Type Theory and Distributional Models of Meaning

Shalom Lappin
King's College London

Workshop on Type Dependency, Type Theory with Records, and Natural-Language Flexibility

Queen Mary University of London
June 17, 2011

Outline

Outline

Outline

Outline

Types, Denotations, and Models

- Classical semantic theories constructed on the basis of Tarski's (1933) formal definition of truth specify a type theory and an associated model theory.
- They define a function that maps the syntactic category of an expression to its semantic type.
- The type of an expression determines the kind of denotation that it receives relative to a model.
- As Bach (1986) observes, the type theory of a language and the class of its possible models jointly specify an ontology for it.

Types, Denotations, and Models

- Classical semantic theories constructed on the basis of Tarski's (1933) formal definition of truth specify a type theory and an associated model theory.
- They define a function that maps the syntactic category of an expression to its semantic type.
- The type of an expression determines the kind of denotation that it receives relative to a model.
- As Bach (1986) observes, the type theory of a language and the class of its possible models jointly specify an ontology for it.

Types, Denotations, and Models

- Classical semantic theories constructed on the basis of Tarski's (1933) formal definition of truth specify a type theory and an associated model theory.
- They define a function that maps the syntactic category of an expression to its semantic type.
- The type of an expression determines the kind of denotation that it receives relative to a model.
- As Bach (1986) observes, the type theory of a language and the class of its possible models jointly specify an ontology for it.

Types, Denotations, and Models

- Classical semantic theories constructed on the basis of Tarski's (1933) formal definition of truth specify a type theory and an associated model theory.
- They define a function that maps the syntactic category of an expression to its semantic type.
- The type of an expression determines the kind of denotation that it receives relative to a model.
- As Bach (1986) observes, the type theory of a language and the class of its possible models jointly specify an ontology for it.

Meaning and Denotation in a Model

- A formal semantic theory recursively defines the denotation of an expression in terms of the denotations of its syntactic constituents.
- It computes the semantic values of a sentence as a function of the values of its syntactic constituents.
- Within such a theory the meaning of an expression is identified with a function from indices (the expressions themselves, worlds, situations, times, etc.), to denotations in a model.
- The meaning of a sentence is a function from indices to truth-values.

Meaning and Denotation in a Model

- A formal semantic theory recursively defines the denotation of an expression in terms of the denotations of its syntactic constituents.
- It computes the semantic values of a sentence as a function of the values of its syntactic constituents.
- Within such a theory the meaning of an expression is identified with a function from indices (the expressions themselves, worlds, situations, times, etc.), to denotations in a model.
- The meaning of a sentence is a function from indices to truth-values.

Meaning and Denotation in a Model

- A formal semantic theory recursively defines the denotation of an expression in terms of the denotations of its syntactic constituents.
- It computes the semantic values of a sentence as a function of the values of its syntactic constituents.
- Within such a theory the meaning of an expression is identified with a function from indices (the expressions themselves, worlds, situations, times, etc.), to denotations in a model.
- The meaning of a sentence is a function from indices to truth-values.

Meaning and Denotation in a Model

- A formal semantic theory recursively defines the denotation of an expression in terms of the denotations of its syntactic constituents.
- It computes the semantic values of a sentence as a function of the values of its syntactic constituents.
- Within such a theory the meaning of an expression is identified with a function from indices (the expressions themselves, worlds, situations, times, etc.), to denotations in a model.
- The meaning of a sentence is a function from indices to truth-values.

Expressing Lexical Meaning in Formal Semantic Systems

- Both classical and revised formal semantic theories focus on the combinatorial dimension of meaning.
- They use the type system and a recursive definition of semantic value to compute the interpretations of expressions from their syntactic components.
- The meaning of a lexical item is given through its type assignment and its denotation in a model.
- Semantic relations among lexical items that cannot be encoded in the type system or the interpretation function of a model are expressed through meaning postulates.

Expressing Lexical Meaning in Formal Semantic Systems

- Both classical and revised formal semantic theories focus on the combinatorial dimension of meaning.
- They use the type system and a recursive definition of semantic value to compute the interpretations of expressions from their syntactic components.
- The meaning of a lexical item is given through its type assignment and its denotation in a model.
- Semantic relations among lexical items that cannot be encoded in the type system or the interpretation function of a model are expressed through meaning postulates.

Expressing Lexical Meaning in Formal Semantic Systems

- Both classical and revised formal semantic theories focus on the combinatorial dimension of meaning.
- They use the type system and a recursive definition of semantic value to compute the interpretations of expressions from their syntactic components.
- The meaning of a lexical item is given through its type assignment and its denotation in a model.
- Semantic relations among lexical items that cannot be encoded in the type system or the interpretation function of a model are expressed through meaning postulates.

Expressing Lexical Meaning in Formal Semantic Systems

- Both classical and revised formal semantic theories focus on the combinatorial dimension of meaning.
- They use the type system and a recursive definition of semantic value to compute the interpretations of expressions from their syntactic components.
- The meaning of a lexical item is given through its type assignment and its denotation in a model.
- Semantic relations among lexical items that cannot be encoded in the type system or the interpretation function of a model are expressed through meaning postulates.

Meaning Postulates

- Meaning postulates can be used to characterize meaning implications between classes of lexical items within a given type.

- Montague uses them to identify extensional verbs, nouns, and modifiers, as with MP1 for extensional transitive verbs, cited in Dowty, Wall, and Peters (1981).

MP1. $\exists S \forall x \forall \mathcal{P} \Box [\delta(x, \mathcal{P}) \leftrightarrow \mathcal{P}\{^\wedge \lambda y[S\{x, y\}]\}]$,

where δ denotes a relation in intension for a transitive verb like *find*, S denotes its extensional counterpart, and $\mathcal{P}$ a generalized quantifier.

Meaning Postulates

- Meaning postulates can be used to characterize meaning implications between classes of lexical items within a given type.
- Montague uses them to identify extensional verbs, nouns, and modifiers, as with MP1 for extensional transitive verbs, cited in Dowty, Wall, and Peters (1981).

MP1. $\exists S \forall x \forall \mathcal{P} \Box [\delta(x, \mathcal{P}) \leftrightarrow \mathcal{P}\{^\wedge \lambda y[S\{x, y\}]\}]$,

where δ denotes a relation in intension for a transitive verb like *find*, S denotes its extensional counterpart, and $\mathcal{P}$ a generalized quantifier.

