

Distributional data for constraints-based semantic parsing

Emmanuele Chersoni

Aix-Marseille University
University of Pisa

Research funded by A*Midex Foundation

Grammar as a set of descriptions

- Integration of the notion of construction in linguistic theories (Fillmore *et al.*, 1988)
- A pairing of form and meaning, describing a specific linguistic phenomenon
- **Ditransitive construction** Subj V Obj1 Obj2
She gave him a kiss
- **Idiomatic construction** *kick the bucket, beating a dead horse*
- No separation between lexicon and syntax: constructions can be lexemes, phrases, sentences...

Properties of a construction

- Constructions result from the convergence of different (syntactic, semantic...) properties
- Property Grammars (Blache: 2000): a framework for description of linguistic phenomena in terms of interacting properties
- Syntactic properties of different types (linearity, co-occurrence, exclusion, uniqueness, dependency)
- Each property represented independently from the others
- Construction described in terms of sets of satisfied / violated properties: we can describe ungrammatical input

PG and semantics

- PG describe all syntactic properties by means of constraints
- Semantic side of the constructions has to be represented too
- What kind of "semantic constraints" should be included in a construction? How do they come into play in semantic composition?

PG and semantics

- Constructions include info on the frames they evoke (Sag: 2012)
- Empty slots to be filled with lexical material matching some properties listed in a feature structure
- Symbolic features, properties of objects and events
- Our proposal: to integrate another dimension of meaning, i.e. word usage

Distributional semantic models

- *We argue that there at least two major types of information from which we learn word meanings. The first is what we call experiential information... the second type of information is language-based. In particular, it is derived from the general linguistic context in which words appear. (Vigliocco: 2009)*

Distributional semantic models

- In **Distributional Semantics** word meaning represented by vectors
- The value of each dimension is (a function of) the frequency of occurrence of a target word within a context (data extracted from large corpora)
- Words having similar vectors will also have similar meanings (cosine the most popular measure of vector similarity)
- In **structured Distributional Semantic models**, co-occurrence statistics in the form of triples: word pairs and the syntactic relation linking them (Baroni, Lenci: 2010)

Distributional semantic models

target	<i>lecture-n</i>	<i>gun-n</i>	<i>speech-n</i>	<i>shoot-v</i>	<i>correct-v</i>
teacher-n	28	0	9	0	28
professor-n	31	0	7	1	29
soldier-n	0	20	2	26	0
politician-n	2	0	27	4	0

Distributional Semantic models

target	depType	context	freq
professor-n	subj	teach-v	37
professor-n	subj	claim-v	15
deer-n	obj	shoot-v	22
soldier-n	subj	fight-v	17
good-j	attr	student-n	23
clock-n	obj	beat-v	11

Distributional semantic models

- Dependency as the generic syntactic-semantic relation linking a dependent to its head
- Dependency relation as a bridge between Construction Grammars and Structured Distributional Semantic Models
- Symbolic features still used in the unification process...
- ... but we can also use the distributional vectors of *prototypical fillers* to measure *how much* a given word fits a slot

Constructions and distributional data

BOY			
FORM	BOY		
SYN	[CAT	N]
		VAL	-
		PROPERTIES	-
SEM	[INDEX	I
		FRAMES	-
		SYM_FEATURES	[LIVING_BEING, HUMAN, MALE...]
		SEM_VECTOR	[PLAY-V 36.3, SCHOOL-N 11.2, YOUNG-J 29.7...]

Constructions and distributional data

VP		
FORM	-	
SYN	CAT	VP
	VAL	NP ¹
	PROPERTIES	$V \prec NP, PP, NP \prec PP, V[\text{TRANS}][\text{ACTIVE}] \rightarrow NP[\text{OBJ}] \dots$
SEM	INDEX	Z
	FRAMES	$\begin{bmatrix} \text{VP:} \\ \text{ARG}^1: - \end{bmatrix}$
	SYM_FEATURES	-
	SEM_VECTOR	-
	ARG_VECTORS	-

Constructions and distributional data

TO_EAT			
FORM	TO_EAT		
SYN	[]	
	CAT	V	
	VAL	NP ¹ , NP ²	
	PROPERTIES	-	
SEM	[]	
	INDEX	J	
	FRAMES	[]
		TO_EAT:	
		[]
		ARG ¹ : <i>eater</i>	ARG ² : <i>meal</i>
		LIVING_BEING	PHYSICAL_OBJECT
			EDIBLE
	SYM_FEATURES	[]
		PHYSICAL_ACTION, CONTINUOUS...	
	SEM_VECTOR	[]
		MEAL-N 52.0, FOOD-N 60.5, DRINK-V 31.7...	
	ARG_VECTORS	[]
		ARG ¹ : []
		MAN-N, DOG-N, BOY-N...	
		ARG ² : []
		FOOD-N, MEAT-N, PASTA-N...	

Constructions and distributional data

- Both symbolic features and distributional vectors to characterize the slots
- Each slot gets assigned the vectors of words typically satisfying the dependency constraint represented by the slot
- The more a word resembles the typical fillers, the more it will be fitting the slot
- Methods for the comparison: exemplar-based, or **prototype-based** (Lenci: 2011). E.g. to get the fitting score of 'apple' as obj of 'to eat', compare the 'apple' vector with the avg vector of (food, meat, pasta...)

