Robust Semantic Analysis of Multiword Expressions with FrameNet

Miriam R. L. Petrucci and Valia Kordoni
miriamp@icsi.berkeley.edu
evangelia.kordoni@anglistik.hu-berlin.de
Road Map

• Overview of FrameNet
  – Frames, Frame Elements, Lexical Units, Valence Descriptions, Frame-to-Frame Relations
  – FrameNet and NLP

• Introduction to Multiword Expressions (MWEs)
  – Types of MWEs
  – Syntactic and Semantic Characteristics of MWEs
  – Representational Issues in MWEs

• Multiword Expressions in FrameNet
  – FrameNet’s treatment of (certain) MWEs
  – Navigating Lexicon and Grammar
  – Exploiting FrameNet Information on MWEs
Road Map

✓ Overview of FrameNet
  - Frames, Frame Elements, Lexical Units, Valence Descriptions, Frame-to-Frame Relations
  - FrameNet and NLP

• Introduction to Multiword Expressions (MWEs)
  - Types of MWEs
  - Syntactic and Semantic Characteristics of MWEs
  - Representational Issues in MWEs

• Multiword Expressions in FrameNet
  - FrameNet’s treatment of (certain) MWEs
  - Navigating Lexicon and Grammar
  - Exploiting FrameNet Information on MWEs
What is FrameNet?

• A unique knowledge base with information on the mapping of meaning to form through the theory of Frame Semantics (Fillmore 1975, 1985, Fillmore and Atkins 1986, Fillmore and Baker 2010, Fillmore 2012, Fontenelle 2003, Petruck 1996)

• A resource that provides rich semantics for the core English vocabulary based on manually annotated corpus evidence, including valence descriptions for each item analyzed
What’s “in” FrameNet?

• ~ 1,200 semantic frames (including FEs)
• > 13,100 lexical units
• > 200,400 manually annotated examples
• nearly 1,800 frame-to-frame relations constituting a hierarchy of semantic frames
What’s a Frame?

A Semantic Frame is a script-like structure of inferences, linked by linguistic convention to the meanings of linguistic units - here, lexical items - constituting a schematic representation of a situation, object, event, or relation providing the background structure against which words are understood. Each frame identifies a set of frame elements – participants in the frame.
Semantic Frames in FrameNet

• Situation: Being_attached, Being_necessary, Being_strong, Being_wet, etc.
• Event: Apply_heat, Borrowing, Catching_fire, Cooking_creation, Hiring, Revenge, etc.
• Object: Buildings, Containers, Intoxicants, Offenses, People_by_origin, etc.
• Relations: Locative_relation, Spacial_co-location, Interior_profile_relation, Similarity, etc.
What’s “in” a Frame?

• **Frame Definition**
  a prose description of a situation involving various participants and other conceptual roles, each of which constitutes a frame element

• **Frame Elements (FEs):**
  semantic roles as the basic unit of a frame, defined specifically to each frame

• **Lexical Units (LUs):**
  pairing of a lemma and a frame, i.e. “word” in one of its senses; LU evokes a frame
Frame Elements: I

Triple of Information

Frame Element
• semantic role

Grammatical Function
• External, Object, Dependent

Phrase Type
• full range of PTs for language
Frame Elements: II

• Core Frame Element: uniquely define frame
  – Commercial_transaction: Buyer, Seller, Money, Goods
  – Giving: Donor, Recipient, Theme
  – Opinion: Cognizer, Opinion

• Non-core Frame Element: capture aspects of situations, events, more generally
  – Time
  – Place
  – Manner
  – Circumstances
Lexical Unit (LU)

• Pairing between a lemma and a frame
  - *hot* - It’s hot outside today.
    • hot – ambient temperature
  - *hot* - The curry is really hot.
    • hot – spiceness
  - *hot* - She’s one hot lady.
    • hot – desirability
FrameNet Methodology

• characterize frames
• collect words that fit the frames
• study corpus attestation of words (“lexical units”)
• develop descriptive terminology (frame elements)
• annotate a subset of corpus examples to document syntactic and semantic behavior
• automatically summarize annotations to produce valence descriptions that show the grammatical realization of the frame elements
Example Frame: Revenge

The Revenge concept involves a situation in which

a) A has done something to harm B and

b) B takes action to harm A in turn

c) B's action is carried out independently of any legal or other institutional setting
Revenge: Vocabulary

- Nouns: revenge, sanction, reprisal, retribution, retaliation, vengeance....
- Verbs: avenge, revenge, retaliate, get back (at), get even, pay back, exact revenge, take revenge....
- Adjectives: retributive, vengeful, vindictive
FN work: choosing FE names

• Develop a descriptive vocabulary for the components of each frame, called frame elements (FEs).