The Competence-Performance Distinction in Semantics

- Formal semantic theories model both lexical and phrasal meaning through categorical rules and algebraic systems that cannot accommodate gradience effects.
- This approach is common to theories which sustain compositionality and those with employ underspecified representations.
- It effectively invokes the same strong version of the competence-performance distinction that categorical models of syntax assume.
- This view of linguistic knowledge has dominated linguistic theory for the past fifty years.

The Competence-Performance Distinction in Semantics

- Formal semantic theories model both lexical and phrasal meaning through categorical rules and algebraic systems that cannot accommodate gradience effects.
- This approach is common to theories which sustain compositionality and those with employ underspecified representations.
- It effectively invokes the same strong version of the competence-performance distinction that categorical models of syntax assume.
- This view of linguistic knowledge has dominated linguistic theory for the past fifty years.

The Competence-Performance Distinction in Semantics

- Formal semantic theories model both lexical and phrasal meaning through categorical rules and algebraic systems that cannot accommodate gradience effects.
- This approach is common to theories which sustain compositionality and those with employ underspecified representations.
- It effectively invokes the same strong version of the competence-performance distinction that categorical models of syntax assume.
- This view of linguistic knowledge has dominated linguistic theory for the past fifty years.

The Competence-Performance Distinction in Semantics

- Formal semantic theories model both lexical and phrasal meaning through categorical rules and algebraic systems that cannot accommodate gradience effects.
- This approach is common to theories which sustain compositionality and those with employ underspecified representations.
- It effectively invokes the same strong version of the competence-performance distinction that categorical models of syntax assume.
- This view of linguistic knowledge has dominated linguistic theory for the past fifty years.

Explaining Gradience in Linguistic Representation

- Gradient effects in representation are ubiquitous throughout linguistic and other cognitive domains.
- Appeal to performance factors to explain gradience has no explanatory content unless it is supported by a precise account of how the interaction of competence and performance generates these effects in each case.
- By contrast, gradience is intrinsic to the formal models that information theoretic methods use to represent events and processes.

Explaining Gradience in Linguistic Representation

- Gradient effects in representation are ubiquitous throughout linguistic and other cognitive domains.
- Appeal to performance factors to explain gradience has no explanatory content unless it is supported by a precise account of how the interaction of competence and performance generates these effects in each case.
- By contrast, gradience is intrinsic to the formal models that information theoretic methods use to represent events and processes.

Explaining Gradience in Linguistic Representation

- Gradient effects in representation are ubiquitous throughout linguistic and other cognitive domains.
- Appeal to performance factors to explain gradience has no explanatory content unless it is supported by a precise account of how the interaction of competence and performance generates these effects in each case.
- By contrast, gradience is intrinsic to the formal models that information theoretic methods use to represent events and processes.

Three Views of Natural Language

- Bach (1986) identifies two theses on the character of natural language.
  (a) Chomsky's thesis: natural languages can be described as formal systems.
  (b) Montague's thesis: natural languages can be described as *interpreted* formal systems.
- Recent work in computational linguistics and cognitive modeling suggests a third proposal.
  (c) The Harris-Jelinek thesis: natural languages can be described as information theoretic systems, using stochastic models that express the distributional properties of its elements.

Three Views of Natural Language

- Bach (1986) identifies two theses on the character of natural language.
  (a) Chomsky's thesis: natural languages can be described as formal systems.
  (b) Montague's thesis: natural languages can be described as *interpreted* formal systems.
- Recent work in computational linguistics and cognitive modeling suggests a third proposal.
  (c) The Harris-Jelinek thesis: natural languages can be described as information theoretic systems, using stochastic models that express the distributional properties of its elements.

Three Views of Natural Language

- Bach (1986) identifies two theses on the character of natural language.
  - (a) Chomsky's thesis: natural languages can be described as formal systems.
  - (b) Montague's thesis: natural languages can be described as *interpreted* formal systems.
- Recent work in computational linguistics and cognitive modeling suggests a third proposal.
  - (c) The Harris-Jelinek thesis: natural languages can be described as information theoretic systems, using stochastic models that express the distributional properties of its elements.

Three Views of Natural Language

- Bach (1986) identifies two theses on the character of natural language.
  - (a) Chomsky's thesis: natural languages can be described as formal systems.
  - (b) Montague's thesis: natural languages can be described as *interpreted* formal systems.
- Recent work in computational linguistics and cognitive modeling suggests a third proposal.
  - (c) The Harris-Jelinek thesis: natural languages can be described as information theoretic systems, using stochastic models that express the distributional properties of its elements.

Three Views of Natural Language

- Bach (1986) identifies two theses on the character of natural language.
  (a) Chomsky's thesis: natural languages can be described as formal systems.
  (b) Montague's thesis: natural languages can be described as *interpreted* formal systems.
- Recent work in computational linguistics and cognitive modeling suggests a third proposal.
  (c) The Harris-Jelinek thesis: natural languages can be described as information theoretic systems, using stochastic models that express the distributional properties of its elements.

The Language Model Hypothesis

- The Language Model Hypothesis (LMH) for Syntax:

  Grammatical knowledge is represented as a stochastic language model.

- On the LMH, a speaker acquires a probability distribution $D : \Sigma^* \to [0, 1]$, over the strings $s \in \Sigma^*$, where Σ is a set of words (phonemes, morphemes, etc.) of the language, and, for any finite subset of Σ^*, $\sum p_D(s) = 1$.
- This distribution is generated by a probabilistic automaton or a probabilistic grammar.

The Language Model Hypothesis

- The Language Model Hypothesis (LMH) for Syntax:

  Grammatical knowledge is represented as a stochastic language model.

- On the LMH, a speaker acquires a probability distribution $D : \Sigma^* \rightarrow [0, 1]$, over the strings $s \in \Sigma^*$, where Σ is a set of words (phonemes, morphemes, etc.) of the language, and, for any finite subset of Σ^*, $\sum p_D(s) = 1$.
- This distribution is generated by a probabilistic automaton or a probabilistic grammar.

The Language Model Hypothesis

- The Language Model Hypothesis (LMH) for Syntax:

  Grammatical knowledge is represented as a stochastic language model.

- On the LMH, a speaker acquires a probability distribution $D : \Sigma^* \rightarrow [0, 1]$, over the strings $s \in \Sigma^*$, where Σ is a set of words (phonemes, morphemes, etc.) of the language, and, for any finite subset of Σ^*, $\sum p_D(s) = 1$.
- This distribution is generated by a probabilistic automaton or a probabilistic grammar.

