# The semantics of poetry: a distributional reading

Aurélie Herbelot

University of Cambridge, Computer Laboratory

J.J. Thomson Avenue, Cambridge

CB1 8AZ

United Kingdom

aurelie.herbelot@cantab.net

February 2, 2014

**Abstract**

Poetry is rarely a focus of linguistic investigation. This is far from surprising, as poetic language, especially in modern and contemporary literature, seems to defy the general rules of syntax and semantics. This paper assumes, however, that linguistic theories should ideally be able to account for creative uses of language, down to their most difficult incarnations. It proposes that at the semantic level, what distinguishes poetry from other uses of language may be its ability to trace conceptual patterns which do not belong to everyday discourse but are latent in our shared language structure. Distributional semantics provides a theoretical and experimental basis for this exploration. First, the notion of a specific 'semantics of poetry' is discussed, with some help from literary criticism and philosophy. Then, distributionalism is introduced as a theory supporting the notion that the meaning of poetry comes from the meaning of ordinary language. In the second part of the paper, experimental results are provided showing that a) distributional representations can model the link between ordinary and poetic language, b) a distributional model can experimentally distinguish between poetic and randomised textual output, regardless of the complexity of the poetry involved, c) there is a stable, but not immediately transparent, layer of meaning in poetry, which can be captured distributionally, across different levels of poetic complexity.

# 1 Introduction

Poetry is not a genre commonly discussed in the linguistics literature. This is not entirely surprising, as poetical language – especially in its contemporary form – is expected to defy the accepted rules of ordinary language and thus, is not a particularly good example of the efficient medium we use to communicate in everyday life.

Still, it would seem misguided to argue that poetry does not belong to the subject of linguistics. It is, after all, made of a *certain type* of language which, at least at some level, relates to ordinary language (it is for instance rare – although not impossible – to read the word *tree* in a poem and find that it has nothing to do with the concept(s) we normally refer to as *tree*).

If we accept that poetical language is not completely dissociated from our everyday utterances, then an ideal linguistic theory should be able to explain how the former can be interpreted in terms of the latter. That is, it should have a model of how poetry uses and upsets our linguistic expectations to produce texts which, however difficult, can be recognised as human and *make sense*.

In this paper, I assume that poetry is a form of language which can be linguistically analysed along the usual dimensions of prosody, syntax, semantics, etc. Focusing on semantics, I investigate whether a particular theory, **distributional semantics (DS)**, is fit to model meaning in modern and contemporary poetry. DS is explicitely based on ordinary language use: the theory assumes that meaning is given by usage, in a Wittgensteinian tradition. So by building a computational model of word meaning over an ordinary language corpus and applying it to poetic text, it is possible to explore in which ways, if any, poetry builds on everyday discourse.

The paper is structured in two parts, one theoretical, one experimental. Starting off with questions surrounding the nature of meaning in everyday language and poetry, I introduce standard semantic theories and some of their historical relationships with particular approaches to poetics. Focusing on distributionalist theories, I then present modern computational work in distributional semantics as a (rough) implementation of the Wittgensteinian account of meaning, and relate this work to experiments in computer-aided poetry dating back to the 1970s. In this process, I argue – with the support of some work in philosophy and literary criticism – that the semantics of poetry does derive from everyday language semantics, and that computational models of ordinary language should let us uncover aspects of poetical meaning.

In the experimental part of the paper, I report on an implementation of a distributional system to compute so-called 'topic coherence', that is, a measure of how strongly words of a text are associated. I show that in terms of coherence, poetry can be quantitatively distinguished from both factual and random texts. More interestingly, the perceived level of difficulty of a text, according to human annotators, is *not* correlated with its overall coherence. That is, complex poetry, when analysed with distributional techniques, is shown to be just as coherent as more conventionally metaphorical texts. I interpret this result as evidence that poetry uses associations which are latent, but usually not explicit, in ordinary language.

# 2 The meaning of poetry

## 2.1 Ordinary language and poetry

One of the most influential theories of meaning in philosophy and linguistics is the theory of reference, as formalised in set-theoretic semantics. The core proposal of set theory is that words denote (refer to) things in the world (Frege, 1892; Tarski, 1944; Montague, 1973). So the word *cat*, for instance, has a so-called 'extension' which is the set of all cats in some world. Set theory is closely related to truth theory in that it is possible to compute the truth or falsity of a statement for a particular world just by looking at the extension of that statement in that world. For instance, the sentence *All unicorns are black* is true if and only if, in our reference world, the set of unicorns is included in the set of black things. The basic notion of extension is complemented by the concept of 'intension' which, under the standard account, is a mapping from possible worlds to extensions, i.e. a function which, given a word, returns the things denoted by that word in a particular world. Intension allows us to make sense of the fact that *Evening Star* and *Morning Star* have different connotations, although they denote the same object in the world: they simply have different intensions.

The question of whether poetry has meaning, and if so, what kind of meaning, has long been debated. An enlightening example of the issues surrounding the discussion can be found in a 1940 exchange between the philosopher Philip Wheelwright and the poet Josephine Miles in the *Kenyon Review* (Wheelwright, 1940; Miles, 1940). Wheelwright wrote an article entitled *On the Semantics of Poetry* where he clearly distinguished the language of poetry from the language of science. According to him, meaning in scientific language was to be identified with conceptual meaning, itself guided by the principles of formal logic and propositional truth. Poetry, on the other hand, was endowed with what he called 'metalogical' meaning, that is, with a semantics not driven by logic. The core of his argument was that signs in poetical language ('plurisigns') were highly ambiguous while words in science ('monosigns') must have the same meaning in all their occurrences. Miles replied to this article with a short letter entitled *More Semantics of Poetry* where she argued that ambiguity could be found in all language; that it was misguided to take scientific language as mostly denotational and poetry as mostly connotational; that some stability of meaning was vital for poetry: "Poetry, as formalizing of thoughts and consolidating of values, works firmly in the material of the common language of the time, limited by its own conventions" (Miles, 1940).

It would be hard to blame Wheelwright for rejecting the thesis that poetry is denotational. Expressions such as *music is the exquisite knocking of the blood* (Brooke, 1911b), *Your huge mortgage of hope* (Hughes, 1999), or *skeleton bells of trees* (Slater,

2008) do not have a natural interpretation in set-theoretic semantics. Still, it also seems difficult to argue that poetical meaning is not related to ordinary language. Without competence in the latter, it is hard to interpret the semantics (if there is any) of a poem.