• Use FE names in labeling the constituents of sentences exhibiting the frame.
Revenge: Frame Elements

• Frame Definition: Because of some injury to something-or-someone important to an avenger (maybe himself), the avenger inflicts a punishment on the offender. The offender is the person responsible for the injury.

• Frame Elements:
  • Avenger,
  • Offender,
  • Injury,
  • Injured_party,
  • Punishment.
Annotating Examples

- Select sentences that exhibit common collocations and show all major syntactic contexts.
- Use the names assigned to FEs in the frame, and label the constituents of sentences that express these FEs.
Annotated Sentence

\[ \text{Nora} \text{ Avenger} \] retaliated \[ \text{against her} \text{ boss} \text{ Offender} \] \[ \text{for being dismissed} \text{ Injury} \] \[ \text{by leaving with the office keys} \text{ Punishment} \].
Summarizing Results

• Automatic processes summarize the results, linking FE\s with information about their grammatical realization.

• Present results in the form of various reports in the public website, in XML format in the data release.
Revenge retaliate.v

Definition:
COD: make an attack or assault in return for a similar attack.

<table>
<thead>
<tr>
<th>Frame Element</th>
<th>Number Annotated</th>
<th>Realization(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avenger</td>
<td>(39)</td>
<td>CNI.-- (2) NP.Ext (37)</td>
</tr>
<tr>
<td>Injured_Party</td>
<td>(1)</td>
<td>PP[on].Dep (1)</td>
</tr>
<tr>
<td>Injury</td>
<td>(38)</td>
<td>DNI.-- (35) PP[against].Dep (2) PP[for].Dep (1)</td>
</tr>
<tr>
<td>Instrument</td>
<td>(3)</td>
<td>PP[with].Dep (3)</td>
</tr>
<tr>
<td>Manner</td>
<td>(1)</td>
<td>AVP.Dep (1)</td>
</tr>
<tr>
<td>Offender</td>
<td>(39)</td>
<td>DNI.-- (36) PP[against].Dep (2) PP[on].Dep (1)</td>
</tr>
<tr>
<td>Place</td>
<td>(1)</td>
<td>PP[at].Dep (1)</td>
</tr>
<tr>
<td>Punishment</td>
<td>(39)</td>
<td>PP[in].Dep (2) PP[with].Dep (3) INI.-- (19) AVP.Dep (2) PP[ing].Dep (12) DNI.-- (1)</td>
</tr>
<tr>
<td>Time</td>
<td>(2)</td>
<td>AVP.Dep (2)</td>
</tr>
</tbody>
</table>
Valence Description

• semantico-syntactic combinatorial possibilities
  – meaning-form-function mappings
  – FrameNet Valence Description
    • Frame Element
    • Grammatical Function
    • Phrase Type
Valence: Mapping Meaning to Form

Revenge retaliate.v

<table>
<thead>
<tr>
<th>Number Annotated</th>
<th>Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 TOTAL</td>
<td>Avenger</td>
</tr>
<tr>
<td>(1)</td>
<td>NP Ext</td>
</tr>
<tr>
<td>3 TOTAL</td>
<td>Avenger</td>
</tr>
<tr>
<td>(3)</td>
<td>NP Ext</td>
</tr>
<tr>
<td>1 TOTAL</td>
<td>Avenger</td>
</tr>
<tr>
<td>(1)</td>
<td>NP Ext</td>
</tr>
<tr>
<td>1 TOTAL</td>
<td>Avenger</td>
</tr>
<tr>
<td>(1)</td>
<td>NP Ext</td>
</tr>
<tr>
<td>31 TOTAL</td>
<td>Avenger</td>
</tr>
<tr>
<td>(1)</td>
<td>CNI</td>
</tr>
<tr>
<td>(1)</td>
<td>CNI</td>
</tr>
<tr>
<td>(2)</td>
<td>NP Ext</td>
</tr>
<tr>
<td>(9)</td>
<td>NP Ext</td>
</tr>
<tr>
<td>(1)</td>
<td>NP Ext</td>
</tr>
</tbody>
</table>
Road Map