Reformulating the Competence-Performance Distinction

- Representing linguistic knowledge stochastically does not eliminate the competence-performance distinction.
- It is still necessary to distinguish between a probabilistic grammar or automaton that generates a language model, and the parsing algorithm that implements it.
- However, a probabilistic characterization of linguistic knowledge does alter the nature of this distinction.
- The gradience of linguistic judgements and the defeasibility of grammatical constraints are now intrinsic to linguistic competence, rather than distorting factors contributed by performance mechanisms.

Reformulating the Competence-Performance Distinction

- Representing linguistic knowledge stochastically does not eliminate the competence-performance distinction.
- It is still necessary to distinguish between a probabilistic grammar or automaton that generates a language model, and the parsing algorithm that implements it.
- However, a probabilistic characterization of linguistic knowledge does alter the nature of this distinction.
- The gradience of linguistic judgements and the defeasibility of grammatical constraints are now intrinsic to linguistic competence, rather than distorting factors contributed by performance mechanisms.

Reformulating the Competence-Performance Distinction

- Representing linguistic knowledge stochastically does not eliminate the competence-performance distinction.
- It is still necessary to distinguish between a probabilistic grammar or automaton that generates a language model, and the parsing algorithm that implements it.
- However, a probabilistic characterization of linguistic knowledge does alter the nature of this distinction.
- The gradience of linguistic judgements and the defeasibility of grammatical constraints are now intrinsic to linguistic competence, rather than distorting factors contributed by performance mechanisms.

Reformulating the Competence-Performance Distinction

- Representing linguistic knowledge stochastically does not eliminate the competence-performance distinction.
- It is still necessary to distinguish between a probabilistic grammar or automaton that generates a language model, and the parsing algorithm that implements it.
- However, a probabilistic characterization of linguistic knowledge does alter the nature of this distinction.
- The gradience of linguistic judgements and the defeasibility of grammatical constraints are now intrinsic to linguistic competence, rather than distorting factors contributed by performance mechanisms.

Gradience in Semantic Properties and Relations

- Lexically mediated relations like synonymy, antinomy, polysemy, and hyponymy are notoriously prone to clustering and overlap effects.
- They hold for pairs of expressions over a continuum of degrees [0,1], rather than Boolean values {1,0}.
- Moreover, the denotations of major semantic types, like the predicates corresponding to NPs and VPs, can rarely, if ever, be identified as sets with determinate membership.
- The case for abandoning the categorical view of competence and adopting a probabilistic model is at least as strong in semantics as it is in syntax (as well as in other parts of the grammar).

Gradience in Semantic Properties and Relations

- Lexically mediated relations like synonymy, antinomy, polysemy, and hyponymy are notoriously prone to clustering and overlap effects.
- They hold for pairs of expressions over a continuum of degrees [0,1], rather than Boolean values {1,0}.
- Moreover, the denotations of major semantic types, like the predicates corresponding to NPs and VPs, can rarely, if ever, be identified as sets with determinate membership.
- The case for abandoning the categorical view of competence and adopting a probabilistic model is at least as strong in semantics as it is in syntax (as well as in other parts of the grammar).

Gradience in Semantic Properties and Relations

- Lexically mediated relations like synonymy, antinomy, polysemy, and hyponymy are notoriously prone to clustering and overlap effects.
- They hold for pairs of expressions over a continuum of degrees [0,1], rather than Boolean values {1,0}.
- Moreover, the denotations of major semantic types, like the predicates corresponding to NPs and VPs, can rarely, if ever, be identified as sets with determinate membership.
- The case for abandoning the categorical view of competence and adopting a probabilistic model is at least as strong in semantics as it is in syntax (as well as in other parts of the grammar).

Gradience in Semantic Properties and Relations

- Lexically mediated relations like synonymy, antinomy, polysemy, and hyponymy are notoriously prone to clustering and overlap effects.
- They hold for pairs of expressions over a continuum of degrees [0,1], rather than Boolean values {1,0}.
- Moreover, the denotations of major semantic types, like the predicates corresponding to NPs and VPs, can rarely, if ever, be identified as sets with determinate membership.
- The case for abandoning the categorical view of competence and adopting a probabilistic model is at least as strong in semantics as it is in syntax (as well as in other parts of the grammar).

Vector Space Models

- Vector Space Models (VSMs) (Turney and Pantel (2010)) offer a fine-grained distributional method for identifying a range of semantic relations among words and phrases.
- They are constructed from matrices in which words are listed vertically on the left, and the environments in which they appear are given horizontally along the top.
- These environments specify the dimensions of the model, corresponding to words, phrases, documents, units of discourse, or any other objects for tracking the occurrence of words.
- They can also include data structures encoding extra-linguistic elements, like visual scenes and events.

Vector Space Models

- Vector Space Models (VSMs) (Turney and Pantel (2010)) offer a fine-grained distributional method for identifying a range of semantic relations among words and phrases.
- They are constructed from matrices in which words are listed vertically on the left, and the environments in which they appear are given horizontally along the top.
- These environments specify the dimensions of the model, corresponding to words, phrases, documents, units of discourse, or any other objects for tracking the occurrence of words.
- They can also include data structures encoding extra-linguistic elements, like visual scenes and events.

Vector Space Models

- Vector Space Models (VSMs) (Turney and Pantel (2010)) offer a fine-grained distributional method for identifying a range of semantic relations among words and phrases.
- They are constructed from matrices in which words are listed vertically on the left, and the environments in which they appear are given horizontally along the top.
- These environments specify the dimensions of the model, corresponding to words, phrases, documents, units of discourse, or any other objects for tracking the occurrence of words.
- They can also include data structures encoding extra-linguistic elements, like visual scenes and events.

Vector Space Models

- Vector Space Models (VSMs) (Turney and Pantel (2010)) offer a fine-grained distributional method for identifying a range of semantic relations among words and phrases.
- They are constructed from matrices in which words are listed vertically on the left, and the environments in which they appear are given horizontally along the top.
- These environments specify the dimensions of the model, corresponding to words, phrases, documents, units of discourse, or any other objects for tracking the occurrence of words.
- They can also include data structures encoding extra-linguistic elements, like visual scenes and events.

A Word-Context Matrix

	context 1	context 2	context 3	context 4
financial	0	6	4	8
market	1	0	15	9
share	5	0	0	4
economic	0	1	26	12
chip	7	8	0	0
distributed	11	15	0	0
sequential	10	31	0	1
algorithm	14	22	2	1

Matrices and Vectors

- The integers in the cells of the matrix give the frequency of the word in an environment.
- A vector for a word is the row of values across the dimension columns of the matrix.
- The vectors for *chip* and *algorithm* are [7 8 0 0] and [14 22 2 1], respectively.