A position which stays clear from the poetical/ordinary language distinction is that of Gerald Bruns (2005) who argues that "poetry is made of language but is not a use of it". Bruns clarifies this statement by adding that:

> Poetry is made of words but not of what we use words to produce: meanings, concepts, propositions, descriptions, narratives, expressions of feeling, and so on. The poetry I have in mind does not exclude these forms of usage – indeed, a poem may "exhibit" different kinds of meaning in self-conscious and even theatrical ways – but what the poem is, is not to be defined by these things. (p7)

In other words, poetry uses the basic building blocks of ordinary language, but with an aim radically different from the one they are normally associated with. I will call Bruns's position the 'pragmatic' view of poetry, where language is at the core of the investigation, but is deeply dependent on (and playing with) context, intention and meta-linguistic factors.

Pushing the focus of poetry into pragmatics has the advantage that Bruns's account can cover all forms of poetry, including the less 'linguistic' ones (sound poetry in particular), but at first sight, it seems to also lessen the role of semantics – meaning being one of those aspects of language use which poetry is not concerned with. This move, however, is not so clearly intended. In fact, there is a natural bridge between semantics and pragmatics in a theory of meaning which Bruns casually alludes to:

> Basically my approach is to apply to poetry the principle that Wittgenstein applied to things like games, numbers, meanings of words, and even philosophy itself. The principle is that the extension of any concept cannot be closed by a frontier. (p5)

The reference here is to Wittgenstein's *Philosophical Investigations* (1953) and the idea that 'meaning is use', i.e. that meaning comes from engaging in a set of (normative) human practices; in a word, that semantics emerges from pragmatics.

Anchoring semantics in context makes meaning boundaries much less clear than they are in set theory. Still, it is possible to formalise the idea in a linguistic framework. The linguistic theory of meaning closest to Wittgenstein's line of argumentation is distributionalism. In this approach, the meaning of *cat* is not directly linked to real cats but rather to the way people talk about cats. The collective of language users acts as a normative force by restricting meaning to a set of uses appropriate in certain pragmatic situations. The roots of distributionalism can perhaps be found in Bloomfield (1933), but the theory grew to have much influence in the 1950s (Harris, 1954; Firth, 1957). Some time later, in the 1990s, the advent of very large corpora and the increase in available computing power made the claims (to some extent) testable. Both psychologists and linguists started investigating the idea that a word's meaning could be derived from its observed occurrences in text (Landauer & Dumais, 1997; Grefenstette, 1994; Schütze, 1998). These empirical efforts would soon lead to a very active area of research in computational linguistics called 'distributional semantics': a field which attempts to model lexical phenomena using 'distributions', i.e. patterns of word usage in large corpora.

Interestingly, there are historical links between Wittgenstein's distributionalism, distributional semantics and computer-generated poetry. One of Wittgenstein's students, Margaret Masterman, was very influenced by his theory of meaning and by the idea that studying language use could give an insight into semantics. Foreseeing the potential of applying computers to this type of philosophical and linguistic investigation, she founded the Cambridge Language Research Unit (CLRU), a research group which would become engaged in early computational linguistics work in the UK. In parallel, she also took an interest in the creative processes involving language and produced an early version of a computer program to support poetry generation.[1] The program was not actually producing poetry but rather presenting word choices to the user, allowing them to fill in a preset haiku frame. Masterman's idea of using computers to produce poetry was not to replace the human poet. In fact, she clearly differentiated the 'real' poet from the machine: "The true poet starts with inspired fragments, emerging fully formed from his subconscious" (Masterman, 1971). She also didn't believe that randomness could be a foundation for poetry:

> To put a set of words on disc in the machine, program the machine to make a random choice between them, constrained only by rhyming requirements, and to do nothing else, this is to write idiot poetry. [...] In poetry, we have not as yet got the generating formulae; though who would doubt that a poem, any poem, has in fact an interior logic of its own?

Masterman thought that there was a structure underlying language use. Uncovering that structure formed an important part of the work at the CLRU. Computing resources in those days were extremely limited so, instead of directly studying linguistic utterances in their natural environment, part of the CLRU's research focused on producing so-called 'semantic networks' by analysing thesauri (Spärck Jones, 1964). Still, this work prefigured what would become distributional semantics, and the automatic construction of 'semantic spaces' out of statistical information from real language data (see §2.2).

The notion of a semantic structure extractable from language data by computational means was also behind Masterman's belief that machines could support the work of poetry:

> Larger vocabularies and unusual connexions between the words in them, together with intricate devices hitherto unexplored forms of word-combination, all these can be inserted into the machine, and still leave the live poet, op-

erating the console, free to choose when, how and whether they should be employed

It is possible to summarise Masterman's position as follows: poetry is not random, but the stuff of poetry, the 'inspired fragments' found in the subconscious of the poet, are already there, latent in language use, and an appropriate semantic theory should be able to uncover them. This is compatible with Miles's argument that poetry is anchored in ordinary language, and also of Bruns's reading of Lyn Hejinian's poetics (2000): "The poet [...] does not so much use language as interact with uses of it, playing these uses by ear in the literal sense that the poet's position with respect to language is no longer simply that of the speaking subject but also, and perhaps mainly, that of one who listens." (p30)

Fifty years after the first CLRU experiments on distributional semantics, computational linguistics is still working towards the perfect model of meaning that Masterman wished for. Further, little has been done to linguistically formalise the relation between the semantics of ordinary language and that of poetry. In what follows, I will attempt to capture this relation: first, intuitively, by discussing examples of poetry based on distributional semantics models (§2.2); and later, more formally, by giving experimental evidence that distributional models constructed from ordinary language can account for (at least) a layer of meaning in modern and contemporary poetry (§3).

## 2.2 Distributional Semantics and Poetry

The core assumption behind distributional semantics is that meaning comes from usage. A fully distributionalist picture includes both linguistic and non-linguistic features in the definition of 'usage'. So the context in which an utterance is observed comprises not only the other utterances that surround it, but possibly also sensorial input, human activities and so on. Although research on distributional semantics is slowly starting to include visual features in its study of meaning (e.g. Feng & Lapata, 2010), I will concentrate here on the bulk of the work which makes the simplifying assumption that the meaning of a word can be defined in terms of its close linguistic context.[2] The representation of a lexical item in the framework is a vector, or simply put, a list of numbers. So the meaning of *dragon* – its so-called ‘distribution’ – might look like this:

| | |
|---|---|
| $dungeon$ | 0.8 |
| $eat$ | 0.1 |
| $fire$ | 0.3 |
| $knight$ | 0.5 |
| $political$ | 0.001 |
| $scale$ | 0.08 |
| $very$ | 0.0001 |

The numbers represent the average strength of association of the lexical item with other words appearing in its close context (say, a window of 10 words around its occurrences in a large corpus). There are many ways to compute strength of association – for a technical introduction, I refer to Turney & Pantel (2010) and Evert (2004). I will assume here the use of measures which give strong weights to ‘characteristic contexts’ (e.g. Pointwise Mutual Information, PMI). In such a setting, a word which appears frequently with a lexical item $t$ and not so frequently with other items has a strong association with $t$ (e.g. *meow* with respect to *cat*); a word which appears frequently with $t$ but also very frequently with other things has low association with $t$ (e.g. *the* with respect to *cat*); a word which does not appear frequently with $t$ also has low association with $t$ (e.g. *autobiography* with respect to *cat*).