• Overview of FrameNet
  – Frames, Frame Elements, Lexical Units, Valence Descriptions, Frame-to-Frame Relations
  – FrameNet and NLP

• Introduction to Multiword Expressions (MWEs)
  – Types of MWEs
  – Syntactic and Semantic Characteristics of MWEs
  – Representational Issues in MWEs

• Multiword Expressions in FrameNet
  – FrameNet’s treatment of (certain) MWEs
  – Navigating Lexicon and Grammar
  – Exploiting FrameNet Information on MWEs
Frame-to-Frame Relations in FN

- Inheritance
- Using
- Subframes
- Precedes
- Perspective_on
- See also
- Inchoative_of
- Causative_of

regular lexical relations
Inheritance

• Relationship between a more general frame, the **parent** frame, and a more specific one, the **child** frame
• Child frame **elaborates** parent frame
• **Corresponding entities**, FE, frame relation, and semantic characteristics, in both child and parent
• Child frame entity is the same as or more specific than in parent frame

Apply_heat **inherits** Intentionally_affect
Using (weak inheritance)

• ...a relationship between a more general frame \textit{(parent)} and a more specific frame \textit{(child)} in which only \textit{some} of the FEs in the parent frame have a corresponding entity in the child frame; if correspondences exist, they are more specific.

\texttt{Cooking\_creation \textcolor{green}{uses} Apply\_heat}
FrameGrapher

![Diagram of FrameGrapher with nodes and connections]

View Frame Reports

**Current Frame:**

Cooking_creation

Legend:

Parent frame ➔ Child frame

Parent ➔ Child Relation Types:
- Inheritance
- Subframe
- Perspective On
- Using
- Constituent Of
- Inclusive Of
- See Also

Ordering Relation:
Subframes

• ...a relationship that characterizes the different (typically, ordered) parts of a complex event in terms of the sequences of states of affairs and transitions between them, each of which can itself be described as a frame.

Getting_a_job is a subframe of Employee_scenario

Hiring is a subframe of Employer_scenario
Precedes

...captures the temporal ordering of subevents within a complex event. The relation holds between component subframes of a single complex frame, and provides additional information to the set of Subframe relations

Being_awake precedes Falling_asleep
Subframes and Precedes
FrameGrapher
Road Map

• Overview of FrameNet
  – Frames, Frame Elements, Lexical Units, Valence Descriptions, Frame-to-Frame Relations
  ✓ FrameNet and NLP

• Introduction to Multiword Expressions (MWEs)
  – Types of MWEs
  – Syntactic and Semantic Characteristics of MWEs
  – Representational Issues in MWEs

• Multiword Expressions in FrameNet
  – FrameNet’s treatment of (certain) MWEs
  – Navigating Lexicon and Grammar
  – Exploiting FrameNet Information on MWEs
NLP needs Frame Semantics

• Frames provide generalizations about lexical units at a useful level of abstraction, e.g. Operate vehicle covers drive.v, fly.v, paddle.v, sail.v, etc. useful for paraphrase

• Roles (Frame Elements) are also more meaningful than traditional semantic role labels, e.g. Driver in Operate vehicle for all the types of vehicle tells us more than just Agent.

• Frames represent conceptual gestalts--more than just the sum of their parts
NLP and FrameNet

• Automatic Semantic Role Labeling (ASRL)
  - Gildea and Jurafsky 2002
  - Chen et al. NAACL-HLT 2010. SEMAFOR
  - Hermann et al. ACL 2014. Automatic Frame Induction
  - Chang et al. LAW 2015. Controlled crowd-sourcing of annotation (work with Google)
Decisive Analytics Corporation

• Long-term collaboration with FrameNet via a series of subcontracts, e.g. current work on
  – Spatial relations
  – Negation, tense, mood and aspect

• Some of DAC’s products:
  – Network extraction
  – Attitude analysis
  – Semantic search
Decisive Analytics Corporation