Matrices and Vectors

- The integers in the cells of the matrix give the frequency of the word in an environment.
- A vector for a word is the row of values across the dimension columns of the matrix.
- The vectors for *chip* and *algorithm* are [7 8 0 0] and [14 22 2 1], respectively.

Matrices and Vectors

- The integers in the cells of the matrix give the frequency of the word in an environment.
- A vector for a word is the row of values across the dimension columns of the matrix.
- The vectors for *chip* and *algorithm* are [7 8 0 0] and [14 22 2 1], respectively.

Measuring Semantic Distance

- A pair of vectors from a matrix can be projected as lines from a common point on a plane.
- The smaller the angle between the lines, the greater the similarity of the terms, as measured by their co-occurrence across the dimensions of the matrix.
- Computing the *cosine* of this angle is a convenient way of measuring the angles between vector pairs.
- If $\vec{x} = \langle x_1, x_2, ..., x_n \rangle$ and $\vec{y} = \langle y_1, y_2, ..., y_n \rangle$ are two vectors, then

$$cos(\vec{x}, \vec{y}) = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \cdot \sum_{i=1}^{n} y_i^2}}$$

Measuring Semantic Distance

- A pair of vectors from a matrix can be projected as lines from a common point on a plane.
- The smaller the angle between the lines, the greater the similarity of the terms, as measured by their co-occurrence across the dimensions of the matrix.
- Computing the *cosine* of this angle is a convenient way of measuring the angles between vector pairs.
- If $\vec{x} = \langle x_1, x_2, ..., x_n \rangle$ and $\vec{y} = \langle y_1, y_2, ..., y_n \rangle$ are two vectors, then

$$cos(\vec{x}, \vec{y}) = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \cdot \sum_{i=1}^{n} y_i^2}}$$

Measuring Semantic Distance

- A pair of vectors from a matrix can be projected as lines from a common point on a plane.
- The smaller the angle between the lines, the greater the similarity of the terms, as measured by their co-occurrence across the dimensions of the matrix.
- Computing the *cosine* of this angle is a convenient way of measuring the angles between vector pairs.
- If $\vec{x} = \langle x_1, x_2, ..., x_n \rangle$ and $\vec{y} = \langle y_1, y_2, ..., y_n \rangle$ are two vectors, then

$$cos(\vec{x}, \vec{y}) = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \cdot \sum_{i=1}^{n} y_i^2}}$$

Measuring Semantic Distance

- A pair of vectors from a matrix can be projected as lines from a common point on a plane.
- The smaller the angle between the lines, the greater the similarity of the terms, as measured by their co-occurrence across the dimensions of the matrix.
- Computing the *cosine* of this angle is a convenient way of measuring the angles between vector pairs.
- If $\vec{x} = \langle x_1, x_2, ..., x_n \rangle$ and $\vec{y} = \langle y_1, y_2, ..., y_n \rangle$ are two vectors, then

$$cos(\vec{x}, \vec{y}) = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \cdot \sum_{i=1}^{n} y_i^2}}$$

Measuring Semantic Distance

- The cosine of $\vec{x}$ and $\vec{y}$ is their internal product, formed by summing the products of the corresponding elements of the two vectors and normalizing the result relative to the lengths of the vectors.
- In computing $cos(\vec{x}, \vec{y})$ it may be desirable to apply a smoothing function to the raw frequency counts in each vector to compensate for sparse data, or to filter out the effects of high frequency terms.
- A higher value for $cos(\vec{x}, \vec{y})$ correlates with greater semantic relatedness of the terms associated with the $\vec{x}$ and $\vec{y}$ vectors.

Measuring Semantic Distance

- The cosine of $\vec{x}$ and $\vec{y}$ is their internal product, formed by summing the products of the corresponding elements of the two vectors and normalizing the result relative to the lengths of the vectors.
- In computing $cos(\vec{x}, \vec{y})$ it may be desirable to apply a smoothing function to the raw frequency counts in each vector to compensate for sparse data, or to filter out the effects of high frequency terms.
- A higher value for $cos(\vec{x}, \vec{y})$ correlates with greater semantic relatedness of the terms associated with the $\vec{x}$ and $\vec{y}$ vectors.

Measuring Semantic Distance

- The cosine of $\vec{x}$ and $\vec{y}$ is their internal product, formed by summing the products of the corresponding elements of the two vectors and normalizing the result relative to the lengths of the vectors.
- In computing $cos(\vec{x}, \vec{y})$ it may be desirable to apply a smoothing function to the raw frequency counts in each vector to compensate for sparse data, or to filter out the effects of high frequency terms.
- A higher value for $cos(\vec{x}, \vec{y})$ correlates with greater semantic relatedness of the terms associated with the $\vec{x}$ and $\vec{y}$ vectors.

VSMs as Representations of Lexical Meaning and Learning

- VSMs provide highly successful methods for identifying a variety of lexical semantic relations, including synonymy, antinomy, polysemy, and hypernym classes.
- They also perform very well in unsupervised sense disambiguation tasks.
- VSMs offer a distributional view of lexical semantic learning.
- On this approach speakers acquire lexical meaning by estimating the environments (linguistic and non-linguistic) in which the words of their language appear.

VSMs as Representations of Lexical Meaning and Learning

- VSMs provide highly successful methods for identifying a variety of lexical semantic relations, including synonymy, antinomy, polysemy, and hypernym classes.
- They also perform very well in unsupervised sense disambiguation tasks.
- VSMs offer a distributional view of lexical semantic learning.
- On this approach speakers acquire lexical meaning by estimating the environments (linguistic and non-linguistic) in which the words of their language appear.

VSMs as Representations of Lexical Meaning and Learning

- VSMs provide highly successful methods for identifying a variety of lexical semantic relations, including synonymy, antinomy, polysemy, and hypernym classes.
- They also perform very well in unsupervised sense disambiguation tasks.
- VSMs offer a distributional view of lexical semantic learning.
- On this approach speakers acquire lexical meaning by estimating the environments (linguistic and non-linguistic) in which the words of their language appear.

VSMs as Representations of Lexical Meaning and Learning

- VSMs provide highly successful methods for identifying a variety of lexical semantic relations, including synonymy, antinomy, polysemy, and hypernym classes.
- They also perform very well in unsupervised sense disambiguation tasks.
- VSMs offer a distributional view of lexical semantic learning.
- On this approach speakers acquire lexical meaning by estimating the environments (linguistic and non-linguistic) in which the words of their language appear.