Constructions and distributional data

- The result of evaluation of symbolic constraints is a boolean value
- The evaluation of a 'distributional constraint' returns a scalar value
- Even in case of violation, the input still gets a description

Constructions and distributional data

	TO_EAT+AN_APPLE		TO_EAT_AN_APPLE
FORM			
SYN			$\left[\begin{array}{ll} \text{CAT} & \text{VP} \\ \text{VAL} & \text{NP}^1 \\ \text{PROPERTIES} & \left[\begin{array}{ll} \text{INHERITS_FROM} & \text{VP} \\ \text{VIOLATIONS} & \{ \dots \} \end{array} \right] \end{array} \right]$
SEM			$\left[\begin{array}{ll} \text{INDEX} & \text{V} \\ \text{FRAMES} & \left[\begin{array}{ll} \text{INHERITS_FROM: TO_EAT} & \\ \left[\begin{array}{ll} \text{ARG}^1: \textit{eater} & \left[\begin{array}{ll} \text{ARG}^2: \text{AN_APPLE} \\ \text{INDEX_REF} & \text{A} \end{array} \right] \\ \text{LIVING_BEING} & \left[\begin{array}{ll} \text{CONST_SAT} & 1 \end{array} \right] \end{array} \right] \end{array} \right] \\ \text{SYM_FEATURES} & \left[\text{TO_EAT: } \dots , \text{APPLE: } \dots \right] \\ \text{SEM_VECTOR} & \left[\text{BREAKFAST-N } 16.0, \text{KITCHEN-N } 10.5, \text{HUNGRY-N } 2.7\dots \right] \\ \text{ARG_VECTORS} & \left[\text{ARG}^1: \left[\text{MAN-N}, \text{DOG-N}\dots \right], \text{ARG}^2: \left[\text{SEM_SIM} \quad 0.7 \right] \right] \end{array} \right]$

Constructions and distributional data

$\left[\begin{array}{l} \text{THE_BOY} + \text{TO_EAT} + \text{AN_APPLE} \\ \text{FORM} \\ \\ \text{SYN} \\ \\ \\ \text{SEM} \\ \\ \text{SEM_VECTOR} \\ \\ \text{ARG_VECTORS} \end{array} \right.$	$\left[\begin{array}{l} \text{THE_BOY_EATS_AN_APPLE} \\ \\ \left[\begin{array}{l} \text{CAT} \quad \text{S} \\ \text{VAL} \quad - \\ \text{PROPERTIES} \left[\begin{array}{l} \text{INHERITS_FROM} \quad \text{TRANSCONSTRUCTION} \\ \text{VIOLATIONS} \quad \{ \dots \} \end{array} \right] \end{array} \right] \\ \\ \left[\begin{array}{l} \text{INDEX} \quad \text{S} \\ \text{FRAMES} \left[\begin{array}{l} \text{INHERITS_FROM: TO_EAT} \\ \left[\begin{array}{l} \text{ARG}^1: \text{THE_BOY} \quad \left[\begin{array}{l} \text{INDEX_REF} \quad \text{I} \\ \text{CONST_SAT} \quad 1 \end{array} \right] \left[\begin{array}{l} \text{ARG}^2: \text{AN_APPLE} \\ \text{INDEX_REF} \quad \text{A} \\ \text{CONST_SAT} \quad 1 \end{array} \right] \end{array} \right] \\ \text{SYM_FEATURES} \left[\text{THE_BOY: } \dots, \text{ TO_EAT: } \dots \right] \end{array} \right] \end{array} \right] \\ \\ - \\ \left[\text{ARG}^1: \left[\text{SEM_SIM} \quad 0.67 \right], \text{ ARG}^2: \left[\text{SEM_SIM} \quad 0.7 \right] \right] \end{array} \right.$
--	---

Constructions and distributional data: the advantages

- Possibility of taking into account another 'dimension' of meaning: word usage
- Concepts have internal structure: different degrees of prototypicality
- Easy to automatically derive word representations from corpora

Constructions and distributional data: future work

- A method for modeling semantic compositionality with vectors
- Model of activation integrating both PG and semantic constraints

Bibliography

- M. Baroni, A. Lenci, *Distributional Memory: a general framework for corpus-based semantics*, in *Computational Linguistics*, vol. 36, n. 4, 2010, pp. 673-721;
- P. Blache, *Property Grammars and the problem of constraint satisfaction*, in *Proceedings of the ESSLLI Workshop on Linguistic Theory and Grammar Implementations*, 2000;
- C. Fillmore et al., *Regularity and idiomaticity in grammatical constructions*, in *Language*, vol. 64, no. 3, 1988, pp. 501-538;
- A. Lenci, *Composing and updating verb argument expectations: a distributional semantic model*, in *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, 2011, pp. 58-66;
- I. Sag, *Sign-based Construction Grammar: an informal synopsis*, in *Sign-based Construction Grammar*, 2012, pp. 61-197;
- G. Vigliocco et al., *Toward a theory of semantic representation*, in *Language and Cognition*, vol. 1, n. 2, 2009.