, see Leslie, 2008). Testing the felicity of generic statements, however, is non-trivial as it is dependent on many factors, from human inter-annotator agreement to the distributional composition function used to combine the subject and predicate. So instead, we propose in what follows a related but straightforward test.

We note that, while the most characteristic contexts of a kind can be extensionally exclusive, those of an individual (at least the static predicates) should be less so. For instance, both *rich* and *poor* may appear in the top contexts of *man* (making the corresponding **Men are rich/poor* infelicitous), but we would only expect one or the other at the top of an individual's distribution. More generally, we suggest that the characteristic contexts of an individual will be significantly more coherent than those of a kind. That is, the linguistic items strongly associated with an individual should be on the whole related to each other because that individual cannot embody the range of experiences covered by many members of a group. To test individuality, we thus propose to calculate the coherence of the top 50 characteristic contexts for each name distribution and compare it to the coherence of the kind(s) it instantiates. As in Newman et al. (2010), we define the coherence of a set of words $w_1...w_n$ as the mean of their pairwise similarities (where similarity is the cosine distance between two vectors):

$$Coherence(w_{1...n}) = mean\{Sim(w_i, w_j), ij \in 1...n, i < j\}$$

We illustrate our claims by closely inspecting the distributions of a small range of individuals in the two novels under study. These include some major and more minor protagonists in *Pride and prejudice* (Mr Darcy, Mr Bingley, Elizabeth and Jane Bennett, Mr Collins and Mr Denny) and the main protagonists in *The wind in the willows* (Mr Toad, Mole, Rat, Badger).

4 Individuals as they come

The obvious way to start building distributions for names is simply to regard them as any other lexical item and output their vector in the standard way. Here are the top characteristic contexts of Mr Darcy (*P&P*) and Mr Toad (*WitW*):

Darcy (Mr, acquaint, Miss, interest, eye, Pemberley, degree, stand, wholly, walk, nephew, sake, civility, surprise)

Toad (Hall, sternly, ho, paddock, smash, whack, popular, nonsense, trot, seize, disguise, cushion, terror, necessaries)

The problem with such a representation is that it is typically very sparse. *Darcy* occurs 416 times in *Pride and prejudice*, but *Denny*, for example, only occurs 11 times. In the following, we show how such representations fare against the design requirements highlighted in §3.2.

4.1 Analysis

4.1.1 Instantiation

Tables 1 and 3 report the top 5 invCL scores for various characters. Relevant concepts are shown in bold. For *P&P*, we note that Darcy and Bingley are correctly classified as gentlemen, at the top of the table,

Darcy	Elizabeth	Bingley	Jane	Collins	Denny
0.47 **gentleman**	0.47 moment	0.48 **gentleman**	0.48 feeling	0.50 daughter	0.47 news
0.47 word	0.46 subject	0.48 lady	0.47 **sister**	0.48 house	0.47 intention
0.46 manner	0.46 feeling	0.46 sister	0.46 pleasure	0.47 family	0.47 aunt
0.46 feeling	0.46 pleasure	0.46 party	0.46 aunt	0.46 **cousin**	0.45 journey
0.46 conversation	0.45 house	0.46 answer	0.46 letter	0.46 lady	0.44 home

Table 1: Top invCL scores for various characters in *Pride and prejudice* – standard distributions

Toad	Rat	Mole	Badger
0.41 **animal**	0.40 water	0.40 **animal**	0.43 time
0.38 **toad**	0.39 **animal**	0.38 time	0.43 **animal**
0.38 time	0.39 time	0.37 thing	0.40 thing
0.37 way	0.37 thing	0.36 way	0.39 friend
0.36 thing	0.36 way	0.35 water	0.38 toad

Table 2: Top invCL scores for various characters in *Wind in the willows* – standard distributions

and that other characters have a relevant hypernym amongst the highest scores: Jane is a sister, Collins a cousin. However, two characters (Elizabeth and Denny) fail to return any concept relevant to their actual status. Moreover, we observe a high proportion of irrelevant gendered/family relations (Mr Collins is no daughter or lady, Mr Denny no aunt), and even non-human items in the lists. The characters of *WitW* present similar issues, although all are classified as 'animals'. Only Toad, however, returns the common noun *toad* as hypernym.