• Network Extraction
  – use frame labeled data to produce entity network
  – filtering focuses analysis
  – relational modeling reorganizes network into meaningful clusters based on frame data

• Attitude Analysis
  – map FN to Attitudes semi-manually
  – exploit FN hierarchy of frames to prepopulate Holder/Target mapping for Frame Elements
  – generate FN-based queries from simple text

• Semantic Search
  – execute queries over frames, frame elements, and “terms”
  – results in several different forms
Road Map

• Overview of FrameNet
  – Frames, Frame Elements, Lexical Units, Valence Descriptions, Frame-to-Frame Relations
  – FrameNet and NLP

• Introduction to Multiword Expressions (MWEs)
  – Types of MWEs
  – Syntactic and Semantic Characteristics of MWEs
  – Representational Issues in MWEs

• Multiword Expressions in FrameNet
  – FrameNet’s treatment of (certain) MWEs
  – Navigating Lexicon and Grammar
  – Exploiting FrameNet Information on MWEs
Introduction

• Challenge of MWEs for NLP
• Defining MWE
• Distinguishing MWE from construction
Sag et al. 2002

• MWEs: A Pain in the Neck for NLP
  ▪ rough definition: “idiosyncratic interpretations that cross word boundaries (or spaces)”
  ▪ ubiquitous in language and across genres
  ▪ “words with spaces” treatment poses problems
    o flexibility
    o lexical proliferation

• Relevant Issues
  o Idiomaticity
  o Compositionality
  o Productivity
    o ...

EMNLP-Lisbon
Baldwin and Kim (2010): Idiomaticity of MWEs

• lexical: components not part of language
  – *ad hoc* (for this < Latin) for a specific purpose
  – *plus ça change* (more it changes < French)
  – *qué tal* (how are you < Spanish)

• syntactic: “non-compositional” syntax
  – *by and large* (prep conj adj) – adv.
  – *what’s up?* (Q-word-cop v. + adv.) – interjection (“Hi”)

18 September 2015
Baldwin and Kim (2010): Idiomaticity of MWEs

• semantic: varying degrees of compositionality
  – back and forth
  – taxi driver (NN compounds generally)
  – blow hot and cold
  – middle of the road

• pragmatic: tied to specific situation or context
  – good evening
  – lights out

• statistical: high frequency, relative to component words or alternative phrasings of same expression
  – immaculate performance vs. spotless performance
  – black and white vs. white and black
Baldwin and Kim (2010): Other Characteristics of MWEs

• crosslingual variation
  – Committee on Culture
    o Spanish: Comisión de la Cultura (...of the...)
    o French: Commission de la Culture (...of the...)
    o Italian: Commissione per la Cultura (...for the...)

• paraphrasable with one word
  – take advantage of \(\rightarrow\) exploit
  – blow the whistle on \(\rightarrow\) report

• proverbiality: describe/explain recurrent situation of social interest
  – piss off = annoy
  – drop off = fall asleep
  \begin{itemize}
  \item \text{informal}
  \end{itemize}

• prosody: related to semantic idiomaticity
  – sóft spot (vs. soft spót)
MWEs in NLP

• Workshops:
  – 11th Workshop on MWEs (2015 NAACAL/HLT)
  – 12th Workshop on MWEs (2016 ACL)

• Additional Publications:
  – *ACM Transactions on Speech and Language Processing (TSLP)* - Special issue on multiword expressions: From theory to practice and use, pt.1 V 10.2, June 2013
  – *ACM Transactions on Speech and Language Processing (TSLP)* - Special issue on multiword expressions: From theory to practice and use, pt.2 V.10.3, June 2013
Definition of MWE

• Fillmore & Ide (2002)
  – any expression made up of more than one lexical item which
does not fit a canonical syntactic pattern and/or which exhibits
some features of meaning, form, or distribution that cannot be
predicted from its component parts and its syntactic
organization.