To come back to our example, the numbers in the dragon vector tell us that dragons are strongly related to dungeons and knights, but only moderately to eating and fire (because a lot of other animals eat and fire has a strong relation to burning houses and firemen). It also shows that they are moderately related to scales, although the

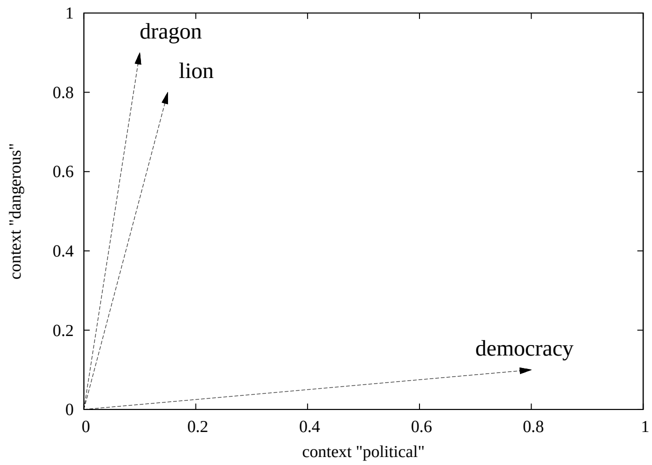


Figure 1: An example of a semantic space with two dimensions, *dangerous* and *political*

prototypical dragon is a scaly creature, because of the polysemy of scale (meaning, for instance, a measurement range). Finally, dragons are not strongly associated with *very* at all, due to the fact that it is such a common word. Contexts with high association figures are said to be 'characteristic' of the lexical item under consideration.

Being a vector, the distribution of *dragon* can be represented in a mathematical space. Such a space, where dimensions correspond to possible contexts for a lexical item, is commonly called a 'semantic space'. One of the benefits of this representation is that similar words can be shown to naturally cluster in the same areas of the semantic space. Typically, the vectors for *dragon* and *lion* will end up close together in the semantic space while *dragon* and *democracy* are much further apart, confirming Harris's hypothesis that 'similar words appear in similar contexts' (1954). Fig. 1 shows a highly simplified illustration of this effect, in a two-dimensional space where words are expressed in terms of the contexts *dangerous* and *political*.

It would be beyond the scope of this work to describe the range of phenomena which can be modelled using approaches based on the framework introduced here. But some examples will be helpful in showing its relevance to semantics as a whole. Distributions, for instance, can be further analysed to reveal the various senses of a word: the vector for *scale* can be combined with its close context via various mathematical operations to produce a new vector distinguishing, say, scales as measurement, scales as weighing machines and dragon scales (Schütze, 1998; Erk & Padó, 2008; Thater et al., 2010). They also capture some inferential properties of language related to hyponymy (e.g. if Molly is a cat, Molly is an animal) and quantification (e.g. *many cats* entails *some cats*) (Baroni et al., 2012). It is even possible to derive compositional frameworks which show how the lexical meaning of individual words combine to form phrasal meaning (see Erk, 2012 for an overview). But this type of work is generally evaluated against human linguistic judgements over ordinary language utterances. So can it tell us anything about poetical language?

I will first approach the question intuitively and consider the features of a poetry built out of distributional representations, in the way Masterman envisaged. *discourse.cpp* (le Si, 2011), a little volume of poems mostly deriving from distributional techniques, will provide suitable examples for my observations. The texts in *discourse.cpp* are more or less edited versions of two kinds of output: the first one consists of words that are similar to a given input (for instance, *dog* or *horse* for the input *cat*), while the second one is a list of 'characteristic contexts' for the input (for instance, *meow* or *purr* for *cat*). The background corpus for the system was a subset of 200,000 Wikipedia[3] pages, fairly small by the standards of 2014, but sufficient to produce the semantic clustering effects expected from a distributional framework. Context was taken to be the immediate semantic relations in which a given lexical item appeared – that is, instead of just considering single words around a target in the text, the system relied on syntactic and semantic information describing 'who did what' in a sentence. For instance, in the statement *The black cat chased the grey mouse*, the context of *chase* would be marked as '– ARG1 cat' (*cat* as first semantic argument of the verb) and '– ARG2 mouse' (*mouse* a second semantic argument of the verb) while the context of

**Illness**

S/he nearly died of a psychosomatic food-borne psychotic-depressive near-fatal
episodic epidemic, diagnosed as HIV-related
and

naturally

undisclosed.

Figure 2: Illness, *discourse.cpp*, O.S. le Si (2011)

mouse would be 'grey ARG1 –' (*grey* as semantic head) and 'chase ARG2 –' (chase a semantic head).[4]

One straightforward example of the program's output is the poem *Illness* (Fig. 2), produced using some of the characteristic contexts of the lexical item *illness*. The editing of this poem, as reported in the appendix of *discourse.cpp*, involved adding coordinations, prepositions and punctuation to the raw output, together with the words *S/he nearly died of a* and *naturally*. The adjective *epidemic* was substantivised.

Unsurprisingly, concepts which have a strong association with *illness* are adjectives such as *psychosomatic* or *diagnosed*. Even in this simple example, however, it is clear that some aspects of 'the discourse' (i.e. the way that people choose to talk about things) is reflected: given the number of very various conditions and illnesses in medical dictionaries, it is striking that *HIV-related* makes it into the top contexts, and we can hypothesise that it explains the presence of the adjective *undisclosed*. In other words, despite the range of medical conditions experienced in everyday life, it is HIV which dominated the thoughts of the Wikipedia contributors responsible for the pages constituting the *discourse.cpp* corpus, and not the common cold or malaria.

More distant – but fully interpretable – associations are found throughout *discourse.cpp*. *Politics*, for instance, is compared to the Japanese puppet theatre Bunraku, probably picking up on the wide-ranging, disenchanted view of government as ‘a circus’. *Pride*, although it does not involve any direct metaphorical association, is pointedly described as a list of ‘status symbols’:

> Pride is your clothes,
> your girlfriend,
> a meal.

Less obvious connections are also found. The poem *The Handbag* is a list of objects commonly found in women’s handbags. The last item in the list, however, is the noun *household*. Whether there is a natural interpretation for this association can be debated, but it picks out a relationship between the handbag and the notion of a home – perhaps a sense of safety, or else a ‘realm’ over which control is exerted.