4.1.2 Individuality

We report the individual coherence of all characters under consideration, and compare them to the coherence of the concepts they (should) instantiate. In order to get good conceptual representations, we use the BNC vectors which we take to average many more instances of a kind than are available in either novel. For the same reason, BNC distributions are taken as representations of the characteristic contexts. These choices mean that in effect, we are comparing the characteristic contexts of a character, as given in the corpus where s/he appears, with the characteristic contexts of a generic man/woman/toad, etc. as given by a large, all-purpose corpus. Table 3 shows all results. Humans in *P&P* are compared with both *man/woman* and *gentleman/lady*.

We note that in general, names are a little less coherent than the concepts they instantiate, indicating that their distributions do not satisfy the individuality property very well.

	Individual coherence	Kind coherence								
		woman	lady	man	gentleman	toad	rat	mole	badger	animal
Darcy	0.22	-	-	0.24	0.25	-	-	-	-	-
Elizabeth	0.24	0.24	0.28	-	-	-	-	-	-	-
Bingley	0.23	-	-	0.24	0.25	-	-	-	-	-
Jane	0.23	0.24	0.28	-	-	-	-	-	-	-
Collins	0.22	-	-	0.24	0.25	-	-	-	-	-
Denny	0.23	-	-	0.24	0.25	-	-	-	-	-
Toad	0.21	-	-	-	-	0.24	-	-	-	0.22
Rat	0.23	-	-	-	-	-	0.24	-	-	0.22
Mole	0.22	-	-	-	-		-	0.22	-	0.22
Badger	0.21	-	-	-	-		-	-	0.23	0.22

Table 3: Coherence values for some characters and the concepts they instantiate – standard distributions

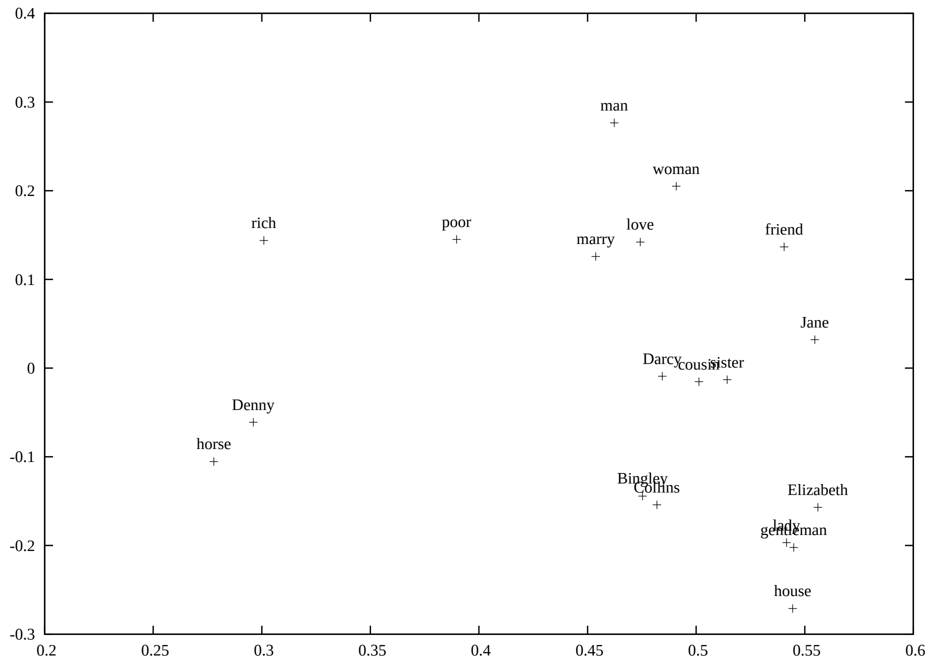

Figure 1: The *Pride and prejudice space* – standard distributions. The figure was produced by selecting relevant names and concepts from the semantic space and applying SVD dimensionality reduction to 'flatten' the space to two dimensions.