• Baldwin & Kim (2010) following Sag et al. (2002)
  – Multiword expressions (MWEs) are lexical items that:
  (a) can be decomposed into multiple lexemes; and (b)
display lexical, syntactic, semantic, pragmatic and/or
statistical idiomaticity
Road Map

• Overview of FrameNet
  – Frames, Frame Elements, Lexical Units, Valence Descriptions, Frame-to-Frame Relations
  – FrameNet and NLP

• Introduction to Multiword Expressions (MWEs)
  ✓ Types of MWEs
    – Syntactic and Semantic Characteristics of MWEs
    – Representational Issues in MWEs

• Multiword Expressions in FrameNet
  – FrameNet’s treatment of (certain) MWEs
  – Navigating Lexicon and Grammar
  – Exploiting FrameNet Information on MWEs
Types of MWEs

Baldwin & Kim (2010)

– identify MWEs in formal terms (nominal, verbal, prepositional)

– classify MWEs based on their syntactic and semantic properties, distinguishing between *lexicalized* MWEs and *institutionalized* MWEs
Baldwin & Kim (2010): MWE Classification

- MWE
  - Lexicalised Phrase
    - fixed
    - semi-fixed
      - non-decomposable VNICs
      - nominal MWEs
  - syntactically-flexible
    - VPCs
    - LVCs
    - decomposable VNICs
  - Institutionalised Phrase
Baldwin & Kim: MWE Classification

• lexicalized: explicitly encoded in the lexicon
  – Fellbaum 20XX identifies those MWUs that must be included in the lexicon

• institutionalized: only statistically idiomatic
Baldwin & Kim: Lexicalized MWEs

• fixed MWEs: do not undergo morphosyntactic or internal modification
  – by and large (cf. *by and larger)
  – *ad hominem (*ad quamplurimos homines)
  – The Bronx (*Bronx, *A Bronx)
• semi-fixed MWEs: lexically-variable forms with hard restrictions on word order and composition, allowing variation in inflection, pronoun and determiner choice
  – shoot the breeze (shot the breeze, shooting the breeze)
  – The Rolling Stones (vs. A Rolling Stones’ concert)
  – find my/your/his/her place
  – NN compounds
• syntactically flexible
  – Verb-Particle: turn the blanket down/turn down the blanket
  – Light Verbs: make a decision, give a lecture, take revenge
  – decomposable VP idioms: kick the bucket, spill the beans
Taxonomy of MWEs
(Fillmore and Ide 2002)

• Grammatically Regular Idioms

• Idiomatic Syntactic Constructions

• Extragrammatical Idioms
# Grammatically Regular Idioms

<table>
<thead>
<tr>
<th>Type</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-sentence idiom</td>
<td><em>The fur is flying.</em></td>
</tr>
<tr>
<td>Full-sentence idiom with variable</td>
<td><em>Somebody up there likes me.</em></td>
</tr>
<tr>
<td>VP idiom</td>
<td><em>Somebody let the cat out of the bag.</em></td>
</tr>
<tr>
<td>Preposition selection</td>
<td><em>We object to your proposal.</em></td>
</tr>
<tr>
<td></td>
<td>I am quite <em>fond of</em> cats.</td>
</tr>
<tr>
<td></td>
<td>After the <em>attack on</em> the station</td>
</tr>
<tr>
<td></td>
<td>Get <em>out of</em> here</td>
</tr>
<tr>
<td>Particle selection</td>
<td><em>Let's cut out</em> early</td>
</tr>
<tr>
<td>Particle and preposition selection</td>
<td><em>Why put up with</em> that?</td>
</tr>
<tr>
<td>Support verb plus noun.</td>
<td><em>She took</em> little <em>advantage of</em> the opportunity.</td>
</tr>
<tr>
<td></td>
<td>Let's <em>pay</em> careful <em>attention to</em> their needs.</td>
</tr>
<tr>
<td>Pertinative adjective + Noun</td>
<td>military policy (cf. military demeanor)</td>
</tr>
<tr>
<td></td>
<td>educational practices (cf. educational experience)</td>
</tr>
<tr>
<td></td>
<td>economic board (cf. economical housewife)</td>
</tr>
</tbody>
</table>
Idiomatic Syntactic Constructions

- structure goes beyond the canonical, requiring appeal to special interpretation principles
- “peripheral” constructions with varying degrees of productivity and lexical restrictions
- parsable if grammar has details of constructions, requires recognition of patterns expressed in terms of grammatical categories and lexical sets, cannot depend on combinatorial requirements of lexical heads
- examples
  - *day in day out, year in year out*
    - CU-in-CU-out
  - my gem of a wife, her jerk of a husband
    - $N_1$ evaluates $N_2$ in $N_1$ of $N_2$ phrase
  - another five pages (*another many pages), a mere thirty dollars
    - singular determiner + quantified plural $N$
Extragramatical Constructions

• Exclusively identified and characterized by lexical form, don’t have canonical syntax

• Examples
    • He just wants to be *let alone*.
  – *First off*, Molly needs a place to live.
    • She always insists on being *first off* the plane.
Questions for Linguistics and NLP

Where is the dividing line?