It is probably clear that *discourse.cpp* is not computer-generated poetry, in the sense that human input is removed. The presentation of the book, its materiality, the typesetting, and of course the editing of the poems were human choices.[5] Calling upon the notion of ‘intentionality’, Emerson (2008) reminds us that the programmer who gets a computer to output data for the aim of producing poetry remains the driving creative force behind the enterprise. From a linguistic point of view, however, the intention of the programmer may be read in terms of pragmatics, as a speech act (Searle, 1969), i.e. as an act of communication with a particular goal. It does not preclude the semantics of the finished product – the meaning produced by the composition of particular lexical items – to come from a fully computational model of part of language.[6]

So we may have traces of a computational semantics of ordinary language in *discourse.cpp*, but is it poetical semantics? I have tried so far to argue, with Masterman, that the ‘structure of language’ – the distributional semantics space–, together with its ‘unusual connexions’ and ‘unexplored forms of word-combination’, can form the basis of poetry production. But is the output comparable to what an actual poet would have produced?

At this point, it may be helpful to think of semantics not as something that texts *have*, but as something that people *do* with texts. If in distributionalism, meaning is 'the use of a word', or 'the things the word is *associated* with', then producing/finding meaning is about producing/finding associations (see Hofstadter, 2001 for the related argument that cognition is anchored in analogy). Arguably, it is impossible for a speaker of a language not to associate when presented with a word sequence: whether speaking/writing or hearing/reading, they are drawn towards specific individual and cultural conceptual connections. It can be shown, in fact, that the neurological response of an individual presented with a word or word sequence includes an activation of relevant associations. Molinaro et al. (2012) write: *while composing the meaning of an expression, comprehenders actively pre-activate semantic features that are likely to be expressed in the next parts of the sentence/discourse*. From this, it follows that:

1. human poetry, however complex, should always be experimentally distinguishable from randomised word sequences, where the latent structure of language is ignored;

2. a certain level of associativity should be identifiable in all human-produced poetry, regardless of complexity (i.e. both a semantically opaque poem and a more straightforward text will make use of the underlying, shared structure of language).

The next section puts this hypothesis to the test by using a distributional model of semantics to quantify the associational strength of a range of poems, as well as random and factual texts.

# 3 Semantic Coherence in Modern and Contemporary Poetry

If we are to show that the semantics of poetry uses the structure of ordinary language to produce meaning, we need to demonstrate that a computational model built on non-poetic language can account for at least some aspects of that semantics, regardless of the apparent difficulty of the text under consideration.

I will now turn to the issue of *topic coherence*, a measure of the semantic relatedness of the items in a given set of words. Topic coherence has been studied from the point of view of so-called 'topic modelling' techniques, that is, computational methods that take a set of documents and classify them within particular topics (e.g. Mimno et al., 2011). But the proposed measures can be applied to any set of words, and might for instance highlight that the set *chair, table, office, team* is more coherent than *chair, cold, elephant, crime*. As such, it is well suited to model the general strength of semantic association in a text.

I will investigate topic coherence in a number of poems written in the period 1881-2008. The general idea is to compare texts of varying 'difficulty' (from metaphorical but transparent lyrics to opaque, contemporary poetry) and analyse how they behave in terms of coherence, using distributions extracted from ordinary language corpora as word representations. Intuitively, it seems that more complex fragments such as *the reaches of turning aside remind* (Coolidge, 1990) should be less coherent than transparent verses such as *The grey veils of the half-light deepen* (Brooke, 1911a). As argued in the last section, however, we are looking for a stable level of associativity across all poetry. Our model should capture associations of (roughly) equal strength in transparent and opaque fragments, making explicit connections which a human reader might not consciously recognise when first presented with a text.

Following Newman et al. (2010), I define the coherence of a set of words $w_1 ... w_n$

as the mean of their pairwise similarities:

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

For example, if we were to calculate the coherence of *the reaches of turning aside*, we would calculate the similarities of *reach* with *turn*, *reach* with *aside* and *turn* with *aside*, and average over the three obtained scores, ignoring closed-class words.

The representations for single words are distributionally obtained from the British National Corpus (BNC). The corpus is lemmatised and each lemma is followed by a part of speech according to the CLAWS tagset format (Leech et al., 1994). For the experiments reported here, parts of speech are grouped into broad classes like N for nouns or V for verbs. Furthermore, I only retain words in the following categories: nouns, verbs, adjectives and adverbs (punctuation is ignored). Each text/poem is converted into a 11-word window format, that is, context is defined by the five words preceding and the five words following the target word.

To calculate co-occurrences, the following equations are used:

$$\mathit{freq}_{c_i} = \sum_t \mathit{freq}_{c_i,t} \qquad \mathit{freq}_t = \sum_{c_i} \mathit{freq}_{c_i,t} \qquad \mathit{freq}_{total} = \sum_{c_i,t} \mathit{freq}_{c_i,t}$$

The quantities in these equations represent the following:

| | |
|---|---|
| $\mathit{freq}_{c_i,t}$ | frequency of the context word $c_i$ with the target word $t$ |
| $\mathit{freq}_{total}$ | total count of word tokens |
| $\mathit{freq}_t$ | frequency of the target word $t$ |
| $\mathit{freq}_{c_i}$ | frequency of the context word $c_i$ |

The weight of each context term in the distribution is given by the function suggested in Mitchell & Lapata (2010) (PMI without log):

$$v_i(t) = \frac{p(c_i|t)}{p(c_i)} = \frac{freq_{c_i,t} \times freq_{total}}{freq_t \times freq_{c_i}} \tag{2}$$

Finally, the 2000 most frequent words in the corpus are taken as the dimensions of the semantic space (this figure has been shown to give good performance in other studies: see again Mitchell & Lapata, 2010).

The similarity between two distributions is calculated using the cosine measure:

$$Sim(A,B) = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \tag{3}$$

where $A$ and $B$ are vectors and $i...n$ are the dimensions of the semantic space.

## 3.1 Experimental setup

Eight poems were selected (all written in modern English), intended to cover a range of ‘difficulty’. That is, some have a straightforward meaning while others require much more interpretation. Two additional texts were added to the sample: one is a subset of a Wikipedia article while the other is randomly-generated by piecing together words from the British National Corpus and inserting random punctuation. These two texts were meant to provide an upper and lower bound on associativity (the assumption being that a factual text makes heavy use of the more common turns of phrase in language). Table 3.1 gives an impression of the content of the sample by showing the beginning of each text.

