Multiword Expressions and Collocations

Valia Kordoni
kordonie@anglistik.hu-berlin.de

Humboldt-Universität zu Berlin (Germany)

LACompLing2017, Stockholm 17.08.2017

Road Map I

1. MWEs: Theoretical Background & Motivation
   - Definitions
   - Characteristics
   - Linguistic and CL Theories
2. MWEs: Computational Methods
   - “Discovering” MWEs
   - Syntax-based Extraction
   - MWE Identification in Context
   - Interpretation
     - Detecting a Continuum of compositionality in PVs
   - NLP Tasks and Applications
3. At the intersection of Deep learning and NLP
   - Beyond learning word vectors

Road Map II

- RNNs for MWEs

4. Resources, tasks and applications
   - Tools
   - Resources
   - Tasks and applications
   - Evaluation

5. Future challenges and open problems

Road Map I

1. MWEs: Theoretical Background & Motivation
   - Definitions
   - Characteristics
   - Linguistic and CL Theories
2. MWEs: Computational Methods
   - "Discovering" MWEs
   - Syntax-based Extraction
   - MWE Identification in Context
   - Interpretation
     - Detecting a Continuum of compositionality in PVs
   - NLP Tasks and Applications
3. At the intersection of Deep learning and NLP
   - Beyond learning word vectors

Road Map II

- RNNs for MWEs

4. Resources, tasks and applications
   - Tools
   - Resources
   - Tasks and applications
   - Evaluation

5. Future challenges and open problems

MWEs: Overview

Multiword Expressions:

- their syntactic or semantic properties cannot be derived from their parts [Sag et al., 2002a, Villavicencio, 2005]
- phrasal verbs (e.g., *come along*), nominal compounds (e.g., *frying pan*), institutionalised phrases (e.g., *bread and butter*)
- equivalent in number to single words in speakers' lexicon [Jackendoff, 1997a]
- fixed (*ad hoc*) vs flexible (*touch/find a nerve*) expressions
- opaque (*kick the bucket*) vs transparent (*eat up*) semantics

MWEs: An Attempt at Definitions

What is a MWE? [Church, 2011]

- A unit whose exact meaning cannot be derived directly from the meaning of its parts [Choueka, 1988]
- Arbitrary and recurrent word combinations [Smadja, 1993]
- A MWE has to be listed in a lexicon [Evert, 2004]
- Idiosyncratic interpretations that cross word boundaries (or spaces) [Sag et al., 2002b]
- A combination of lexemes that must be treated as a unit at some level of linguistic processing. [Calzolari et al., 2002]

MWEs: An Attempt at Definitions

What is a MWE? [Church, 2011]

- A unit whose exact meaning cannot be derived directly from the meaning of its parts [Choueka, 1988]
- Arbitrary and recurrent word combinations [Smadja, 1993]
- A MWE has to be listed in a lexicon [Evert, 2004]
- Idiosyncratic interpretations that cross word boundaries (or spaces) [Sag et al., 2002b]
- A combination of lexemes that must be treated as a unit at some level of linguistic processing. [Calzolari et al., 2002]

MWEs: An Attempt at Definitions

Multiword Expressions:*Definition*

A **multiword expression** (MWE) is [Baldwin and Kim, 2010]

- decomposable into multiple simplex words
- lexically, syntactically, semantically, pragmatically and/or statistically idiosyncratic

MWEs: Some Examples

- *San Fancisco, ad hoc, by and large, Where Eagles Dare, kick the bucket, part of speech, in step, the Oakland Raiders, trip the light fantastic, telephone box, call (someone) up, take a walk, do a number on (someone), take (unfair) advantage of, pull strings, kindle excitement, fresh air, ...*

MWEs: Yet another attempt at a definition

MWE or not MWE?

- *... there is no unified phenomenon to describe but rather a complex of features that interact in various, often untidy, ways and represent a broad continuum between non-compositional (or idiomatic) and compositional groups of words.* [Moon, 1998]

MWEs: Characteristics

Lexicosyntactic Idiomaticity

- by and large (???) = by(P) and(conj) large(Adj)
- wine and dine (V [trans]) = wine (V [intrans]) and(conj) dine (V [intrans])
- ad hoc (Adj) = ad(?) hoc(?)

MWEs: Characteristics

Semantic Idiomaticity

- *kick the bucket* = die’
- *spill the beans* = reveal’ (secret’)
- *kindle excitement* = kindle’ (excitement’)

MWEs: Characteristics

Pragmatic Idiomaticity

- Situatedness: the expression is associated with a fixed pragmatic point
    - situated MWEs: *good morning, all aboard*
    - non-situated MWEs: *first off, to and fro*

MWEs: Characteristics

Statistical Idiomaticity

	unblemished	spotless	flawless	immaculate	impeccable
eye	–	–	–	–	+
gentleman	–	–	?	–	+
home	?	+	–	+	?
lawn	–	–	?	+	–
memory	–	–	+	–	?
quality	–	–	–	–	+
record	+	+	+	+	+
reputation	+	–	–	+	+
taste	–	–	–	–	+

Table: Adapted from [Cruse, 1986]

MWEs: Characteristics

MWE Markedness

MWE	Marked				
	Lex	Syn	Sem	Prag	Stat
ad hominem	✔	?	?	?	✔
at first	✗	✔	✗	✗	✗
first aid	✗	✗	✔	✗	?
salt and pepper	✗	✗	✗	✗	✔
good morning	✗	✗	✗	✔	✔
cat's cradle	✔	✔	✔	✗	?

MWEs:Characteristics

Other Indicators of MWE-hood ([Fillmore et al., 1988a], [Liberman and Sproat, 1992], [Nunberg et al., 1994])

- Institutionalisation/conventionalisation: *bread and butter*
- Non-identifiability: meaning cannot be predicted from surface form
  - idiom of decoding (non-identifiable): *kick the bucket, fly off the handle*
  - idiom of encoding (identifiable): *wide awake, plain truth*

MWEs: Characteristics

Other Indicators of MWE-hood ([Fillmore et al., 1988a], [Liberman and Sproat, 1992], [Nunberg et al., 1994])

- Figuration: the expression encodes some metaphor, metonymy, hyperbole, etc.
  - figurative expressions: *bull market, beat around the bush*
  - non-figurative expressions: *first off, to and fro*

MWEs: Characteristics

Other Indicators of MWE-hood ([Fillmore et al., 1988a], [Liberman and Sproat, 1992], [Nunberg et al., 1994])