4.1.3 Discussion

Our results suggest that standard distributional representations of proper names model instantiation to some extent, but fail to show the individuality of the referents.

As a further qualitative analysis of the obtained representation, we show a 2D visualisation of the semantic space for *P&P* in Fig. 1. We observe that there is no principled separation between names and concepts: although individuals tend to roughly occupy the same portion of the space, they are situated in the midst of concepts. We can expect that learning a function that separates kinds from individuals would be extremely challenging.

5 Names as contextualised kinds

We observed in §4.1 that standard distributions do not model instantiation well. But we would expect that in virtue of being called *Badger* or *Mr Darcy*, an individual would naturally acquire the properties of the kind *badger* or the kind *man/gentleman* – without, in fact, needing any context. In this section, we try out another model of names which relies on the contextualisation of a distributional kind.

5.1 System description

We suggest that a speaker of English already has an extensive lexical knowledge when coming to *P&P* or *WitW*. That is, her interpretation of the words in the books relies on prior assumptions with regard to their meaning. For proper names, it is expected that conventions, the context surrounding the first mention and/or world knowledge will anchor them as a named entity of a certain type (*Mr Darcy* is a male person in virtue of his title, the *Tower Bridge* explicitly refers to its nature via a common noun, *London* is supposedly known to be a city and *Pemberley* is understood to be a place name in the sentence *she will always be at Pemberley with you* (*P&P*, Chap. 6)). In computational linguistics, this process is performed via named entity recognition (NER).

In this work, we assume that NER has taken place over our corpus and that the *P&P* characters have been classed as either *man* or *woman*. We also presuppose, trivially, that *Toad* has been recognised as a toad, *Rat* as a rat, etc. We propose that on encountering Mr Darcy for the first time, a reader might simply attribute him the properties of the lexical item *man*, as given by the relevant distribution in a large corpus, and then specialise the representation as per the contexts where *Darcy* occurs.

We note that in this setup, the individuality property is at odds with instantiation. In order to make Darcy a unique individual, more weight must be given to the features that distinguish him from the kind *man*. Doing so, however, may result in a vector that does not stand anymore in a principled relationship with the concept it instantiates.

We formalise a name distribution as follows. Let N be a proper name, instance of kind K. N has a 'standard' distribution $\boldsymbol{v}(N)$, as obtained in §4, with m characteristic contexts $c_1...c_m \in C$ (i.e. the m most highly weighted dimensions in the vector) . K also has a distribution $\boldsymbol{v}(K)$ which lives in a space S with dimensions $d_{1...n} \in D$, as obtained from a large background corpus (in our example, the BNC). We define $\boldsymbol{v}(K)$ in terms of S's basis vectors $\{\boldsymbol{e}_{d'} | d' \in D\}$ and a weighting function w (in our case, PPMI):

$$\boldsymbol{v}(K) = \sum_{d' \in D} w(K, d').\boldsymbol{e}_{d'} \tag{1}$$

We could contextualise $\boldsymbol{v}(K)$ with respect to each context in which the name appears. For efficiency reasons, however, we simply perform the contextualisation with respect to each of the characteristic contexts $c' \in C$ in $\boldsymbol{v}(N)$, using the following function:

$$C(K, c') = \sum_{d' \in D} cos(c', d')^p w(K, d').\boldsymbol{e}_{d'} \tag{2}$$

This is equivalent to one of the models proposed by Thater et al. (2011), but without taking into account the nature of the syntactic relation between K and c'. Further, we introduce a weight p acting on the cosine function to increase or decrease the effect of the contextualisation. The assumption is that a higher p makes the individual more 'unique' but less like its kind (see Erk and Padó, 2008 for a similar use of powers).