Compositional VSMs

- The primary limitation of VSMs is that they measure semantic distances and relations among words independently of syntactic structure (bag of words).
- Coecke et al. (2010) and Grefenstette et al. (2011) propose a procedure for computing vector values for sentences on the basis of the vectors of their syntactic constituents.
- This procedure relies upon a category theoretic representation of the types of a pregroup grammar (PGG, Lambek (2007,2008)), which builds up complex syntactic categories through direction-marked function application in a manner similar to a basic categorial grammar.
- All sentences receive vectors in the same vector space, and so they can be compared for semantic similarity using measures like cosine.

Compositional VSMs

- The primary limitation of VSMs is that they measure semantic distances and relations among words independently of syntactic structure (bag of words).
- Coecke et al. (2010) and Grefenstette et al. (2011) propose a procedure for computing vector values for sentences on the basis of the vectors of their syntactic constituents.
- This procedure relies upon a category theoretic representation of the types of a pregroup grammar (PGG, Lambek (2007,2008)), which builds up complex syntactic categories through direction-marked function application in a manner similar to a basic categorial grammar.
- All sentences receive vectors in the same vector space, and so they can be compared for semantic similarity using measures like cosine.

Compositional VSMs

- The primary limitation of VSMs is that they measure semantic distances and relations among words independently of syntactic structure (bag of words).
- Coecke et al. (2010) and Grefenstette et al. (2011) propose a procedure for computing vector values for sentences on the basis of the vectors of their syntactic constituents.
- This procedure relies upon a category theoretic representation of the types of a pregroup grammar (PGG, Lambek (2007,2008)), which builds up complex syntactic categories through direction-marked function application in a manner similar to a basic categorial grammar.
- All sentences receive vectors in the same vector space, and so they can be compared for semantic similarity using measures like cosine.

Compositional VSMs

- The primary limitation of VSMs is that they measure semantic distances and relations among words independently of syntactic structure (bag of words).
- Coecke et al. (2010) and Grefenstette et al. (2011) propose a procedure for computing vector values for sentences on the basis of the vectors of their syntactic constituents.
- This procedure relies upon a category theoretic representation of the types of a pregroup grammar (PGG, Lambek (2007,2008)), which builds up complex syntactic categories through direction-marked function application in a manner similar to a basic categorial grammar.
- All sentences receive vectors in the same vector space, and so they can be compared for semantic similarity using measures like cosine.

Computing the Vector of a Sentence

- PGGs are modeled as *compact closed categories*.
- A sentence vector is computed by a linear map f on the tensor product for the vectors of its main constituents, where f stores the type categorial structure of the string determined by its PGG representation.
- The vector for a sentence headed by a transitive verb, for example, is computed according to the equation

$$\overrightarrow{subj\ V_{tr}\ obj} = f(\overrightarrow{subj} \otimes \overrightarrow{V_{tr}} \otimes \overrightarrow{obj})$$

Computing the Vector of a Sentence

- PGGs are modeled as *compact closed categories*.
- A sentence vector is computed by a linear map f on the tensor product for the vectors of its main constituents, where f stores the type categorial structure of the string determined by its PGG representation.
- The vector for a sentence headed by a transitive verb, for example, is computed according to the equation

$$\overrightarrow{subj\ V_{tr}\ obj} = f(\overrightarrow{subj} \otimes \overrightarrow{V_{tr}} \otimes \overrightarrow{obj})$$

Computing the Vector of a Sentence

- PGGs are modeled as *compact closed categories*.
- A sentence vector is computed by a linear map f on the tensor product for the vectors of its main constituents, where f stores the type categorial structure of the string determined by its PGG representation.
- The vector for a sentence headed by a transitive verb, for example, is computed according to the equation

$$\overrightarrow{subj\ V_{tr}\ obj} = f(\overrightarrow{subj} \otimes \overrightarrow{V_{tr}} \otimes \overrightarrow{obj})$$

Computing the Vector of a Sentence

- The vector of a transitive verb V_{tr} could be taken to be an element of the tensor product of the vector spaces for the two noun bases corresponding to its possible subject and object arguments $\overrightarrow{V_{tr}} \in N \otimes N$.
- The vector for a sentence headed by a transitive verb could be computed as the point-wise product of the verb's vector, and the tensor product of its subject and its object.

$$\overrightarrow{subj\ V_{tr}\ obj} = \overrightarrow{V_{tr}} \odot (\overrightarrow{subj} \otimes \overrightarrow{obj})$$

Computing the Vector of a Sentence

- The vector of a transitive verb V_{tr} could be taken to be an element of the tensor product of the vector spaces for the two noun bases corresponding to its possible subject and object arguments $\overrightarrow{V_{tr}} \in N \otimes N$.
- The vector for a sentence headed by a transitive verb could be computed as the point-wise product of the verb's vector, and the tensor product of its subject and its object.

$$\overrightarrow{subj\ V_{tr}\ obj} = \overrightarrow{V_{tr}} \odot (\overrightarrow{subj} \otimes \overrightarrow{obj})$$

Advantages of PGG Compositional VSMs

- PGG compositional VSMs (CVSMs) offer a formally grounded and computationally efficient method for obtaining vectors for complex expressions from their syntactic constituents.
- They permit the same kind of measurement for relations of semantic similarity among sentences that lexical VSMs give for word pairs.
- They can be trained on a (PGG parsed) corpus, and their performance evaluated against human annotators' semantic judgements for phrases and sentences.

Advantages of PGG Compositional VSMs

- PGG compositional VSMs (CVSMs) offer a formally grounded and computationally efficient method for obtaining vectors for complex expressions from their syntactic constituents.
- They permit the same kind of measurement for relations of semantic similarity among sentences that lexical VSMs give for word pairs.
- They can be trained on a (PGG parsed) corpus, and their performance evaluated against human annotators' semantic judgements for phrases and sentences.

Advantages of PGG Compositional VSMs

- PGG compositional VSMs (CVSMs) offer a formally grounded and computationally efficient method for obtaining vectors for complex expressions from their syntactic constituents.
- They permit the same kind of measurement for relations of semantic similarity among sentences that lexical VSMs give for word pairs.
- They can be trained on a (PGG parsed) corpus, and their performance evaluated against human annotators' semantic judgements for phrases and sentences.

Problems with CVSMs

- Although the vector of a complex expression is the value of a linear map on the vectors of its parts, it is not obvious what independent property this vector represents.
- Sentential vectors do not correspond to the distributional properties of these sentences, as the data in most corpora is too sparse to estimate distributional vectors for all but a few sentences, across most dimensions.
- Coecke et al. 2010 show that it is possible to encode a classical model theoretic semantics in their system by using vectors to express sets, relations, and truth-values.
- But CVSMs are interesting to the extent that the sentential vectors that they assign are derived from lexical vectors that represent the distributional properties of these expressions.