The 10 texts in the sample are fairly intuitively categorisable into various degrees of complexity. To confirm this, the author and two independent annotators attributed a ‘difficulty’ score in the range 1-5 to each text (where 1 = very easy to understand and 5 = very hard to understand). To help the annotators with the task, they were first presented with simple questions regarding the topic of the text:

*What is this poem about?*

| Author | Year | Excerpt | |
|---|---|---|---|
| Rupert Brooke | Day that I have loved | 1911 | Tenderly, day that I have loved, I close your eyes,/ And smooth your quiet brow, and fold your thin dead hands. |
| Coolidge | Argument over, Amounting | 1990 | In edges, in barriers the tonal light of t/ the one thing removed overemphasizes tonally/ and you could hurry it, and it vanish and plan |
| Carol Ann Duffy | Valentine | 1993 | Not a red rose or a satin heart./ I give you an onion./ It is a moon wrapped in brown paper. |
| Allen Ginsberg | Five A.M. | 1996 | Elan that lifts me above the clouds/ into pure space, timeless, yea eternal/ Breath transmuted into words/ Transmuted back to breath |
| MacCormack | At Issue III | 2001 | Putting shape into getting without perfect in a culture that doesn't think, pumps up, the/ two traits go at the face of rate themselves |
| Avery Slater | Ithaca, Winter. | 2008 | Creaking, skeleton bells of trees/ dissolve in a quilt of pale flurries. |
| Gertrude Stein | If I Told Him, A Completed Portrait of Picasso | 1924 | If I told him would he like it. Would he like it if I told him./ Would he like it would Napoleon would Napoleon would would he like it. |
| Oscar Wilde | In The Gold Room – A Harmony | 1881 | Her ivory hands on the ivory keys/ Strayed in a fitful fantasy,/ Like the silver gleam when the poplar trees/ Rustle their pale-leaves listlessly |
| Wikipedia | 'The Language Poets' | ? | The Language poets (or L=A=N=G=U=A=G=E poets, after the magazine of that name) are an avant garde group or tendency in United States poetry that emerged in the late 1960s and early 1970s. |
| Random text | Psychologist. Strong. | - | tabard, battersea, wolf, coma, acas. hutchinson cap'n. suet. ellesmere. proportionality/ mince. outside, morey folk, cum, willoughby, belligerent, dimension |

Table 1: Excerpts from the sample

| | Author | Annotator 1 | Annotator 2 | Average |
|---|---|---|---|---|
| random | 5 | 5 | 5 | 5.00 |
| MacCormack | 5 | 5 | 5 | 5.00 |
| Coolidge | 4 | 5 | 5 | 4.67 |
| Ginsberg | 5 | 4 | 3 | 4.00 |
| Stein | 5 | 3 | 3 | 3.67 |
| Slater | 2 | 3 | 4 | 3.00 |
| Brooke | 2 | 4 | 3 | 3.00 |
| Wilde | 1 | 1 | 2 | 1.33 |
| Duffy | 1 | 1 | 2 | 1.33 |
| Wikipedia | 1 | 1 | 1 | 1.00 |

Table 2: Difficulty scores for each text in sample

*How confident are you of your answer? (1=not confident at all, 5=absolutely confident)*

*What is the main emotion conveyed by the poem? (e.g. anger, love, disappointment, etc)*

*What are the main images in the poem? (e.g. some people talking, the sun, a busy street, etc)*

*How did you like the poem? (1=not at all, 5=a lot)*

*How difficult did you find it to understand the poem? (1=very easy, 5=very difficult)*

The average Spearman correlation between annotators was 0.81, indicating very good agreement. Table 3.1 shows individual scores for the three annotators, as well as an average of those scores. The table is sorted from the most to the least difficult text.

As expected, the Wikipedia article is annotated as being the most transparent text, while the random poem is considered the most difficult (on a par, however, with MacCormack's 'At Issue III'). When told that one of the texts was randomly produced by a computer, the two independent annotators were able to identify it but also indicated that MacCormack's poem had caused some hesitation.

The poems are POS-tagged with TreeTagger[7], and the tagging is manually checked.

Coherence is calculated between the words in each sentence for poems which have a clear sentence structure (Brooke, Duffy, Slater, Stein, Wilde, the random text and Wikipedia article). The other poems are split into fragments corresponding to the average sentence length in the texts made of sentences. Only content words (nouns, verbs, adjectives and some adverbs) with a frequency over 50 in the BNC are considered: the frequency threshold ensures that good-quality distributions are extracted. For the calculation of coherence, very frequent adverbs and auxiliaries are also disregarded (e.g. *so, as, anymore, be*). In total, 608 distributions are extracted from the BNC, covering around 72% of all content words in the sample. The average sentence length comes to 4 content words.[8] When a word is repeated within a fragment, the similarity of that word with itself is *not* included in the coherence calculation, so that poems with a high level of repetition do not come out as being particularly coherent.

Once coherence figures have been calculated for all sentences/fragments in a text, the average of these figures is taken to give an overall coherence measure for the text.

## 3.2 Results

Table 3.2 shows the average semantic coherence of our ten texts, together with the mean and standard deviation for the sample. Fig. 3 shows the results as a graph.

The horizontal line going through the graph shows the mean of the coherence values, while the greyed out areas highlight the points contained within the standard deviation. The figure clearly shows that the random text and the Wikipedia article are outside of the standard deviation, as would be expected. Randomness results in much lower coherence than for the human-produced poetry, and the factual text displays greater coherence.

To confirm that the sampled poetry could generally be distinguished from both factual and random texts, 8 other texts were introduced (4 random, 4 factual in the form of the first paragraphs of Wikipedia article related to poetry) and their average

| Poem | Average coherence |
|---|---|
| random | 0.17 |
| Slater | 0.23 |
| Duffy | 0.25 |
| Wilde | 0.25 |
| MacCormack | 0.32 |
| Ginsberg | 0.33 |
| Brooke | 0.35 |
| Stein | 0.35 |
| Coolidge | 0.38 |
| Wikipedia | 0.43 |
| MEAN | 0.31 |
| STDV | 0.08 |

Table 3: Semantic coherence for each text in sample

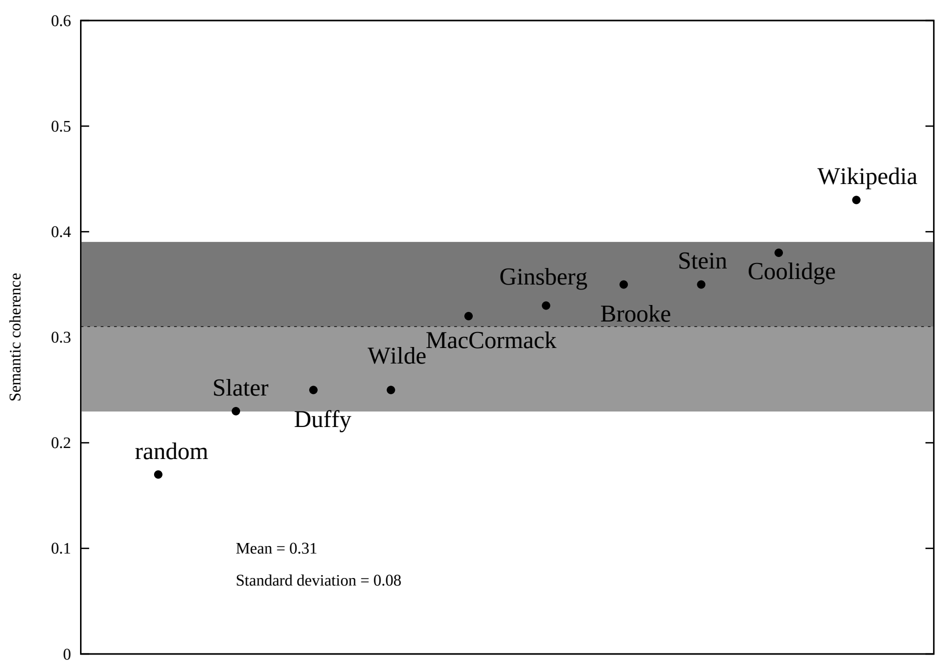


Figure 3: Semantic coherence plot for the 10-text sample.