- Single-word paraphrasability: the expression has a single word paraphrase
  - paraphrasable MWEs: *leave out = omit, take off clothes = undress*
  - non-paraphrasable MWEs: *look up*

MWEs: Theoretical Linguistic Background

The study of MWEs

- is almost as old as linguistics itself
- in the traditional generative grammatical framework, the representation of idioms poses a challenge, e.g., the idiom *first off* is an adverbial locution synonym to *firstly*
- in Construction Grammar, [Fillmore et al., 1988b]
    - suggest that there must be an appendix to the set of lexical units and syntactic rules of a language model for *idiomatic* entries and their specific syntactic, semantic and pragmatic charecteristics
    - this way, idioms can become part of the core of the grammar: that is, a language can be fully described by its idioms and their properties

MWEs: Theoretical Linguistic Background

In *meaning-text theory* (MTT)

- [Mel'čuk and Polguère, 1987] suggest that a dictionary entry contains three zones: (i) the semantic zone, (ii) the syntactic zone, and (iii) the lexical combinatorics zone
- MWEs are present at two points of the computational MTT model: as *phrasemes* and as *lexico-semantic functions (LSF)* in the so called *lexical combinatorics zone*

MWEs: Theoretical Linguistic Background

In psycholinguistics and cognitive linguistics, there has been work on learning

- verb-particle constructions [Villavicencio et al., 2012]
- noun compounds [Devereux and Costello, 2007]
- light verb constructions and
- multiword terms [Lavagnino and Park, 2010] based on corpora evidence and sophisticated cognitive models; these models try to validate computational models for MWE acquisition by checking their correlation with experiments that use similar models for human language acquisition [Joyce and Srdanović, 2008, Rapp, 2008]

MWEs: CL Background

In computational linguistics

- the study of MWEs arose from the availability of very large corpora and of computers capable of analysing them by the end of the 80's and beginning of the 90's
- the aim was to build systems for computer-assisted lexicography and terminography of multiword units [Choueka, 1988]
- [Smadja, 1993] proposed Xtract, a tool for collocation extraction based on some simple POS filters and on mean and standard deviation of word distance
- [Church and Hanks, 1990] suggested a more sophisticated association measure based on mutual information

MWEs: CL Background

In computational linguistics

- later, [Dagan and Church, 1994] proposed a terminographic environment called Termight, which uses this association score, performs bilingual extraction, and provides tools to easily classify candidate terms, find bilingual correspondences, define nested terms and investigate occurrences through a concordancer
- [Justeson and Katz, 1995] proposed a simple approach based on a small set of POS patterns and frequency thresholds

MWEs: CL Background

In computational linguistics

- [Dunning, 1993] proposed a 2-gram measure called *likelihood ratio*. It estimates directly how more likely a 2-gram is than expected by chance. In addition to being theoretically sound, Dunning's score is also easily interpretable. Nowadays, measures based on likelihood ratio (e.g., the log-likelihood score) are still largely employed in several MWE extraction contexts

MWEs: CL Background

In computational linguistics

- At the beginning of the 2000's, the Stanford MWE project (`http://mwe.stanford.edu/`) has revived interest of the NLP community in this topic. One of the most cited publications of the MWE project is the famous "pain-in-the-neck" paper by [Sag et al., 2002b]. It provided an overview of MWE characteristics and types and then presented some methods for dealing with them in the context of grammar engineering. The Stanford MWE project is also at the origin of the MWE workshop series

MWEs: their semantics

MWEs and the Notion of Compositionality: *Definition*

- degree to which the features of the parts of an MWE combine to predict the features of the whole

MWEs: their semantics

MWEs and the Notion of Compositionality

- Generally considered in the context of semantic compositionality, but we can equally talk about:
  - lexical compositionality
  - syntactic compositionality
  - pragmatic compositionality

MWEs: their syntax

Example: Syntactic Compositionality

- **Definition**: *Degree to which the syntactic features of the parts of an MWE combine to predict the syntax of the whole*
  - Fixed expression: *by and large, San Francisco*
  - Verb particles: *eat up* vs. *chicken out*
- Syntactic compositionality binary effect: non-compositional MWEs lexicalised

MWEs: their semantics and syntax

Question

Given that compositionality extends over all aspects of markedness that affect MWEs, is it all we need to take into consideration?

Almost, but there are subtleties due to:

- statistical markedness
- decomposability

MWEs: their semantics and syntax

Question

Given that compositionality extends over all aspects of markedness that affect MWEs, is it all we need to take into consideration?

Almost, but there are subtleties due to:

- statistical markedness
- decomposability

MWEs: back to statistics

Statistical Markedness (Revisited)

- Statistical markedness is not a lack of compositionality:
  1. $p(\text{impeccable N}) \times p(\text{Adj eye}) \approx p(\text{impeccable eye})$

     BUT

  2. $p(\text{unblemished N}) \times p(\text{Adj eye}) \gg p(\text{unblemished eye})$
  3. $p(\text{spotless N}) \times p(\text{Adj eye}) \gg p(\text{spotless eye})$
  4. $p(\text{flawless N}) \times p(\text{Adj eye}) \gg p(\text{flawless eye})$

MWEs: their importance for Linguistics and CL

And why is it that we care about MWEs?

- Because of the role of MWEs in:
  - Lexicography/dictionary making
  - Idiomaticity (coherent semantics)
  - Overgeneration
  - Undergeneration
  - Relevance in NLP and LT applications, including MT, IR, QA, ...

Road Map I

1. MWEs: Theoretical Background & Motivation
   - Definitions
   - Characteristics
   - Linguistic and CL Theories
2. MWEs: Computational Methods
   - "Discovering" MWEs
   - Syntax-based Extraction
   - MWE Identification in Context
   - Interpretation
     - Detecting a Continuum of compositionality in PVs
   - NLP Tasks and Applications
3. At the intersection of Deep learning and NLP
   - Beyond learning word vectors

Road Map II

- RNNs for MWEs

4 Resources, tasks and applications
- Tools
- Resources
- Tasks and applications
- Evaluation

5 Future challenges and open problems

MWEs: Computational Methods

Overview

Adapted from [Anastasiou et al., 2009]

- **Acquisition**
  - **Extraction**
    How can we build a list of MWE types from corpora?
  - **Identification**
    How can we locate the tokens that correspond to MWEs *in context*?

MWEs: Computational Methods

Overview (contd.)