Problems with CVSMs

- Although the vector of a complex expression is the value of a linear map on the vectors of its parts, it is not obvious what independent property this vector represents.
- Sentential vectors do not correspond to the distributional properties of these sentences, as the data in most corpora is too sparse to estimate distributional vectors for all but a few sentences, across most dimensions.
- Coecke et al. 2010 show that it is possible to encode a classical model theoretic semantics in their system by using vectors to express sets, relations, and truth-values.
- But CVSMs are interesting to the extent that the sentential vectors that they assign are derived from lexical vectors that represent the distributional properties of these expressions.

Problems with CVSMs

- Although the vector of a complex expression is the value of a linear map on the vectors of its parts, it is not obvious what independent property this vector represents.
- Sentential vectors do not correspond to the distributional properties of these sentences, as the data in most corpora is too sparse to estimate distributional vectors for all but a few sentences, across most dimensions.
- Coecke et al. 2010 show that it is possible to encode a classical model theoretic semantics in their system by using vectors to express sets, relations, and truth-values.
- But CVSMs are interesting to the extent that the sentential vectors that they assign are derived from lexical vectors that represent the distributional properties of these expressions.

Problems with CVSMs

- Although the vector of a complex expression is the value of a linear map on the vectors of its parts, it is not obvious what independent property this vector represents.
- Sentential vectors do not correspond to the distributional properties of these sentences, as the data in most corpora is too sparse to estimate distributional vectors for all but a few sentences, across most dimensions.
- Coecke et al. 2010 show that it is possible to encode a classical model theoretic semantics in their system by using vectors to express sets, relations, and truth-values.
- But CVSMs are interesting to the extent that the sentential vectors that they assign are derived from lexical vectors that represent the distributional properties of these expressions.

Classical Formal Semantic Theories vs CVSMs

- In classical formal semantic theories the functions that drive semantic composition are supplied by the type theory, where the type of each expression specifies the formal character of its denotation in a model.
- The sequence of functions that determines the semantic value of a sentence exhibits at each point a value that directly corresponds to an independently motivated semantic property of the expression to which it is assigned.
- Types of denotation provide non-arbitrary formal relations between types of expressions and classes of entities specified relative to a model.
- The sentential vectors obtained from distributional vectors of lexical items lack this sort of independent status.

Classical Formal Semantic Theories vs CVSMs

- In classical formal semantic theories the functions that drive semantic composition are supplied by the type theory, where the type of each expression specifies the formal character of its denotation in a model.
- The sequence of functions that determines the semantic value of a sentence exhibits at each point a value that directly corresponds to an independently motivated semantic property of the expression to which it is assigned.
- Types of denotation provide non-arbitrary formal relations between types of expressions and classes of entities specified relative to a model.
- The sentential vectors obtained from distributional vectors of lexical items lack this sort of independent status.

Classical Formal Semantic Theories vs CVSMs

- In classical formal semantic theories the functions that drive semantic composition are supplied by the type theory, where the type of each expression specifies the formal character of its denotation in a model.
- The sequence of functions that determines the semantic value of a sentence exhibits at each point a value that directly corresponds to an independently motivated semantic property of the expression to which it is assigned.
- Types of denotation provide non-arbitrary formal relations between types of expressions and classes of entities specified relative to a model.
- The sentential vectors obtained from distributional vectors of lexical items lack this sort of independent status.

Classical Formal Semantic Theories vs CVSMs

- In classical formal semantic theories the functions that drive semantic composition are supplied by the type theory, where the type of each expression specifies the formal character of its denotation in a model.
- The sequence of functions that determines the semantic value of a sentence exhibits at each point a value that directly corresponds to an independently motivated semantic property of the expression to which it is assigned.
- Types of denotation provide non-arbitrary formal relations between types of expressions and classes of entities specified relative to a model.
- The sentential vectors obtained from distributional vectors of lexical items lack this sort of independent status.

Truth/Probability Conditions vs Sentential Vectors

- An important part of the interpretation of a sentence involves knowing its truth (more generally, its satisfaction or fulfillment) conditions.
- From a probabilistic perspective, we can exchange truth conditions for probability (or plausibility) conditions, the likelihood of a sentence occurring given certain conditions.
- It is not obvious how we can extract such conditions, expressed in Boolean or probabilistic terms, from sentential vector values, when these are computed from vectors expressing the distributional (rather then the model theoretic or conditional probability) properties of their constituent lexical items.

Truth/Probability Conditions vs Sentential Vectors

- An important part of the interpretation of a sentence involves knowing its truth (more generally, its satisfaction or fulfillment) conditions.
- From a probabilistic perspective, we can exchange truth conditions for probability (or plausibility) conditions, the likelihood of a sentence occurring given certain conditions.
- It is not obvious how we can extract such conditions, expressed in Boolean or probabilistic terms, from sentential vector values, when these are computed from vectors expressing the distributional (rather then the model theoretic or conditional probability) properties of their constituent lexical items.

Truth/Probability Conditions vs Sentential Vectors

- An important part of the interpretation of a sentence involves knowing its truth (more generally, its satisfaction or fulfillment) conditions.
- From a probabilistic perspective, we can exchange truth conditions for probability (or plausibility) conditions, the likelihood of a sentence occurring given certain conditions.
- It is not obvious how we can extract such conditions, expressed in Boolean or probabilistic terms, from sentential vector values, when these are computed from vectors expressing the distributional (rather then the model theoretic or conditional probability) properties of their constituent lexical items.

A Semantically Enriched Language Model

- Another way to integrate lexical semantics into combinatorial meaning is to enrich the conditional dependencies of lexicalized probabilistic grammars (such as LPCFG, PCCG) with semantic features specified in terms of the distributional (VSM) properties of lexical heads of constituents.
- The rules of the grammar are specified as probabilities of constituents conditioned by the semantic (and syntactic) features of their lexical heads, and of the lexical heads of their daughters.
- The semantic properties of lexical elements play a direct role in determining a sentence's conditional probability, expressed as a probability determined by the probabilities of its constituents (it is the product of the rules applied in the derivation of the sentence).