The name vector for N is the sum of the contextualisations with respect to all characteristic contexts in C:

$$\sum_{c' \in C} \sum_{d' \in D} cos(c', d')^p w(K, d').\boldsymbol{e}_{d'} \tag{3}$$

The following parameters can be tuned: a) the number m of characteristic contexts used for describing the name; b) the p value. We consider 10-30 characteristic contexts and $p =$ 1-10.

5.2 Analysis

Varying p mostly affects the coherence values used to assess the degree of individuality of a name. We note a general increase in coherence with higher values of p, although it comes to a plateau at $p = 6$ and slowly decreases again. High values of m have the effect that some characters do not satisfy the instantiation property anymore.

In the following, we report our best results, i.e. the system which returns the highest individuality (coherence) figures while leaving the names in the expected instantiation relation with the relevant concepts. This is obtained using $m = 20$ and $p = 6$. We should note that a large quantitative survey would be needed to ascertain which parameter combination best models the cognitive individuation process. We leave this for further work and focus here on a illustrative evaluation of the procedure.

We first note that the instantiation property is straightforwardly satisfied by the model (see Table 4). The shape of the initial kind vector is retained throughout the contextualisation, so that names remain instances of their respective kinds. Interestingly, however, we note that the *WitW* characters also move

Darcy	Elizabeth	Bingley	Jane	Toad	Badger
0.97 **man**	0.97 **woman**	0.98 **man**	0.98 **woman**	0.97 **toad**	0.97 **badger**
0.91 girl	0.90 girl	0.91 boy	0.82 girl	0.75 sea	0.72 sight
0.91 face	0.89 eye	0.90 girl	0.82 man	0.74 desert	0.72 dog
0.91 boy	0.88 man	0.88 eye	0.81 other	0.73 rock	0.71 boy
0.90 smile	0.88 face	0.88 face	0.79 eye	0.73 mountain	0.71 fox

Table 4: Top invCL scores for various characters – contextualised individuals

	Individual coherence	Kind coherence								
		woman	lady	man	gentleman	toad	rat	mole	badger	animal
Darcy	**0.42**	-	-	0.24	0.25	-	-	-	-	-
Elizabeth	**0.40**	0.24	0.28	-	-	-	-	-	-	-
Bingley	**0.42**	-	-	0.24	0.25	-	-	-	-	-
Jane	**0.34**	0.24	0.28	-	-	-	-	-	-	-
Collins	**0.34**	-	-	0.24	0.25	-	-	-	-	-
Denny	**0.40**	-	-	0.24	0.25	-	-	-	-	-
Toad	**0.28**	-	-	-	-	0.24	-	-	-	0.22
Rat	**0.32**	-	-	-	-	-	0.24	-	-	0.22
Mole	**0.24**	-	-	-	-	-	-	0.22	-	0.22
Badger	**0.28**	-	-	-	-	-	-	-	0.23	0.22

Table 5: Coherence values for some characters and the concepts they instantiate – contextualised distributions

towards human concepts: Badger, for instance, has *boy* as its third most likely kind. This tendency gets stronger as p increases, with Mole returning the kinds *mole, human, adult* at $p = 10$. This indicates that in cases where the initial named entity type turns out to be incorrect or partially correct (as in the case of anthropomorphised animals), the model has the potential to rectify the representation.

Further, the produced vectors strongly assert the individuality of the modelled names, in particular for the *P&P* characters. Table 5 shows that all names have higher coherence than their respective kinds (all differences are statistically significant). The names from *WitW* are generally less coherent than those in *P&P*, which can perhaps be explained by the fact that the characters combine properties from two separate areas of the semantic space (animal- and human-related features).