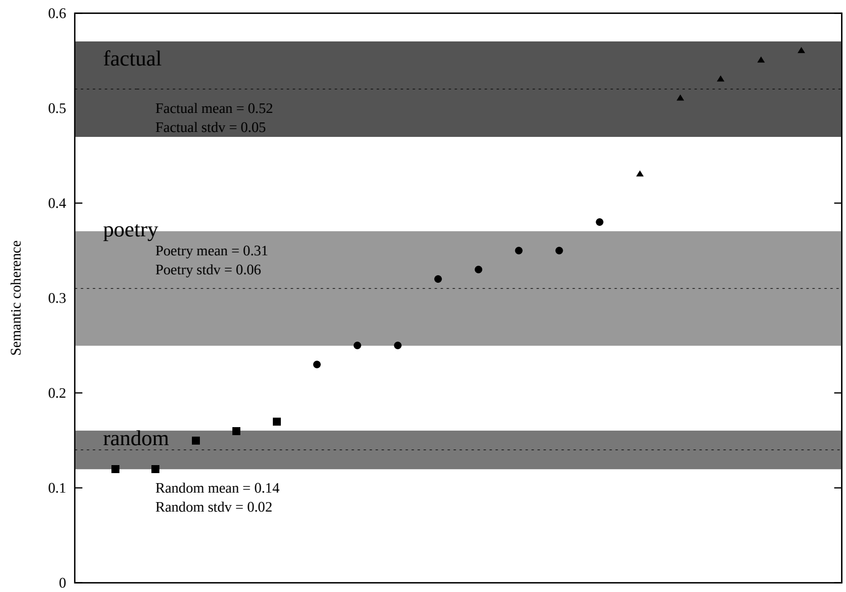


Figure 4: Coherence range for poetry versus random and factual texts

coherence computed. The effect is confirmed, with the coherence of random texts lying in the range [0.12-0.17] and the coherence of the Wikipedia texts in the range [0.43-0.56]. Fig. 4 shows the means and standard deviations of the three types of text. The differences are statistically significant at the 1% level.

These results show that, as hypothesised, human-produced poetry can clearly be differentiated from random texts, even in cases where a human reader might hesitate (e.g. MacCormack's 'At Issue III'). But they also indicate a significant difference between factual and poetic writing. The generally lower coherence of poetry compared to factual prose can presumably be put down to both the linguistic creativity and the greater metaphorical content of the texts. Despite *journey* being a conventional metaphor for *life*, for instance, we cannot expect their distributions to be as similar as, say, *journey* and *travel* because their overall pattern of use differs in significant ways. The creativity of the poet is also at work in that he or she might pick out unheard combinations which, although they make use of the underlying structure of language, do not score so highly in terms of distributional similarity (see §3.3).

Notably, the poems have a standard deviation similar to the factual texts, indicating no marked difference between individual poems, despite their obvious variety in complexity. There is also no correlation between the perceived difficulty of the poems, as given by the annotators, and their semantic coherence. In the top range, we find Coolidge and Brooke close together, despite the fact that Coolidge is apparently fairly opaque and Brooke generally transparent. At the other end of the range, Duffy and Wilde, perceived to be generally 'easy', come slightly below the mean. This disproves that semantic coherence is linked to perceived complexity and thus, the thesis that 'there is not much semantics in complex poetry'.

When looking closer at the results, we find that MacCormack's *play arrived, how large in prompting* is as coherent as Wikipedia's opening sentence *The Language poets [...] are an avant garde group or tendency in United States poetry that emerged in the late 1960s and early 1970s* (both have coherence 0.47). This may well seem puzzling, but again, a detailed analysis of the distributions involved shows that some semantics is clearly shared between the words of MacCormack's fragment. Table 3.2 shows the word pairs involved in the coherence calculation, together with their cosine similarities and all the contexts they share in the most highly weighted subset of their distributions (the shared terms must have a weight of at least 0.2 in both distributions). I have grouped contexts by topic where possible, to make results more readable.

Several topics emerge across the captured contexts. A first one covers performance arts and their audience (*ticket, scene, audience, performance*, etc). A second one concerns temporality (*Wednesday, minute, finally, June*, etc.). A third one relates to policing and violence (*police, troop, army, violence, dominate*, etc). We also find, perhaps less evidently, a topic about news (*tv, story, news*). Tellingly, these themes are echoed in other parts of the poem: we find *gun, combat, push, violence* close to the fragment under consideration, *day, year, postpone, temporary, minute, hour* across the text, *story,*

| word pair | similarity | shared contexts |
|---|---|---|
| play_N–arrive_V | 0.49 | {ticket_N scene_N studio_N hall_N} {australia_N africa_N} {wednesday_N minute_N regularly_A} |
| play_N–large_A | 0.44 | {audience_N hall_N} area_N flat_A dominate_V |
| play_N–prompt_V | 0.43 | {tv_N audience_N performance_N success_N} {write_V version_N scene_N story_N} {violence_N dominate_V} {move_N run_N} united_A rain_N |
| arrive_V–large_A | 0.56 | {ship_N station_N island_N} {crowd_N hall_N} |
| arrive_V–prompt_V | 0.48 | {police_N troop_N army_N warning_N} {finally_A eventually_A june_N march_N weekend_N} {flight_N visit_N} {paris_N germany_N} news_N scene_N miss_N william_N couple_N |
| large_A prompt_V | 0.46 | {audience_N gather_V} {firm_N organization_N europe_N} {coal_N plastic_N fruit_N} complex_A volume_N dominate_V |

Table 4: Similarities and shared contexts for the fragment *play arrived, how large in prompting*

*celebrity, television, radio, glamorous, camera* also throughout the text. Even the apparently unconnected *fruit*, which appears in the shared contexts of *large* and *prompt*, occurs two words after our fragment in the poem.