A Semantically Enriched Language Model

- Another way to integrate lexical semantics into combinatorial meaning is to enrich the conditional dependencies of lexicalized probabilistic grammars (such as LPCFG, PCCG) with semantic features specified in terms of the distributional (VSM) properties of lexical heads of constituents.
- The rules of the grammar are specified as probabilities of constituents conditioned by the semantic (and syntactic) features of their lexical heads, and of the lexical heads of their daughters.
- The semantic properties of lexical elements play a direct role in determining a sentence's conditional probability, expressed as a probability determined by the probabilities of its constituents (it is the product of the rules applied in the derivation of the sentence).

A Semantically Enriched Language Model

- Another way to integrate lexical semantics into combinatorial meaning is to enrich the conditional dependencies of lexicalized probabilistic grammars (such as LPCFG, PCCG) with semantic features specified in terms of the distributional (VSM) properties of lexical heads of constituents.
- The rules of the grammar are specified as probabilities of constituents conditioned by the semantic (and syntactic) features of their lexical heads, and of the lexical heads of their daughters.
- The semantic properties of lexical elements play a direct role in determining a sentence's conditional probability, expressed as a probability determined by the probabilities of its constituents (it is the product of the rules applied in the derivation of the sentence).

Modeling Plausibility and Entailment

- In a semantically enriched language model the probability value of a sentence can (in part) be correlated with its plausibility in non-linguistic contexts.
- Entailment can be reconstructed algebraically, on the model of entailment in a lattice of propositions, as a partial order on conditional probabilities.
- A sentence A probabilistically entails a sentence B, relative to a distribution D, when, for $c \in R$ (R a set of relevant conditions), $p_D(A|c) \leq p_D(B|c)$.
- Clearly the nature of the entailment relation will depend on the specification of R.

Modeling Plausibility and Entailment

- In a semantically enriched language model the probability value of a sentence can (in part) be correlated with its plausibility in non-linguistic contexts.
- Entailment can be reconstructed algebraically, on the model of entailment in a lattice of propositions, as a partial order on conditional probabilities.
- A sentence A probabilistically entails a sentence B, relative to a distribution D, when, for $c \in R$ (R a set of relevant conditions), $p_D(A|c) \leq p_D(B|c)$.
- Clearly the nature of the entailment relation will depend on the specification of R.

Modeling Plausibility and Entailment

- In a semantically enriched language model the probability value of a sentence can (in part) be correlated with its plausibility in non-linguistic contexts.
- Entailment can be reconstructed algebraically, on the model of entailment in a lattice of propositions, as a partial order on conditional probabilities.
- A sentence A probabilistically entails a sentence B, relative to a distribution D, when, for $c \in R$ (R a set of relevant conditions), $p_D(A|c) \leq p_D(B|c)$.
- Clearly the nature of the entailment relation will depend on the specification of R.

Modeling Plausibility and Entailment

- In a semantically enriched language model the probability value of a sentence can (in part) be correlated with its plausibility in non-linguistic contexts.
- Entailment can be reconstructed algebraically, on the model of entailment in a lattice of propositions, as a partial order on conditional probabilities.
- A sentence A probabilistically entails a sentence B, relative to a distribution D, when, for $c \in R$ (R a set of relevant conditions), $p_D(A|c) \leq p_D(B|c)$.
- Clearly the nature of the entailment relation will depend on the specification of R.

The Representation of Linguistic Knowledge as an Integrated Language Model

- An enriched lexicalized probabilistic grammar of this kind will specify an integrated language model that generates a probability distribution for the phrases and sentences of a language that is partially determined by their lexically based semantic properties.
- The resulting language model provides a fully integrated representation of semantic and syntactic (as well as other kinds of) linguistic knowledge.
- One might object that in this framework it is not possible to distinguish precisely the semantic, syntactic, and real world conditions that determine the probability of a sentence.

The Representation of Linguistic Knowledge as an Integrated Language Model

- An enriched lexicalized probabilistic grammar of this kind will specify an integrated language model that generates a probability distribution for the phrases and sentences of a language that is partially determined by their lexically based semantic properties.
- The resulting language model provides a fully integrated representation of semantic and syntactic (as well as other kinds of) linguistic knowledge.
- One might object that in this framework it is not possible to distinguish precisely the semantic, syntactic, and real world conditions that determine the probability of a sentence.

The Representation of Linguistic Knowledge as an Integrated Language Model

- An enriched lexicalized probabilistic grammar of this kind will specify an integrated language model that generates a probability distribution for the phrases and sentences of a language that is partially determined by their lexically based semantic properties.
- The resulting language model provides a fully integrated representation of semantic and syntactic (as well as other kinds of) linguistic knowledge.
- One might object that in this framework it is not possible to distinguish precisely the semantic, syntactic, and real world conditions that determine the probability of a sentence.

The Representation of Linguistic Knowledge as an Integrated Language Model

- This is correct, but it is also true of lexical VSMs.
- The distribution of lexical items depends upon all of these factors, and we can separate them into distinct classes of features only locally to particular contexts.
- The interpenetration of these conditions in the language model is a pervasive aspect of the distributional view of meaning and structure.
- We can focus on the role of certain factors in controlling the probabilities of strings, but there is ultimately no well grounded partitioning of these factors into disjoint classes of syntactic, semantic, and non-linguistic conditions.

The Representation of Linguistic Knowledge as an Integrated Language Model

- This is correct, but it is also true of lexical VSMs.
- The distribution of lexical items depends upon all of these factors, and we can separate them into distinct classes of features only locally to particular contexts.
- The interpenetration of these conditions in the language model is a pervasive aspect of the distributional view of meaning and structure.
- We can focus on the role of certain factors in controlling the probabilities of strings, but there is ultimately no well grounded partitioning of these factors into disjoint classes of syntactic, semantic, and non-linguistic conditions.

The Representation of Linguistic Knowledge as an Integrated Language Model

- This is correct, but it is also true of lexical VSMs.
- The distribution of lexical items depends upon all of these factors, and we can separate them into distinct classes of features only locally to particular contexts.
- The interpenetration of these conditions in the language model is a pervasive aspect of the distributional view of meaning and structure.
- We can focus on the role of certain factors in controlling the probabilities of strings, but there is ultimately no well grounded partitioning of these factors into disjoint classes of syntactic, semantic, and non-linguistic conditions.

The Representation of Linguistic Knowledge as an Integrated Language Model

- This is correct, but it is also true of lexical VSMs.
- The distribution of lexical items depends upon all of these factors, and we can separate them into distinct classes of features only locally to particular contexts.
- The interpenetration of these conditions in the language model is a pervasive aspect of the distributional view of meaning and structure.
- We can focus on the role of certain factors in controlling the probabilities of strings, but there is ultimately no well grounded partitioning of these factors into disjoint classes of syntactic, semantic, and non-linguistic conditions.