Adapted from [Anastasiou et al., 2009]

- **Classification**
  - **Interpretation**
    How can we discover the syntactic and semantic relations between the units that compose a MWE type?
  - **Disambiguation**
    How can we disambiguate the syntactic and semantic properties of a MWE token *in context*?

MWEs: Computational Methods

Overview (contd.)

Adapted from [Anastasiou et al., 2009]

- **Representation**
  How can we represent complex MWEs in computational lexicons?
- **Tasks and applications**
  How can we integrate MWEs in NLP tasks (parsing, WSD) and applications (IR, MT)?

MWEs: Computational Methods

"Discovering" MWEs: Co-occurrences

- If *a word is characterized by the company it keeps* [Firth, 1957] then we can try to find MWEs using information about how often words co-occur together
- **Hypothesis**: the more frequently some words occur together, the more likely it is that they form a MWE

"Discovering" MWEs: Filtering with Association Measures

Statistical association measures (AMs)

- can give indication of strength of the association between words (or n-grams)
  - based on frequency of words individually and as a group

"Discovering" MWEs - Filtering with Association Measures

AMs for Ranking MWE Candidates

- **Hypothesis**: If the words are dependent then the candidate is a MWE
  1. Determine the probability given by the *Null Hypothesis* (that they are independent)
  2. Compare with the probability given by a statistical measure
     - t-test, Pearson's X^2, Pointwise Mutual Information, Mutual Information, ...
  3. If Null Hypothesis is rejected then they are dependent (MWEs)

"Discovering" MWEs: Alternative Measures: Entropy-based

Permutation Entropy

Hypothesis: MWEs prefer a certain word order (*give a demo* vs *a demo give*)

- If a candidate is result of random combination of words then word order in n-gram is not important: *of alcohol and*, *and of alcohol*, *alcohol and of*, etc
- Entropy: $S = -\frac{1}{\log N} \sum_{perm} P(abc) \log P(abc)$: $S \rightarrow 0$ (prevalent order) $\longrightarrow$ possible MWE

MWE	Pages	S
the burden of	36,600,000	0.366
but also in	27,100,000	0.038
to bring together	25,700,000	0.086
points of view	24,500,000	0.017
and the more	23,700,000	0.512
taking into account the	22,100,000	0.009

Evaluation of the Extraction of MWEs

Factors in MWE Extraction [Evert and Krenn, 2005]

- corpus size and type
- MWE type and language
- AMs

Comparison of AMs

- 84 measures among which some are rank-equivalent to one another [Pecina, 2008]
- comparison of their combination [Ramisch et al., 2008]

More on Evaluation of the Extraction of MWEs

For statistical approaches there are two important questions

- Q1 How reliable/generalizable are the results for a given corpus?
- Q2 How precise an association measure is to distinguish MWEs from noise?

Evaluation of the Extraction of MWEs - Comparing corpora

Q1: How reliable/generalizable are the results for a given corpus?

- **Hypothesis**: relative candidate rankings are preserved across similar corpora
  - If not, different conclusions may be drawn from different corpora
- Evaluation
  - list of MWE candidates
  - 4 corpora
    - standard: BNC vs fragment of the BNC (BNC_f)
    - WACs: Google vs Yahoo

Evaluation of the Extraction of MWEs - Comparing corpora

Relative Frequency Rank for the Trigrams

- Kendall's τ scores between corpora show significant correlation ($p \leq 0.000001$)
- A higher correlation was observed between Yahoo and Google: as corpora sizes increase, so do the correlations between them

	BNC	Google	Yahoo
BNC_f	0.81	0.73	0.78
BNC		0.73	0.77
Google			0.86

Evaluation of the Extraction of MWEs - Comparing AMs

Q2: How precise an association measure is to distinguish MWEs from noise?

- Using a single corpus (BNC_f), comparing MI, χ^2 and Permutation Entropy (PE)
- Kendall's τ for assessing the correlation of the rankings for these AMs and Q is the probability of finding the same ordering in them

	MI $\times \chi^2$	MI $\times$ PE	$\chi^2 \times$ PE
Q	0.71	0.55	0.45

- The correlations found are statistically significant
- The measures order the trigrams differently
  - 70% chance of getting the same order from MI and χ^2

Verb-particle Constructions (VPCs) [Baldwin, 2005]

- VPC = verb + obligatory particle(s) (e.g.*hand in*, *battle on*)
  - **intransitive**: e.g. *the team battled on*
  - **transitive**: e.g. *Kim handed the paper in*
- Variable word order for transitive VPCs:
  - **joined**: *<u>hand</u> <u>in</u> the paper*
  - **split**: *<u>hand</u> the paper <u>in</u>*
- Structural/analytical ambiguity:
  - *hand [the paper] [in] [here]* vs. *hand [the paper] [in here]* vs. *hand [the paper in here]*
  - *hand [in] [the paper]* vs. *hand [in the paper]*

Token Identification

MWE or not?

- *The cook/journalist **spilled the beans***
- [Kim and Baldwin, 2010] identify token instances of VPCs from output of parser
  - head nouns can help distinguish if V and P in a given sentential context are
    - VPC: *Kim **handed in** the paper* or
    - verb-PP: *Kim **walked in** the room*
  - post-processing VPC identification
  - combine syntactic and semantic features
  - using sentential context of instances of VPCs and verb-PPs

Token Identification

MWE or not?

- Semantics of VPC may be
  - derived from semantics of verb and particle: *walk off*
  - different from them: *look up*;
- Selectional preferences of VPCs may
  - mirror those of the verbs: *clean* and *clean up*
  - diverge: *put the book on the table* and *put on a sweater*;

Different selectional preferences for verb in isolation or in VPC

Token Identification

Method

1. From parser output identify:
   - verbs and particles and transitive prepositions
   - head nouns of subject and object of each verb
2. Obtain lexical semantics of the head nouns
   - based on WordNet 2.1 [Fellbaum, 1998]
   - using the first sense for that word in SemCor (Landes et al., 1998)
3. Build a classifier
   - feature vector for VPC and verb-PP: *(verb, preposition, subject WN class, object WN class)*

Detecting a Continuum of compositionality in PVs

MWE Identification with Distributional Thesaurus

- Compositionality detection [McCarthy et al., 2003] via automatically acquired thesaurus [Lin, 1998]
  - Thesaurus contains 500 neighbors for verb and VPC
  - from subject and object grammatical relations of simplex verb and VPC
- Indicators of VPC Compositionality:
  - overlap of neighbors of simplex and VPC
  - neighbors of VPC include other VPCs with same particle
  - verb neighbor of VPC, etc