Does identifying a line matter?

Does identifying a line matter for NLP?
Construction vs. MWE
June 2015 MWE Workshop

• Baldwin: Where is the dividing line between idiomatic constructions and MWEs?

• Michaelis: I don’t know.
Useful Heuristic?

- Highly abstract forms (e.g. Subject-Predicate) tend to be viewed as constructions.
- Forms with one or more fixed lexical items tend to be viewed as MWEs.
- Where is the dividing line?
- Does identifying a dividing line matter for NLP?
Road Map

• Overview of FrameNet
  – Frames, Frame Elements, Lexical Units, Valence Descriptions, Frame-to-Frame Relations
  – FrameNet and NLP

• Introduction to Multiword Expressions (MWEs)
  – Types of MWEs
    ✓ Syntactic and Semantic Characteristics of MWEs
  – Representational Issues in MWEs

• Multiword Expressions in FrameNet
  – FrameNet’s treatment of (certain) MWEs
  – Navigating Lexicon and Grammar
  – Exploiting FrameNet Information on MWEs
Syntactic Characteristics of MWEs

• range of syntactic configurations
  – nominal: surgeon general, airline employee complaint
  – verbal: take a shower, run the bath
  – adverbial: in short, first and foremost

• need not be well-formed
  ✓ kick the bucket, answer the door (cf. *answer a door)
  – by and large (cf. thick and thin, heart and soul, etc.)
  – on top (cf. on the top, *on bottom) on leave, in school, in court, to hospital
  – say when (*say whether), and then some (*and then any)
Syntactic Characteristics of MWEs

• may not allow modification
  – in medical school,
  – *in appellate court,
  – *to local hospital

• vary in degree of fixedness
  – spic and span (cf. *spic and very span), on air
  – kick the bucket (cf. *the bucket was kicked), fill one’s shoes
  – *turn in the work/*turn the work in, made a decision/a decision was made)
Semantic Characteristics

- reduced semantic transparency
- reduced or absent compositionality
- highly idiomatic
Road Map

• Overview of FrameNet
  – Frames, Frame Elements, Lexical Units, Valence Descriptions, Frame-to-Frame Relations
  – FrameNet and NLP

• Introduction to Multiword Expressions (MWEs)
  – Types of MWEs
  – Syntactic and Semantic Characteristics of MWEs
    ✓ Representational Issues in MWEs

• Multiword Expressions in FrameNet
  – FrameNet’s treatment of (certain) MWEs
  – Navigating Lexicon and Grammar
  – Exploiting FrameNet Information on MWEs
Representation of MWEs

- Sag et al. (2002) Lexical Representation
  - words with spaces: only works for fixed MWEs
  - other

  - \([\text{by}_{\text{prep.}} \text{ and}_{\text{conj.}} \text{large}_{\text{adj.}}]\)_{\text{adv}}

- Schneider (2014)
  - formal representation of shallow token groupings into “strong” MWEs (noncompositional expressions and proper names included) and “weak” collocations
Representational Issues: Creating Standards

• International Standard for Language Engineering
  – Calzolari, Lenci, and Zampolli (2001)
    • includes proposals for the representation of support verbs and noun-noun compounds cross-linguistically

• Cross-lingual Multi-word Expression Lexicons for Language Technology (XMELLT)
  – N. Ide (Vassar) 2000-2001 NSF Grant
  – Calzolari et al. 2002
Road Map

• Overview of FrameNet
  – Frames, Frame Elements, Lexical Units, Valence Descriptions, Frame-to-Frame Relations
  – FrameNet and NLP

• Introduction to Multiword Expressions (MWEs)
  – Types of MWEs
  – Syntactic and Semantic Characteristics of MWEs
  – Representational Issues in MWEs

• Multiword Expressions in FrameNet
  – FrameNet’s treatment of (certain) MWEs
  – Navigating Lexicon and Grammar
  – Exploiting FrameNet Information on MWEs
Revenge MWEs

• *get back (at)*
  – Tim got back at Peter for...
  – *Tim got back.