It is worth pointing out that, although the highlighted shared contexts are amongst the most highly weighted for the corresponding distributions (the 0.2 threshold means that we are effectively considering the top 7% of the *play*, *arrive* and *large* distributions, and the top 13% of *prompt*), they are not the most salient features *overall* for those words. That is, they probably do not correspond to features that a native speaker of English would readily associate with *play*, *arrive*, *large* or *prompt*. However, a closer inspection of the type usually practised by literary criticism would certainly uncover such associational threads in the poem. In other words, if meaning is not immediately present when reading the poem for the first time, it is also not closed to the reader able to disregard the broader pathways of the semantic space.

## 3.3 Making Sense

In linguistics, the term 'semantic transparency' is used to refer to how easy or difficult it is for the speaker of a language to guess what a particular combination of words means. According to Zwitserlood (1994), "[t]he meaning of a fully transparent compound is synchronically related to the meaning of its composite words". So a noun phrase like *vulnerable gunman* might be said to be semantically transparent while *sharp glue* would not (Vecchi et al., 2011). Transparency is not directly related to acceptability in language. Some well-known linguistic compounds are not 'compositional', that is, the meaning of the compound is not given by the meaning of their parts (e.g. *ivory tower*), but they are usually fixed phrases which are frequent enough that their meaning is learnt in the same way as the meaning of single words. The line between semantically transparent and opaque phrases is naturally very blurred. Bell & Schäfer (2013) point out, for instance, that sharp glue, which is neither transparent nor fixed in English, could be glue with a pH less than 7, i.e., they can come up with an interpretation for a noun phrase without an obvious compositional – or previously known – meaning.

The study of poetry has consequences for general linguistics. It may be possible to say that semantic transparency is not a fixed attribute of a word combination, but rather a state in the mind of the hearer. ‘Making sense’ of a text, or ‘doing semantics’, becomes the process of making the text more transparent by investigating less-travelled pathways in the semantic space. In the same way that we are aware of the similarity between cats and mongooses – even though we hardly ever encounter the utterance *Cats and mongooses are similar* – it is likely that we capture ‘hidden’ relations in the semantic space, leading us to recognise the connection between *handbag* and *household*, between *pride* and *girlfriend*, or again between *mortgage* and *hope* (that which can be lent and taken away). The task of ‘making sense’ of poetry may then be seen as a type of disambiguation, where the dimensions of a word’s distribution are re-weighted to take context into account (see e.g. Erk & Padó, 2008 for a distributional model of sense disambiguation).

A last word should be reserved for the study of linguistic creativity. Although a large body of work exists on the topic of modelling metaphorical language (see Shutova, 2010 for an overview), the study of poetical semantics has not been a focus of investigation so far. In spite of this, the claims that apply to metaphor and other well-studied productive phenomena can arguably be made for more complex creative processes: simply put, there is nothing in the present paper that would invalidate the claim that creativity in language can be traced back to its very ordinary use (see Veale, 2012 for an extensive, computationally-based account of this). Language can be seen as the result of profoundly individual and yet ultimately collective phenomena. The semantics we ascribe to very mundane objects like cups and mugs varies widely, depending on speakers (Labov, 1973). Still, speakers of a language communicate effortlessly with each other, and general evolutionary effects can be observed in any language, be it at the phonetic, syntactic or semantic level. Distributions capture the common denominator which allows communication to take place. They are in essence a model of the 'normative' effects of language use: the repeated utterance of a word in a particular type of context, across a wide range of speakers, fixes its meaning in a way that makes its usage predictable and fit for successful communication. Each new utterance entering the language contributes to these norming effects by either reinforcing the status quo or, possibly, modifying it – thereby accounting for language change. Now, if ordinary language is a collective construction, so is its underlying semantic structure and we could expect the latent conceptual associations in this structure to be roughly shared across a specific language and cultural background. The hidden, uncommon associations invoked in poetical semantics may be said to come from the very normative force of everyday speech.

# 4 Conclusion

In his literary criticism, Richards (1970) suggests the existence of an intuitive process which guides the poet towards particular linguistic combinations.:

> The poet [...] makes the reader pick out the precise particular senses required from an indefinite number of possible senses which a word, phrase or sentence may carry. The means by which he does this are many and varied. [...] [T]he way in which he uses them is the poet's own secret, something which cannot be taught. He knows how to do it, but he does not himself necessarily know how it is done. (p.32)

A possible linguistic translation of this intuition, based on distributionalism, is to say that the poet, as a speaker of a language, has access to its semantic structure. The 'secret' hypothesised by Richards is perhaps simply the special skill of some individuals to analyse that structure. A poet's work provides exemplars of his/her observations, where the observed data consists of many actual snippets of language use, placing the work of poetry within a collective linguistic intuition.

Using insights from computational linguistics, we can model the ways in which certain types of poetical output might emerge. The implementation of such models follows some prior claims about the nature of language (Wittgenstein, 1953), about semantic structure and poetry (Masterman, 1971), and about the connection between ordinary language and poetical expression (Miles, 1940). In this paper, I argued that:

1. assuming a distributional view of meaning, it is possible to show the relation between ordinary language and the 'extraordinary' language of poetry;

2. the distributional model clearly captures the distinction between human and randomised production, regardless of the immediate semantic transparency of the text;

3. the distributional model shows a stable layer of semantic associativity across poems, regardless of complexity.

A natural next step for the investigation presented here would be to explore the annotators' judgements on semantic complexity. It is unclear what exactly makes a fairly straightforward text such as Duffy's 'Valentine' comparatively less coherent than the complex 'Argument over. Amounted.' by Coolidge. A more fine-grained analysis of the results would be necessary to make any hypothesis.

As a final note, it may be worth pointing out that, although 'big data' has so far mostly been used for the analysis of large phenomena in the digital humanities, this paper shows that one of its incarnations (distributional representations) may have a role to play as a background linguistic model for close reading.

# Notes

[1]An example output can be seen at http://www.chart.ac.uk/chart2004/papers/clements.html.

[2]This must not invalidate distributional semantics techniques as essentially Wittgensteinian constructs. A corpus which is coherent from the point of view of 'speech acts' (Searle, 1969) can be seen as a particular language game: the meaning representations obtained from it are just specific to that language game. So for instance, the online encyclopedia Wikipedia might be said to collate language games where one participant gives information about a particular topic to a hearer, in a regulated written form.

[3]http://www.wikipedia.org/

[4]See Dowty et al. (1981) for an introduction to Montague semantics and a description of verbs and adjectives as 'functions' taking arguments.

[5]See Emerson (2012) for a review of *discourse.cpp* covering this issue.

[6]There is a subtlety here. In a distributional account, the semantics of words comes from pragmatics, that is, from an indefinite number of situations where, collectively, words are used in particular situations, with a particular intent. These situations are separate from (or more accurately, a very large superset of) the pragmatic situation and intent behind the creation of a specific poem.