Detecting a Continuum of compositionality in PVs

Evaluation: against human judgements

- 116 annotated VPC
- 3 native speakers
  - scale from 0 (non-compositional) to 10 (fully compositional)
- Correlation with
  - frequency of VPC and verb
  - resources
    - Wordnet, Alvey Tools Lexicon
  - AMs
    - χ^2, log likelihood ratio, PMI

Grammar Engineering and Parsing

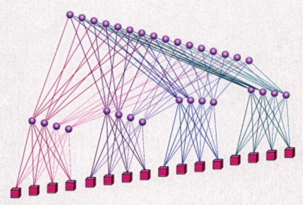
- Lexical coverage is a major barrier to broad-coverage linguistically deep processing
  - 40% parsing failures caused by missing lexical entries [Baldwin et al., 2004]
- MWEs are a significant part of the lexicon
  - Detect potential errors in parsing involving sequences of words
  - Identify MWE candidates
  - Generate new lexical entries based on corpus data

Extension of a hand-crafted linguistic resource with MWEs: English Resource Grammar [Flickinger, 2000]

- A large scale broad coverage precision HPSG grammar
- Lexicon coverage is a major problem
- MWEs comprise a large portion of the missing lexical entries

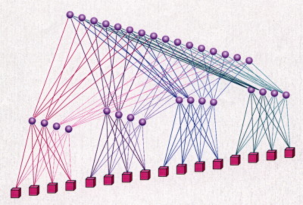
Lexical hierarchy and atomic lexical types

- The lexical information is encoded in atomic lexical types
- A lexicon is a $n : n$ mapping between lexemes and atomic lexical type



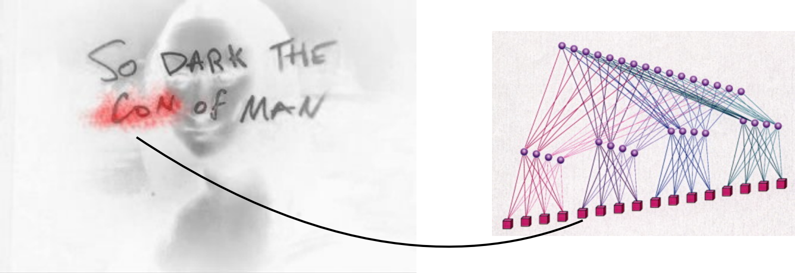
Lexical hierarchy and atomic lexical types

- The lexical information is encoded in atomic lexical types
- A lexicon is a $n : n$ mapping between lexemes and atomic lexical type



Lexical hierarchy and atomic lexical types

- The lexical information is encoded in atomic lexical types
- A lexicon is a $n : n$ mapping between lexemes and atomic lexical type



Maximum Entropy Model-based Lexical Type Predictor

- A statistical classifier that predicts for each occurrence of an unknown word or a missing lexical entry
- Input: features from the context
- Output: atomic lexical types

$$p(t,c) = \frac{exp(\sum_i \theta_i f_i(t,c))}{\sum_{t' \in T} exp(\sum_i \theta_i f_i(t',c))}$$

"Words-with-spaces" vs. compositional approaches

Words-with-spaces approach [Zhang et al., 2006]

- Assign lexical types for the entire MWE
- Grammar coverage significantly improves
- Loss in generality for productive MWEs

Compositional approach

- Assign new lexical entries for the head word to treat the MWE as compositional
- Hopefully the grammar coverage improves without drop in accuracy

"Words-with-spaces" vs. compositional approaches

Words-with-spaces approach [Zhang et al., 2006]

- Assign lexical types for the entire MWE
- Grammar coverage significantly improves
- Loss in generality for productive MWEs

Compositional approach

- Assign new lexical entries for the head word to treat the MWE as compositional
- Hopefully the grammar coverage improves without drop in accuracy

Experiment

- Rank all the MWE candidates according to the three statistical measures: MI, χ^2, PE, and select the top 30 MWE with highest average ranking
- Extract sub-corpus from BNC_f which contains at least one of the MWE for evaluation (674 sentences)
- Use heuristics to extract head words (20 head words)
- Run lexical acquisition for head words on the sub-corpus (21 new entries)

Grammar Coverage

	item #	parsed #	avg. analysis #	coverage %
ERG	674	48	335.08	7.1%
ERG + MWE	674	153	285.01	22.7%

- The coverage improvement is largely compatible with the results of "words-with-spaces" approach reported in [Zhang et al., 2006] (about 15%)
- Great reduction in lexical entries added

Grammar Accuracy

- 153 parsed sentences are analyzed by hand
- 124 (81.0%) of them receive at least one correct/acceptable analysis (comparable to the accuracy reported by [Baldwin et al., 2004])
- Parse selection model finds best analysis in top-5 for 66% of the cases, and top-10 for 75%

Outlook

- Hand-crafted precision grammars usually face coverage/robustness challenges when applied to unseen data with unknown words/MWEs, unknown constructions, etc., all over the place
- [Baldwin et al., 2004] reported parsing coverage of **18**% on unseen BNC data parsed with the ERG, with the majority of parsing failures related to missing lexical entries
- The Lexical Type Prediction model presented as an example above is used to handle unknown words (simplex and MWE) on-the-fly
- With the use of this model the ERG achieves around **84**% parsing coverage on unseen WSJ data

Road Map I

1 MWEs: Theoretical Background & Motivation
- Definitions
- Characteristics
- Linguistic and CL Theories

2 MWEs: Computational Methods
- "Discovering" MWEs
- Syntax-based Extraction
- MWE Identification in Context
- Interpretation
  - Detecting a Continuum of compositionality in PVs
- NLP Tasks and Applications

3 At the intersection of Deep learning and NLP
- Beyond learning word vectors

Road Map II

- RNNs for MWEs

4 Resources, tasks and applications
- Tools
- Resources
- Tasks and applications
- Evaluation

5 Future challenges and open problems

DL for phrases

Semantic Compositionality through Recursive Matrix-Vector (RNN) spaces [Socher et al., 2012]

- an RNN model which learns compositional vector representations for phrases and sentences of arbitrary syntactic type and length
- the model assigns a vector and a matrix to every node in a parse tree: the vector captures the inherent meaning of the constituent, while the matrix captures how it changes the meaning of neighbouring words or phrases
- it may handle compositional MWEs

DL for phrases

Motivation

- despite their success, single word vector models are severely limited since they do not capture compositionality, the important quality of natural language that allows speakers to determine the meaning of a longer expression based on the meanings of its words and the rules used to combine them
- this prevents them from gaining a deeper understanding of the semantics of longer phrases, like MWEs, or sentences