• *get even*
  – Tim got with Peter for...
  – Tim got even for...

• *pay back*

• *take revenge*

• *exact revenge*

N B. register difference
MWEs in FrameNet

Support Constructions: ~ 2750

Support Vs: make a decision; host a reception; launch an attack

Support Ps: under construction; with success; in doubt

As Lexical Units in FN database: ~830

Not in Count

Noun-Noun Compounds: wine bottle, armchair, etc.

Transparent Nouns: glass of milk, herd of cows, etc.
FrameNet’s treatment of MWEs

• Support Verbs
  – make decision
  – take revenge
  – give advice
  – turn blue
  – get happy

• Transparent Nouns
  – herd of sheep
  – box of toys
  – lock of hair

• Compound Nouns

discrepencies between syntactic and semantic head
Support Verbs

• syntactic object idiosyncratically selects the verb (not reverse)
  – make a decision
  – say a prayer
  – file a complaint

• may profile phase of complex event
  – make a promise
  – keep a promise

• lexical functions that present (different) subjects of transitive actions
  – give a test vs. take a test
  – perform surgery vs. undergo a surgery
FrameNet Treatment of Support Verb Constructions

• **NOUN** evokes the frame
  
  – hold a discussion Discussion
  
  – conduct research Research
  
  – make a deal **Make_agreement_on_action**

• Analyze Support Vs in terms of **evoked** frame
Discussion Frame

Core Frame Elements
- Interlocutor_1
- Interlocutor_2
- Interlocutors
- Topic

Non-Core Frame Elements
- Amount of Discussion
- Time
- Means

Last week the President \( \text{[held} \quad \text{supp}] \) [exhaustive DISCUSSIONS] with the Foreign Minister via Skype
FrameNet Treatment of Support Verb Constructions

- Adjective *evokes* the frame
  - get happy  \textit{Emotion\_directed}
  - turn blue  \textit{Color}

- Analyze Support Vs in terms of *evoked* frame

  Jasper’s face \textit{[turned}^{\text{Supp}} \textit{[ [a dark}^{\text{Color\_Descriptor}} \textit{] ] \textit{BLUE}^{\text{Color\_Color}}\text{]}} in the cold lake.
Types of Transparent Nouns

- **Aggregates**
  - bunch, group, collection, herd, school, flock
- **Quantities**
  - flood, number, scores, storm
- **Types**
  - breed, class, ilk, kind, type, sort
- **Portions and Parts**
  - half, segment, top, bottom, part
- **Unitizers**
  - glass, bottle, box, serving
- **Evaluations**
  - gem, idiot, prince
Transparent Nouns

• Aggregates
  – bunch of grapes, group of problems, flock of birds
• Quantities
  – flood of email, number of calls, scores of papers,
• Types
  – breed of dog, class of words, type of flower
• Portions and Parts
  – half an ounce, piece of paper, top of mountain
• Unitizers
  – glass of juice, bottle of perfume, serving of soup
• Evaluations
  – jerk of a husband, gem of a wife, dream of a house
FrameNet Treatment of Transparent Nouns

• Analyzes $[N_1 \text{ of } N_2]$ from the perspective of $N_1$
  $N_1 = \text{transparent } N \text{ and syntactic head}$
  determines integration of semantics
  $N_2 = \text{semantic head}$

$...[\text{a piece} \text{Part_Piece.Piece} \text{of cake} \text{Part_Piece.Whole}]]$
Transparency

• facilitates recognizing some types of discrepancies between syntactic and semantic structure in
  – support verb constructions
    • V + N
    • V governs N syntactically, but N is semantic head
  – N₁ of N₂ Construction
    • N₂ is semantic head
    • round of golf
Compound Nouns

• lexicalized compounds
  – picture frame, bookstore
    • w/o regard to typographical convention

• productive compounds

bookstore salesperson responsibility agreement
FrameNet Treatment of Compound Nouns