5.2.1 Discussion

A visualisation of the semantic space is provided in Fig. 2. Due to layout restrictions, we show the space for $m = 20$ and $p = 3$. Our best model ($p = 6$) results in essentially the same configuration, but with the names much further apart from the concepts. It is however clear from the illustration that individuals are separated from kinds and occupy their own subspace.

If it is true that kind distributions are a rough representation of the linguistic contexts generally associated with individual members of a group (i.e. if we consider kinds as supremums – see §3.2), existential plurals should be situated somewhere between individuals and kinds in the semantic space. That is, we would assume that the distribution of '*Darcy, Bingley, Collins and Denny*' would highlight some properties common to the individuals but also lose some coherence in virtue of representing their differences, i.e. be closer to a kind. We can in fact illustrate this property by adding name vectors together and inspecting the position of the resulting plural in the semantic space.

Fig. 2 shows that indeed, the plurals '*Darcy and Bingley*', '*Darcy, Bingley and Collins*' and '*Darcy, Bingley, Collins and Denny*' come progressively closer to the kind *man*. Although this result is very tentative and should be shown to be replicable, we take it as support for the idea that there is a principled relation between the distributions of individuals, plurals and kinds (seen as supremums).

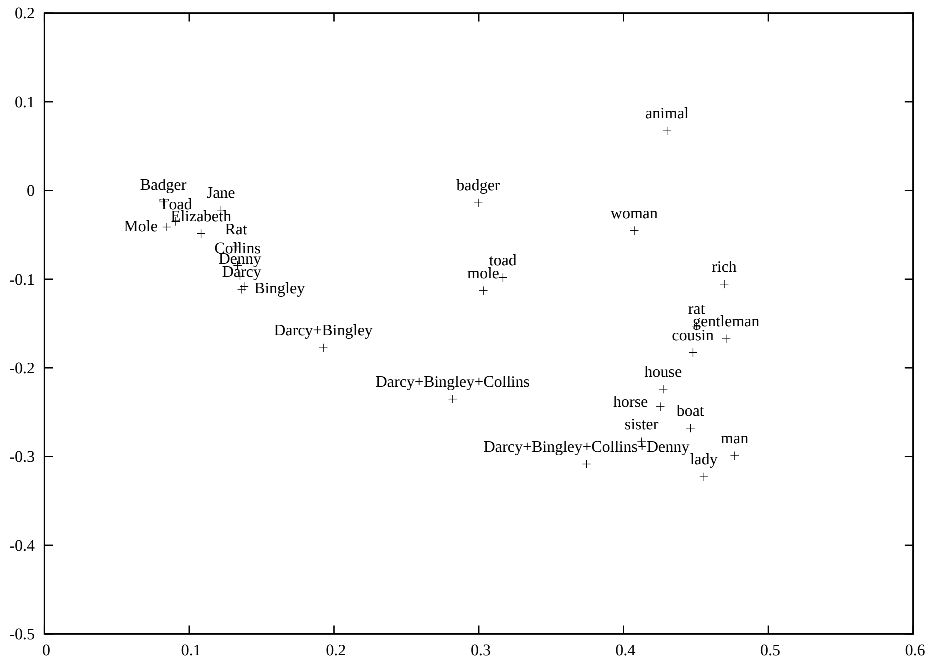

Figure 2: The BNC space with names as contextualised kind distributions. The names have clustered on the left of the space, kinds on the right. Plurals consisting of men's names are roughly situated on a line between the individuals and the concept *man*.

6 Conclusion

In this paper, we have investigated the notion of a distributional proper name and proposed a model which satisfies some natural properties of individuals.

We would like to conclude by adding that we have preliminary results indicating that our contextualisation method can be applied to any type of individual in a co-reference chain with similar effect. That is, the technique can be applied to instances of common nouns (*boat, car, letter*). We intend to pursue this work further, and provide a large-scale evaluation of our proposal involving different types of individuals. We will also integrate the system in a pipeline involving co-reference resolution and test the robustness of this setup.