DL for phrases

Overview of RNN Model Variations

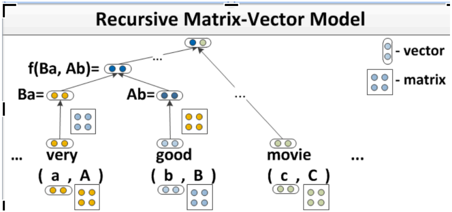
- Objective Functions
  - Supervised Scores for Structure Prediction
  - Unsupervised autoencoding immediate children or entire tree structure
- Composition Functions
  - Syntactically-Untied Weights
  - Matrix Vector RNN / Tensor-Based Models
- Tree Structures
  - Constituency Parse Trees
  - Dependency Parse Trees
  - CCG Trees
  - Fixed Tree Structures (Connections to CNNs)

Learning vector representations for MWEs

How

- an RNN learns compositional vector representations of vectors or sentences of arbitrary length or syntactic type
- a vector and a matrix are assigned to every node in the parse tree:
  - the vector captures the inherent meaning of the word
  - the matrix captures how the word modifies the neighbouring words
- A representation for a longer phrase, like a MWE, is computed in a bottom-up manner by recursively combining children words according to the syntactic structure in the parse tree

Recursive Matrix-Vector Model



Road Map I

1. MWEs: Theoretical Background & Motivation
   - Definitions
   - Characteristics
   - Linguistic and CL Theories
2. MWEs: Computational Methods
   - “Discovering” MWEs
   - Syntax-based Extraction
   - MWE Identification in Context
   - Interpretation
     - Detecting a Continuum of compositionality in PVs
   - NLP Tasks and Applications
3. At the intersection of Deep learning and NLP
   - Beyond learning word vectors

Road Map II

- RNNs for MWEs

4. Resources, tasks and applications
   - Tools
   - Resources
   - Tasks and applications
   - Evaluation

5. Future challenges and open problems

Tools for acquisition

mwetoolkit

- Multi-level patterns for candidate generation
- Several filtering methods
- Focused on genericity and flexibility

`http://mwetoolkit.sourceforge.net`

[Ramisch et al., 2010a, Ramisch et al., 2010b]

Tools for acquisition

Embedded

- FIPS parser [Seretan and Wehrli, 2009, Seretan and Wehrli, 2011]
- Stanford parser [Green et al., 2011]
- Phrasal verbs in RASP
- Most parsers include (minimal) MWE processing

Tools for acquisition

Related tools

- Complex corpus searches: CQP [Christ, 1994] and Manatee [Rychlý and Smrz, 2004]
- Terminology extraction
  - TermoStat
    http://olst.ling.umontreal.ca/~drouinp/termostat_web/
  - AntConc
    http://www.antlab.sci.waseda.ac.jp/software.html
  - TerMine
    http://www.nactem.ac.uk/software/termine/
- Named entity recognition

Tools for acquisition

Which one to choose? [Ramisch et al., 2012]

	LocMax	mwetk	NSP	UCS
Cand. extr.	+	+	+	–
N-grams $n > 2$	+	+	+	–
Non-adjacent	–	+	+	
Ling. filter	–	+	–	–
Robust measures	–	–	+	+
Large corpora	Partly	+	+	–
Language independent	+	Partly	+	+
Token identification	–	–	–	–
Availability	Free	Free	Free	Free

Resources

Why do we need MWE acquisition?

- MWEs are very frequent in human languages [Jackendoff, 1997b]
- Computational resources (corpora, grammars, lexicons) do not reflect this

Resources

Corpora

- At least 17% of Europarl sentences contain a phrasal verb
- 70% of terms in Genia are multiwords
- Flat annotation of noun compounds in treebanks (PTB, French treebank, etc)

Resources

Wordnet

	Non-MWE	**MWE**
Nouns	57535	60292
Verbs	8729	2829
Adverbs	3796	714
Adjectives	21012	496

- Other languages?
- Missing MWE types (e.g. support verb constructions)?
- New expressions?

Resources

Wordnet

	Non-MWE	MWE
Nouns	57535	60292
Verbs	8729	2829
Adverbs	3796	714
Adjectives	21012	496

- Other languages?
- Missing MWE types (e.g. support verb constructions)?
- New expressions?

Tasks and applications

- Parsing
- Information retrieval
- Word sense disambiguation
- Machine translation
- Educational testing
- Sentiment analysis

Tasks and applications

Parsing

- Small set of fixed MWEs (e.g. conjunctions) in most parsers, chunkers and POS taggers
- Joining contiguous nominal expressions with an underscore prior to parsing [Korkontzelos and Manandhar, 2010]
- Extend handcrafted computational grammars with MWEs [Zhang and Kordoni, 2006, Villavicencio et al., 2007]
- Named entities replaced by placeholders [Hogan et al., 2011]

Tasks and applications

Parsing (contd.)

- Feature for disambiguation [Wehrli et al., 2010]
- Tree substitution grammars [Green et al., 2011]
- Joint MWE identification and POS tagging with CRF [Constant and Sigogne, 2011]

Overgeneration vs undergeneration

Tasks and applications

Machine translation (statistical)

- Phrases in Moses [Koehn et al., 2007]
- Static and dynamic strategies for English MWEs from Wordnet [Carpuat and Diab, 2010]
- Monolingual paraphrases for increasing training data [Nakov, 2008a]
- Pre- and post-processing for German compounds [Stymne, 2011, Stymne, 2009]
- Named entities and compound verbs tokenisation [Pal et al., 2010]
- Corpus and phrase-table artificial extensions [Ren et al., 2009]

Evaluation context

1. What are the acquisition goals (that is, the target applications) of the resulting MWEs?
2. What is the nature of the evaluation measures that we intend to use?
3. What is the cost of the resources (dictionaries, reference lists, human experts) required for the desired evaluation?
4. How ambiguous are the target MWE types?

Evaluation context

Acquisition goals

- **Intrinsic**: Evaluate the MWEs per se, using human annotation or gold standard dictionaries.
- **Extrinsic**: Evaluate an application output which includes MWE acquisition.

Evaluation context

Nature of evaluation measures

- **Quantitative**: Objective measures (precision, recall, P@100, MAP)
- **Qualitative**: More fine-grained characterisation of errors.