Head of the compound evokes the frame

$N_1N_2$ (where $N_2$ is Head)

[wine bottle] = Containers

The[wine Containers.Use] [BOTTLE Containers.Container] stood on the shelf.
Road Map

• Overview of FrameNet
  – Frames, Frame Elements, Lexical Units, Valence Descriptions, Frame-to-Frame Relations
  – FrameNet and NLP

• Introduction to Multiword Expressions (MWEs)
  – Types of MWEs
  – Syntactic and Semantic Characteristics of MWEs
  – Representational Issues in MWEs

• Multiword Expressions in FrameNet
  – FrameNet’s treatment of (certain) MWEs
  – Navigating Lexicon and Grammar
  – Exploiting FrameNet Information on MWEs
Traditional Distinction

• Lexicon: set of items associated with categories and denotations

• Grammar: set of rules about combining items in lexicon
Lexicon-Constructicon

• FrameNet Lexicon: repository of information about “words” in contemporary English based on the semantic frames, or common scenes and situations that the words describe.

• FrameNet Constructicon: repository of information about grammatical constructions in contemporary English that constitute the basic building blocks of the language.
Lexicon-Constructicon

Capturing meaningful units in language requires both lexicon and constructicon (Fillmore 2006), as does characterizing MWEs for identification and representation in natural language processing.
Road Map

• Overview of FrameNet
  – Frames, Frame Elements, Lexical Units, Valence Descriptions, Frame-to-Frame Relations
  – FrameNet and NLP

• Introduction to Multiword Expressions (MWEs)
  – Types of MWEs
  – Syntactic and Semantic Characteristics of MWEs
  – Representational Issues in MWEs

• Multiword Expressions in FrameNet
  – FrameNet’s treatment of (certain) MWEs
  – Navigating Lexicon and Grammar
  ✓ Exploiting FrameNet Information on MWEs
NLP Applications

• Information Retrieval
• Event Tracking
• Question-Answering
• FrameNet provides information about events and their participants, also for MWEs:
  – support verbs
  – transparent nouns
  – compound nouns
Example

Horatio took a bit of a dirt nap.

Support V:  

$take$ a dirt nap

Transparent N:  

a bit of a dirt nap

Compound N:  

dirt nap
Support Verb

take a dirt nap

take a nap = nap.v

cf. have a nap, get a nap

Analyzed in terms of Sleep frame, one of whose LUs is nap.n
Transparent Nouns

...a bit of a dirt nap

- \( N_1 \) of \( N_2 \), where \( N_2 \) identifies the whole of which \( N_1 \) is a part; \( N_2 \) = semantic head
- \( N_1 \) and \( N_2 \) also happen to be MWEs
  * He took bit of dirt nap
Transparent Nouns

\[ \text{a bit}_{\text{Part}} \text{ [of a dirt nap}_{\text{Whole}} \]}

Core Frame Elements

\textbf{Part}: identifies the part of the larger whole

\textbf{Whole}: identifies the undivided entity
Compound Nouns

*dirt nap*: $N_1 N_2$ where $N_2 =$ semantic head

non-compositional

dirt + nap = ????

non-productive

*sand nap

catnap, afternoon nap
Example

Horatio $[^{\text{took}}_{\text{Supp}}\{A\ \text{BIT} \ [\text{of a} \ DIRT\ NAP_{\text{Target}}\}]$}

[ ] Support Verb Construction

{} {} Transparent Noun

NN Compound
Example

1. [Horatio]_Dead_or_alive.Protagonist._ [took]_Supp_
   \{a bit of a DIRT NAP Target\}]

2. Horatio [took]_Supp_ \{A BIT TARGET [[of a dirt nap Hedging.Hedged_content]]\}
Conclusions

• FrameNet provides a wealth of information about the semantics of MWEs
• NLP would benefit from exploiting that information
• FrameNet plans major reconfiguration of data presentation

STAY TUNED!
Opportunity!

SemEval 2016 Task 10: Detecting Minimal Semantic Units and their Meanings (DiMSUM)

http://dimsum16.github.io/

In the open condition, systems may use any and all available resources.
Thanks!

Miriam R. L. Petruck
miriamp@icsi.berkeley.edu
References


References


References


References


References


References


References