We believe that having a distributional model of individuals is crucial for several reasons. First, if distributional semantics is to claim psycholinguistic validity, it should account for the fact that many of the words/phrases we use repeatedly (and therefore might build a distribution for) refer to individuals. Consider the name of the city a speaker lives in, the company he/she works for, phrases such as *my boss, Kim's dog*, etc. Second, having access to distributional individuals may help us solve problems that DS has been struggling with. For instance, we may make progress on the topic of antonymy, as (static) antonyms cannot be applied to the same individual (e.g. a city cannot be small and large at the same time). We have also briefly shown that there is potential for developing distributional theories of plurality and genericity by studying the principled relationships between individual, plural and kind distributions.

More generally, we note that our considerations on proper names lead to a less static view of the semantic space. While it is fair to extract distributions from large corpora as lexical representations, the exercise does not teach us much about the way people attribute meaning to new linguistic entities, whether they refer to individuals or concepts (see unknown words, neologisms, second language acquisition, etc). We hope that by viewing the semantic space as a dynamic system, where prior knowledge combined with new linguistic input produces new and updated representations, we can make progress on those issues.

References

Baroni, M., R. Bernardi, and R. Zamparelli (2014). Frege in space: A program of compositional distributional semantics. *Linguistic Issues in Language Technology 9*.

Bruni, E., G. Boleda, M. Baroni, and N.-K. Tran (2012). Distributional semantics in technicolor. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pp. 136–145.

Carlson, G. N. and F. J. Pelletier (1995). *The generic book.* University of Chicago Press.

Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics 6*, 339–405.

Cumming, S. (2013). Names. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

Dinu, G., S. Thater, and S. Laue (2012). A comparison of models of word meaning in context. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT2012)*, pp. 611–615.

Erk, K. (2013). Towards a semantics for distributional representations. In *Proceedings of the Tenth International Conference on Computational Semantics (IWCS2013)*, Potsdam, Germany.

Erk, K. (2014). What do you know about an alligator when you know the company it keeps? Draft.

Erk, K. and S. Padó (2008). A structured vector space model for word meaning in context. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP2008)*, Honolulu, HI, pp. 897–906.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik 100*, 25–50.

Harris, Z. (1954). Distributional structure. *Word 10*(2-3), 146–162.

Herbelot, A. and A. Copestake (2011). Formalising and specifying underquantification. In *Proceedings of the Ninth International Conference on Computational Semantics (IWCS 2011)*, Oxford, UK.

Lenci, A. and G. Benotto (2012). Identifying hypernyms in distributional semantic spaces. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics (*SEM)*, pp. 75–79.

Leslie, S.-J. (2008). Generics: Cognition and acquisition. *Philosophical Review 117*(1), 1–47.

McNally, L. (2014). Kinds, descriptions of kinds, concepts, and distributions. Draft.

Mill, J. S. (1843). A system of logic, ratiocinative and inductive. In J. Robson (Ed.), *Collected Works of John Stuart Mill*, Volume 7-8. Toronto: University of Toronto Press.

Newman, D., J. H. Lau, K. Grieser, and T. Baldwin (2010). Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (ACL-HLT2010)*, pp. 100–108.

Rapp, R. (2004). A practical solution to the problem of automatic word sense induction. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*.

Russell, B. (1911). Knowledge by acquaintance and knowledge by description. In *Proceedings of the Aristotelian Society*, pp. 108–128.

Searle, J. R. (1958). Proper names. *Mind 67*(266), 166–173.

Thater, S., H. Fürstenau, and M. Pinkal (2011). Word meaning in context: A simple and effective vector model. In *Proceedings of the 5th International Joint Conference on Natural Language Processing*, Chiang Mai, Thailand.