Evaluation context

Availability of resources

- **Manual annotation.** Sample annotated by group of (expert) native speakers. Realiability depends on quality of guidelines, agreement, sample size, etc.
- **Automatic annotation.** Compare extracted MWEs with existing gold standard (dictionary, thesaurus). Assumes the gold standard is complete.

- A mixture of both is commonly employed.

Evaluation context

Type of target MWEs

- **Type-based evaluation.** Non-ambiguous expressions, can be evaluated out of context. Several resources available on MWE website [Laporte and Voyatzi, 2008, Krenn, 2008, Nicholson and Baldwin, 2008, Nakov, 2008b]
- **Token-based evaluation.** Target MWEs are ambiguous (e.g. phrasal verbs, idioms). Fewer resources available [Cook et al., 2007, Cook et al., 2008, Baldwin, 2008, Fritzinger et al., 2010].

Acquisition context

Generalisation of evaluation results depends on parameters of acquisition context:

- Characteristics of target MWEs
  - Type
  - Language
  - Domain
- Characteristics of corpora
  - Size
  - Nature
  - Level of analysis
- Existing resources

Road Map I

1. MWEs: Theoretical Background & Motivation
   - Definitions
   - Characteristics
   - Linguistic and CL Theories
2. MWEs: Computational Methods
   - “Discovering” MWEs
   - Syntax-based Extraction
   - MWE Identification in Context
   - Interpretation
     - Detecting a Continuum of compositionality in PVs
   - NLP Tasks and Applications
3. At the intersection of Deep learning and NLP
   - Beyond learning word vectors

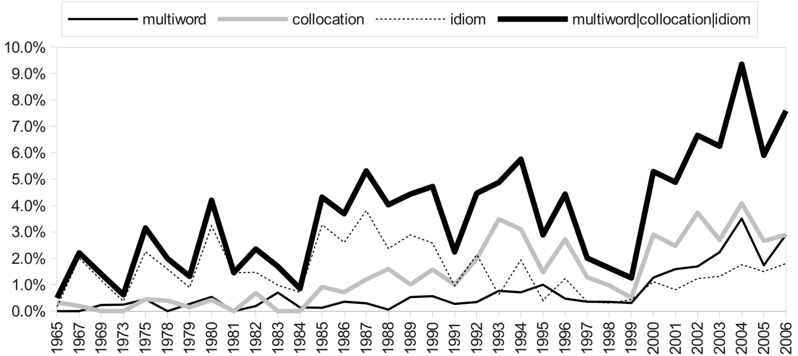
Road Map II

- RNNs for MWEs

4. Resources, tasks and applications
   - Tools
   - Resources
   - Tasks and applications
   - Evaluation

5. Future challenges and open problems

MWE community



MWE community

Trending topics

- Semantics
- Multilingualism
- Applications
- Evaluation
- Machine learning

MWE community

Current and future activities

- MWE workshop series
- ACM TSLP Special Issue on MWEs in 2 parts (`http://dl.acm.org/citation.cfm?id=2483691&picked=prox`)
- SIGLEX-MWE Section (`http://multiword.sourceforge.net/`)
- ICT COST Action IC1207: Parsing and multiword expressions - Towards linguistic precision and computational efficiency in natural language processing (PARSEME; `http://www.cost.eu/domains_actions/ict/Actions/IC1207`)

MWE community

Future challenges

- Identification is not a solved problem
- Integration and representation in applications
- Robust methods for new MWEs in web texts

Further reading

Please refer to complete list of references :-)

Further reading

Thank you!

For Further Reading I

(1993).
Comp. Ling., 19(1).

(2005).
Comp. Speech & Lang. Special issue on MWEs, 19(4).

(2010a).
Proc. of HLT: The 2010 Annual Conf. of the NAACL (NAACL 2003), Los Angeles, California. ACL.

(2010b).
Proc. of the Seventh LREC (LREC 2010), Valetta, Malta. ELRA.

Anastasiou, D., Hashimoto, C., Nakov, P., and Kim, S. N., editors (2009).
Proc. of the ACL Workshop on MWEs: Identification, Interpretation, Disambiguation, Applications (MWE 2009), Suntec, Singapore. ACL.
70 p.

Baldwin, T. (2005).
Deep lexical acquisition of verb-particle constructions.
In [jou, 2005], pages 398–414.

Baldwin, T. (2008).
A resource for evaluating the deep lexical acquisition of English verb-particle constructions.
In [Grégoire et al., 2008], pages 1–2.

For Further Reading II

Baldwin, T., Bender, E. M., Flickinger, D., Kim, A., and Oepen, S. (2004).
Road-testing the English Resource Grammar over the British National Corpus.
In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal.

Baldwin, T. and Kim, S. N. (2010).
Multiword expressions.
In Indurkhya, N. and Damerau, F. J., editors, *Handbook of Natural Language Processing*, pages 267–292. CRC Press, Taylor and Francis Group, Boca Raton, FL, USA, 2 edition.

Calzolari, N., Fillmore, C., Grishman, R., Ide, N., Lenci, A., Macleod, C., and Zampolli, A. (2002).
Towards best practice for multiword expressions in computational lexicons.
In *Proc. of the Third LREC (LREC 2002)*, pages 1934–1940, Las Palmas, Canary Islands, Spain. ELRA.

Carpuat, M. and Diab, M. (2010).
Task-based evaluation of multiword expressions: a pilot study in statistical machine translation.
In [con, 2010a], pages 242–245.

Choueka, Y. (1988).
Looking for needles in a haystack or locating interesting collocational expressions in large textual databases.
In Fluhr, C. and Walker, D. E., editors, *Proceedings of the 2nd International Conference on Computer-Assisted Information Retrieval (Recherche d'Information et ses Applications - RIA 1988)*, pages 609–624, Cambridge, MA, USA. CID.

For Further Reading III

Christ, O. (1994).
A modular and flexible architecture for an integrated corpus query system.
In *COMPLEX 1994*, pages 23–32, Budapest, Hungary.

Church, K. (2011).
How many multiword expressions do people know?
In [Kordoni et al., 2011], pages 137–144.

Church, K. and Hanks, P. (1990).
Word association norms mutual information, and lexicography.
Comp. Ling., 16(1):22–29.

Constant, M. and Sigogne, A. (2011).
MWU-aware part-of-speech tagging with a CRF model and lexical resources.
In [Kordoni et al., 2011], pages 49–56.

Cook, P., Fazly, A., and Stevenson, S. (2007).
Pulling their weight: Exploiting syntactic forms for the automatic identification of idiomatic expressions in context.
In Grégoire, N., Evert, S., and Kim, S. N., editors, *Proc. of the ACL Workshop on A Broader Perspective on MWEs (MWE 2007)*, pages 41–48, Prague, Czech Republic. ACL.