[7]Available at http://www.cis.uni-muenchen.de/∼schmid/tools/TreeTagger/

[8]This fairly short sentence length is due to Stein's poem containing many single word sentences. However, experiments with different fragment lengths (up to 10 content words) did not signicantly change the results reported here.

# References

Baroni, M., Bernardi, R., Do, N.-Q., & Shan, C.-c. (2012). Entailment above the word level in distributional semantics. In *Proceedings of the 15th conference of the European Chapter of the Association for Computational Linguistics (EACL12)*.

Bell, M. J., & Schäfer, M. (2013). Semantic transparency: challenges for distributional semantics. In *Proceedings of the 'Towards a Formal Distributional Semantics' workshop (collocated with IWCS2013, Potsdam, Germany)*.

Bloomfield, L. (1933). *Language*. New York: Henry Holt.

Brooke, R. (1911a). Day that I have loved. In *Poems*. Sidgwick & Jackson.
URL http://www.rupertbrooke.com/poems/1905-1908/day_that_i_have_loved

Brooke, R. (1911b). The fish. In *Poems*. Sidgwick & Jackson.

Bruns, G. (2005). *The material of poetry: sketches for a philosophical poetics*. University of Georgia Press.

Coolidge, C. (1990). Argument over, amounting. In *Sound as thought: Poems 1982-1984*. Green Integer.
URL http://www.poetryfoundation.org/poem/243682

Dowty, D. R., Wall, R., & Peters, S. (1981). *Introduction to Montague semantics*. Springer.

Duffy, C. A. (1993). Valentine. In *Mean Time*. London, England: Anvil Press.

Emerson, L. (2008). Materiality, intentionality, and the computer-generated poem: reading Walter Benn Michaels with Erin Mour's Pillage Laud. *ESC: English Studies in Canada*, *34*(4), 45–69.

Emerson, L. (2012). Review of *discourse.cpp* by OS le Si. *Computational Linguistics*, *38*(4), 923–925.

Erk, K. (2012). Vector space models of word meaning and phrase meaning: a survey. *Language and Linguistics Compass*, *6:10*, 635–653.

Erk, K., & Padó, S. (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.

Evert, S. (2004). *The statistics of word cooccurrences: word pairs and collocations*. Ph.D. thesis, University of Stuttgart.

Feng, Y., & Lapata, M. (2010). Visual information in semantic representation. In *Human Language Technologies: the 2010 annual conference of the North American Chapter of the Association for Computational Linguistics (NAACL2010)*, (pp. 91–99). Los Angeles, California.

Firth, J. R. (1957). *A synopsis of linguistic theory, 1930–1955*. Oxford: Philological Society.

Frege, G. (1892). über sinn und bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, *100*, 25–50.

Ginsberg, A. (1996). 5a.m. In *Death and Fame:Last Poems 1993-1997*. Harper Perennial (reprint 2000).

Grefenstette, G. (1994). *Explorations in automatic thesaurus discovery*. Springer.

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

Hejinian, L. (2000). *The language of inquiry*. Berkeley: University of California Press.

Hofstadter, D. R. (2001). Analogy as the core of cognition. In *The analogical mind: perspectives from cognitive science*, (pp. 499–538). Cambridge MA: The MIT Press/Bradford Books.

Hughes, T. (1999). Wuthering Heights. In *Birthday Letters*. Faber & Faber.

Labov, W. (1973). The boundaries of words and their meanings. In C.-J. Bailey, & R. W. Shuy (Eds.) *New ways of analysing variation in English*, (pp. 340–371). Georgetown University Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, (pp. 211–240).

le Si, O. (2011). *discourse.cpp*. Berlin: Press Press.

Leech, G., Garside, R., & Bryant, M. (1994). CLAWS4: The Tagging Of The British National Corpus. In *Proceedings of the 15th international Conference on Computational Linguistics (COLING 94)*, (pp. 622–628). Kyoto, Japan.

Masterman, M. (1971). Computerized haiku. In J. Reichardt (Ed.) *Cybernetics, art and ideas*, (pp. 175–184). London, England: Studio Vista.

Miles, J. (1940). More semantics of poetry. *The Kenyon Review*, *2*(4), 502–507.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*, (pp. 262–272).

Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science*, *34*(8), 1388–1429.

Molinaro, N., Carreiras, M., & Duñabeitia, J. A. (2012). Semantic combinatorial processing of non-anomalous expressions. *NeuroImage*, *59*(4), 3488–3501.

Montague, R. (1973). The proper treatment of quantification in ordinary english. In J. Hintikka, J. Moravcsik, & P. Suppes (Eds.) *Approaches to Natural Language*, (pp. 221–242). Dordrecht.

Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (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 (NAACL2010)*, (pp. 100–108). Los Angeles, CA.

Richards, I. A. (1970). *Poetries and Sciences: a reissue of Science and Poetry (1926, 1935) with commentary*. London: Routledge & Kegan Paul.

Schütze, H. (1998). Automatic word sense discrimination. *Computational Linguistics*, *24*(1), 97–123.

Searle, J. (1969). *Speech Acts*. Cambridge University Press.

Shutova, E. (2010). Models of metaphor in NLP. In *Proceedings of the 48th annual meeting of the Association for Computational Linguistics (ACL2010)*, (pp. 688–697).

Slater, A. (2008). Ithaca, Winter. *The Cortland Review*, *41*.
URL http://www.cortlandreview.com/issue/41/slater.html

Spärck Jones, K. (1964). *Synonymy and semantic classification*. Ph.D. thesis, University of Cambridge. Cambridge Language Research Unit.

Stein, G. (1924). If I Told Him: A Completed Portrait of Picasso. *Vanity Fair*.

Tarski, A. (1944). The semantic conception of truth. *Philosophy and Phenomenological Research*, *4*, 341–375.

Thater, S., Fürstenau, H., & Pinkal, M. (2010). Contextualizing semantic representations using syntactically enriched vector models. In *Proceedings of the 48th annual meeting of the Association for Computational Linguistics (ACL2010)*, (pp. 948–957). Uppsala, Sweden.

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188.

Veale, T. (2012). *Exploding the creativity myth: The computational foundations of linguistic creativity*. A&C Black.

Vecchi, E. M., Baroni, M., & Zamparelli, R. (2011). (Linear) maps of the impossible: Capturing semantic anomalies in distributional space. In *Proceedings of the ACL DISCo (Distributional Semantics and Compositionality) Workshop*. Portland, Oregon.

Wheelwright, P. (1940). On the semantics of poetry. *The Kenyon Review*, *2*(3), pp. 263–283.

Wittgenstein, L. (1953). *Philosophical investigations*. Wiley-Blackwell (2010 reprint).

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of dutch compounds. *Language and Cognitive Processes*, *9*(3), 341–368.