Grammar as Type Theory

- The grammar constitutes the combinatorial mechanism for computing the semantically conditioned probabilities of complex constituents from the lexically dependent probabilities of their constituents.
- No additional type theory is required as a device for producing semantic values for complex expressions.
- Syntactic categories are semantic types, where these are units of semantic value, expressed as conditional probabilities.

Grammar as Type Theory

- The grammar constitutes the combinatorial mechanism for computing the semantically conditioned probabilities of complex constituents from the lexically dependent probabilities of their constituents.
- No additional type theory is required as a device for producing semantic values for complex expressions.
- Syntactic categories are semantic types, where these are units of semantic value, expressed as conditional probabilities.

Grammar as Type Theory

- The grammar constitutes the combinatorial mechanism for computing the semantically conditioned probabilities of complex constituents from the lexically dependent probabilities of their constituents.
- No additional type theory is required as a device for producing semantic values for complex expressions.
- Syntactic categories are semantic types, where these are units of semantic value, expressed as conditional probabilities.

Semantic Learning in an Integrated Language Model

- In an integrated language model semantic learning is assimilated to distributionally driven learning of syntax, enriched with semantic features.
- Probabilistic theories of grammar induction (Clark and Lappin 2011) can be extended to the representations that generate the probability distributions of an integrated language model.
- Therefore, we avoid the problematic choice of encoding the mechanisms of a formal semantic theory as strong learning biases on language learning, or positing them as general constraints on cognitive representation and processing.

Semantic Learning in an Integrated Language Model

- In an integrated language model semantic learning is assimilated to distributionally driven learning of syntax, enriched with semantic features.
- Probabilistic theories of grammar induction (Clark and Lappin 2011) can be extended to the representations that generate the probability distributions of an integrated language model.
- Therefore, we avoid the problematic choice of encoding the mechanisms of a formal semantic theory as strong learning biases on language learning, or positing them as general constraints on cognitive representation and processing.

Semantic Learning in an Integrated Language Model

- In an integrated language model semantic learning is assimilated to distributionally driven learning of syntax, enriched with semantic features.
- Probabilistic theories of grammar induction (Clark and Lappin 2011) can be extended to the representations that generate the probability distributions of an integrated language model.
- Therefore, we avoid the problematic choice of encoding the mechanisms of a formal semantic theory as strong learning biases on language learning, or positing them as general constraints on cognitive representation and processing.

Conclusions

- Classical formal semantic theories compute the semantic value of an expression from its type, and the denotations of its constituents in a model.
- They sustain a strong form of the competence-performance distinction that excludes gradience effects for semantic properties.
- These are attributed to performance factors, but no rigorous account is offered of how the interaction of competence and performance systems produces gradience.

Conclusions

- Classical formal semantic theories compute the semantic value of an expression from its type, and the denotations of its constituents in a model.
- They sustain a strong form of the competence-performance distinction that excludes gradience effects for semantic properties.
- These are attributed to performance factors, but no rigorous account is offered of how the interaction of competence and performance systems produces gradience.

Conclusions

- Classical formal semantic theories compute the semantic value of an expression from its type, and the denotations of its constituents in a model.
- They sustain a strong form of the competence-performance distinction that excludes gradience effects for semantic properties.
- These are attributed to performance factors, but no rigorous account is offered of how the interaction of competence and performance systems produces gradience.

Conclusions

- Gradience effects are particularly clear and pervasive in lexical semantic relations.
- Classical and revised theories lack formally elegant procedures for integrating lexical meaning into the combinatorial interpretation of complex expressions determined by the type system.
- They are unable to deal with gradience effects except by invoking performance factors.

Conclusions

- Gradience effects are particularly clear and pervasive in lexical semantic relations.
- Classical and revised theories lack formally elegant procedures for integrating lexical meaning into the combinatorial interpretation of complex expressions determined by the type system.
- They are unable to deal with gradience effects except by invoking performance factors.

Conclusions

- Gradience effects are particularly clear and pervasive in lexical semantic relations.
- Classical and revised theories lack formally elegant procedures for integrating lexical meaning into the combinatorial interpretation of complex expressions determined by the type system.
- They are unable to deal with gradience effects except by invoking performance factors.

Conclusions

- VSMs provide a powerful representation of lexical semantic relations in which gradience is an intrinsic feature of lexical meaning.
- PGG CVSMs permit a formally grounded and efficient way of computing the vectors of sentences from those of their constituents, but the sentential vectors that they generate from lexical vectors that are distributional do not have a straightforward interpretation.
- Language models enriched with probabilities conditioned by distributional lexical semantic features may offer an alternative framework for integrating lexical and combinatorial dimensions of meaning within a stochastic representation of linguistic knowledge.
- They allow semantic learning to be accommodated within probabilistic models of grammar induction.

Conclusions

- VSMs provide a powerful representation of lexical semantic relations in which gradience is an intrinsic feature of lexical meaning.
- PGG CVSMs permit a formally grounded and efficient way of computing the vectors of sentences from those of their constituents, but the sentential vectors that they generate from lexical vectors that are distributional do not have a straightforward interpretation.
- Language models enriched with probabilities conditioned by distributional lexical semantic features may offer an alternative framework for integrating lexical and combinatorial dimensions of meaning within a stochastic representation of linguistic knowledge.
- They allow semantic learning to be accommodated within probabilistic models of grammar induction.

Conclusions

- VSMs provide a powerful representation of lexical semantic relations in which gradience is an intrinsic feature of lexical meaning.
- PGG CVSMs permit a formally grounded and efficient way of computing the vectors of sentences from those of their constituents, but the sentential vectors that they generate from lexical vectors that are distributional do not have a straightforward interpretation.
- Language models enriched with probabilities conditioned by distributional lexical semantic features may offer an alternative framework for integrating lexical and combinatorial dimensions of meaning within a stochastic representation of linguistic knowledge.
- They allow semantic learning to be accommodated within probabilistic models of grammar induction.

Conclusions

- VSMs provide a powerful representation of lexical semantic relations in which gradience is an intrinsic feature of lexical meaning.
- PGG CVSMs permit a formally grounded and efficient way of computing the vectors of sentences from those of their constituents, but the sentential vectors that they generate from lexical vectors that are distributional do not have a straightforward interpretation.
- Language models enriched with probabilities conditioned by distributional lexical semantic features may offer an alternative framework for integrating lexical and combinatorial dimensions of meaning within a stochastic representation of linguistic knowledge.
- They allow semantic learning to be accommodated within probabilistic models of grammar induction.