For Further Reading IV

Cook, P., Fazly, A., and Stevenson, S. (2008).
The VNC-tokens dataset.
In [Grégoire et al., 2008], pages 19–22.

Cruse, A. (1986).
Lexical Semantics.
Cambridge University Press, Cambridge, UK.

Dagan, I. and Church, K. (1994).
Termight: Identifying and translating technical terminology.
In *Proc. of the 4th ANLP Conf. (ANLP 1994)*, pages 34–40, Stuttgart, Germany. ACL.

Devereux, B. and Costello, F. (2007).
Learning to interpret novel noun-noun compounds: evidence from a category learning experiment.
In Buttery, P., Villavicencio, A., and Korhonen, A., editors, *Proc. of the ACL 2007 Workshop on Cognitive Aspects of Computational Language Acquisition*, pages 89–96, Prague, Czech Republic. ACL.

Dunning, T. (1993).
Accurate methods for the statistics of surprise and coincidence.
In [jou, 1993], pages 61–74.

For Further Reading V

Evert, S. (2004).
The Statistics of Word Cooccurrences: Word Pairs and Collocations.
PhD thesis, Institut für maschinelle Sprachverarbeitung, University of Stuttgart, Stuttgart, Germany.
353 p.

Evert, S. and Krenn, B. (2005).
Using small random samples for the manual evaluation of statistical association measures.
In [jou, 2005], pages 450–466.

Fellbaum, C., editor (1998).
WordNet: An Electronic Lexical Database (Language, Speech, and Communication).
MIT Press.
423 p.

Fillmore, C., Kay, P., and O'Connor, M. (1988a).
Regularity and idiomaticity in grammatical constructions.
Language, 64:501–38.

Fillmore, C. J., Kay, P., and O'Connor, M. C. (1988b).
Regularity and idiomaticity in grammatical constructions: The case of let alone.
Language, 64:501–538.

For Further Reading VI

Firth, J. R. (1957).
Papers in Linguistics 1934-1951.
Oxford UP, Oxford, UK.
233 p.

Flickinger, D. (2000).
On building a more efficient grammar by exploiting types.
Natural Language Engineering, 6(1):15–28.

Fritzinger, F., Weller, M., and Heid, U. (2010).
A survey of idiomatic preposition-noun-verb triples on token level.
In [con, 2010b], pages 2908–2914.

Green, S., de Marneffe, M.-C., Bauer, J., and Manning, C. D. (2011).
Multiword expression identification with tree substitution grammars: A parsing tour de force with French.
In Barzilay, R. and Johnson, M., editors, *Proc. of the 2011 EMNLP (EMNLP 2011)*, pages 725–735, Edinburgh, Scotland, UK. ACL.

Grégoire, N., Evert, S., and Krenn, B., editors (2008).
Proc. of the LREC Workshop Towards a Shared Task for MWEs (MWE 2008), Marrakech, Morocco.

Hogan, D., Foster, J., and van Genabith, J. (2011).
Decreasing lexical data sparsity in statistical syntactic parsing - experiments with named entities.
In [Kordoni et al., 2011], pages 14–19.

For Further Reading VII

Jackendoff, R. (1997a).
Twistin' the night away.
Language, 73:534–59.

Jackendoff, R. (1997b).
Twistin' the night away.
Language, 73:534–559.

Joyce, T. and Srdanović, I. (2008).
Comparing lexical relationships observed within Japanese collocation data and Japanese word association norms.
In [Zock and Huang, 2008], pages 1–8.

Justeson, J. S. and Katz, S. M. (1995).
Technical terminology: some linguistic properties and an algorithm for identification in text.
Nat. Lang. Eng., 1(1):9–27.

Kim, S. N. and Baldwin, T. (2010).
How to pick out token instances of English verb-particle constructions.
Lang. Res. & Eval. Special Issue on Multiword expression: hard going or plain sailing, 44(1-2):97–113.

For Further Reading VIII

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (2007).
Moses: open source toolkit for statistical machine translation.
In *Proc. of the 45th ACL (ACL 2007)*, pages 177–180, Prague, Czech Republic. ACL.

Kordoni, V., Ramisch, C., and Villavicencio, A., editors (2011).
Proc. of the ACL Workshop on MWEs: from Parsing and Generation to the Real World (MWE 2011), Portland, OR, USA. ACL.

Korkontzelos, I. and Manandhar, S. (2010).
Can recognising multiword expressions improve shallow parsing?
In [con, 2010a], pages 636–644.

Krenn, B. (2008).
Description of evaluation resource – German PP-verb data.
In [Grégoire et al., 2008], pages 7–10.

Laporte, É., Nakov, P., Ramisch, C., and Villavicencio, A., editors (2010).
Proc. of the COLING Workshop on MWEs: from Theory to Applications (MWE 2010), Beijing, China. ACL.

Laporte, É. and Voyatzi, S. (2008).
An electronic dictionary of French multiword adverbs.
In [Grégoire et al., 2008], pages 31–34.

For Further Reading IX

Lavagnino, E. and Park, J. (2010).
Conceptual structure of automatically extracted multi-word terms from domain specific corpora: a case study for Italian.
In Zock, M. and Rapp, R., editors, *Proc. of the 2nd COGALEX workshop (COGALEX 2010)*, pages 48–55, Beijing, China. The Coling 2010 Organizing Committee.

Liberman, M. and Sproat, R. (1992).
The stress and structure of modified noun phrases in English.
Lexical Matters – CSLI Lecture Notes, 24:99–108.

Lin, D. (1998).
Automatic retrieval and clustering of similar words.
In *Proc. of the 36th ACL and 17th COLING, Volume 2*, pages 768–774, Montreal, Quebec, Canada. ACL.

McCarthy, D., Keller, B., and Carroll, J. (2003).
Detecting a continuum of compositionality in phrasal verbs.
In Bond, F., Korhonen, A., McCarthy, D., and Villavicencio, A., editors, *Proc. of the ACL Workshop on MWEs: Analysis, Acquisition and Treatment (MWE 2003)*, pages 73–80, Sapporo, Japan. ACL.

Mel'čuk, I. and Polguère, A. (1987).
A formal lexicon in the meaning-text theory or (how to do lexica with words).
Comp. Ling., 13(3-4):261–275.

For Further Reading X

Moon, R. (1998).
Fixed Expressions and Idioms in English: A Corpus-based Approach.
Oxford University Press, Oxford, UK.

Nakov, P. (2008a).
Improved statistical machine translation using monolingual paraphrases.
In Ghallab, M., Spyropoulos, C. D., Fakotakis, N., and Avouris, N. M., editors, *Proc. of the 18th ECAI (ECAI 2008)*, volume 178 of *Frontiers in Artificial Intelligence and Applications*, pages 338–342, Patras, Greece. IOS Press.

Nakov, P. (2008b).
Paraphrasing verbs for noun compound interpretation.
In [Grégoire et al., 2008], pages 46–49.

Nicholson, J. and Baldwin, T. (2008).
Interpreting compound nominalisations.
In [Grégoire et al., 2008], pages 43–45.

Nunberg, G., Sag, I., and Wasow, T. (1994).
Idioms.
Language, 70:491–538.

For Further Reading XI

Pal, S., Naskar, S. K., Pecina, P., Bandyopadhyay, S., and Way, A. (2010).
Handling named entities and compound verbs in phrase-based statistical machine translation.
In [Laporte et al., 2010], pages 45–53.

Pecina, P. (2008).
Reference data for Czech collocation extraction.
In [Grégoire et al., 2008], pages 11–14.

Ramisch, C., Araujo, V. D., and Villavicencio, A. (2012).
A broad evaluation of techniques for automatic acquisition of multiword expressions.
In *Proc. of the ACL 2012 SRW*, pages 1–6, Jeju, Republic of Korea. ACL.

Ramisch, C., Schreiner, P., Idiart, M., and Villavicencio, A. (2008).
An evaluation of methods for the extraction of multiword expressions.
In [Grégoire et al., 2008], pages 50–53.

Ramisch, C., Villavicencio, A., and Boitet, C. (2010a).
Multiword expressions in the wild? the mwetoolkit comes in handy.
In Liu, Y. and Liu, T., editors, *Proc. of the 23rd COLING (COLING 2010) — Demonstrations*, pages 57–60, Beijing, China. The Coling 2010 Organizing Committee.

Ramisch, C., Villavicencio, A., and Boitet, C. (2010b).
mwetoolkit: a framework for multiword expression identification.
In [con, 2010b], pages 662–669.

For Further Reading XII

Rapp, R. (2008).
The computation of associative responses to multiword stimuli.
In [Zock and Huang, 2008], pages 102–109.

Ren, Z., Lü, Y., Cao, J., Liu, Q., and Huang, Y. (2009).
Improving statistical machine translation using domain bilingual multiword expressions.
In Anastasiou, D., Hashimoto, C., Nakov, P., and Kim, S. N., editors, *Proc. of the ACL Workshop on MWEs: Identification, Interpretation, Disambiguation, Applications (MWE 2009)*, pages 47–54, Suntec, Singapore. ACL.

Rychlý, P. and Smrz, P. (2004).
Manatee, bonito and word sketches for Czech.
In *Proceedings of the Second International Conference on Corpus Linguisitcs*, pages 124–131, Saint-Petersburg, Russia.

Sag, I., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. (2002a).
Multiword expressions: A pain in the neck for NLP.
In *Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, pages 1–15, Mexico City, Mexico.

Sag, I., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. (2002b).
Multiword expressions: A pain in the neck for NLP.
In *Proc. of the 3rd CICLing (CICLing-2002)*, volume 2276/2010 of *LNCS*, pages 1–15, Mexico City, Mexico. Springer.

For Further Reading XIII

Seretan, V. and Wehrli, E. (2009).
Multilingual collocation extraction with a syntactic parser.
Lang. Res. & Eval. Special Issue on Multilingual Language Resources and Interoperability, 43(1):71–85.

Seretan, V. and Wehrli, E. (2011).
Fipscoview: On-line visualisation of collocations extracted from multilingual parallel corpora.
In [Kordoni et al., 2011], pages 125–127.

Smadja, F. A. (1993).
Retrieving collocations from text: Xtract.
In [jou, 1993], pages 143–177.

Socher, R., Huval, B., Manning, C. D., and Ng, A. Y. (2012).
Semantic compositionality through recursive matrix-vector spaces.
In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL '12, pages 1201–1211, Stroudsburg, PA, USA. Association for Computational Linguistics.

Stymne, S. (2009).
A comparison of merging strategies for translation of German compounds.
In *Proc. of the Student Research Workshop at EACL 2009*, pages 61–69.

For Further Reading XIV

Stymne, S. (2011).
Pre- and postprocessing for statistical machine translation into Germanic languages.
In *Proc. of the ACL 2011 SRW*, pages 12–17, Portland, OR, USA. ACL.

Villavicencio, A. (2005).
The availability of verb-particle constructions in lexical resources: How much is enough?
Journal of Computer Speech and Language Processing, 19.

Villavicencio, A., Idiart, M., Ramisch, C., Araujo, V. D., Yankama, B., and Berwick, R. (2012).
Get out but don't fall down: verb-particle constructions in child language.
In Berwick, R., Korhonen, A., Poibeau, T., and Villavicencio, A., editors, *Proc. of the EACL 2012 Workshop on Computational Models of Language Acquisition and Loss*, pages 43–50, Avignon, France. ACL.

Villavicencio, A., Kordoni, V., Zhang, Y., Idiart, M., and Ramisch, C. (2007).
Validation and evaluation of automatically acquired multiword expressions for grammar engineering.
In Eisner, J., editor, *Proc. of the 2007 Joint Conference on EMNLP and Computational NLL (EMNLP-CoNLL 2007)*, pages 1034–1043, Prague, Czech Republic. ACL.

Wehrli, E., Seretan, V., and Nerima, L. (2010).
Sentence analysis and collocation identification.
In [Laporte et al., 2010], pages 27–35.

For Further Reading XV

Zhang, Y. and Kordoni, V. (2006).
Automated deep lexical acquisition for robust open texts processing.
In *Proc. of the Sixth LREC (LREC 2006)*, pages 275–280, Genoa, Italy. ELRA.

Zhang, Y., Kordoni, V., Villavicencio, A., and Idiart, M. (2006).
Automated multiword expression prediction for grammar engineering.
In *Proceedings of the Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, pages 36–44, Sydney, Australia. Association for Computational Linguistics.

Zock, M. and Huang, C.-R., editors (2008).
Proc. of the COLING 2008 COGALEX workshop (COGALEX 2008), Manchester, UK. The Coling 2008 Organizing Committee.