

# Robust Semantic Analysis of Multiword Expressions with FrameNet

Miriam R. L. Petruck 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 – spiciness
  - *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

[**Nora**<sub>Avenger</sub>] **retaliated** [**against her**  
**boss**<sub>Offender</sub>] [**for being dismissed**<sub>Injury</sub>]  
[**by leaving with the office keys**<sub>Punishment</sub>].

# Summarizing Results

- Automatic processes summarize the results, linking **FEs** 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.

## Frame Elements and Their Syntactic Realizations

The Frame Elements for this word sense are (with realizations):

Frame Element	Number Annotated	Realization(s)
Avenger	(39)	CNI.-- (2) NP.Ext (37)
Injured_Party	(1)	PP[on].Dep (1)
Injury	(38)	DNI.-- (35) PP[against].Dep (2) PP[for].Dep (1)
Instrument	(3)	PP[with].Dep (3)
Manner	(1)	AVP.Dep (1)
Offender	(39)	DNI.-- (36) PP[against].Dep (2) PP[on].Dep (1)
Place	(1)	PP[at].Dep (1)
Punishment	(39)	PP[in].Dep (2) PP[with].Dep (3) INI.-- (19) AVP.Dep (2) PPing[by].Dep (12) DNI.-- (1)
Time	(2)	AVP.Dep (2)

# 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



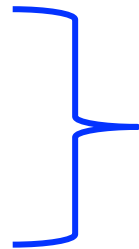
Number Annotated	Patterns				
<u>1</u> TOTAL	Avenger	Injured_Party	Offender	Punishment	
(1)	NP Ext	PP[on] Dep	DNI --	INI --	
<u>3</u> TOTAL	Avenger	Injury	Instrument	Offender	Punishment
(3)	NP Ext	DNI --	PP[with] Dep	DNI --	INI --
<u>1</u> TOTAL	Avenger	Injury	Manner	Offender	Punishment
(1)	NP Ext	DNI --	AVP Dep	DNI --	INI --
<u>1</u> TOTAL	Avenger	Injury	Offender	Place	Punishment
(1)	NP Ext	DNI --	DNI --	PP[at] Dep	PP[with] Dep
<u>31</u> TOTAL	Avenger	Injury	Offender	Punishment	
(1)	CNI --	DNI --	DNI --	PP[in] Dep	
(1)	CNI --	DNI --	DNI --	PP[with] Dep	
(2)	NP Ext	DNI --	DNI --	AVP Dep	
(9)	NP Ext	DNI --	DNI --	INI --	
(1)	NP Ext	DNI --	DNI --	PP[in] Dep	

# 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

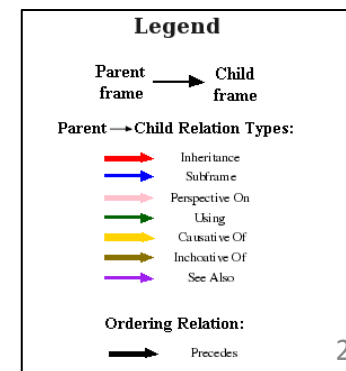
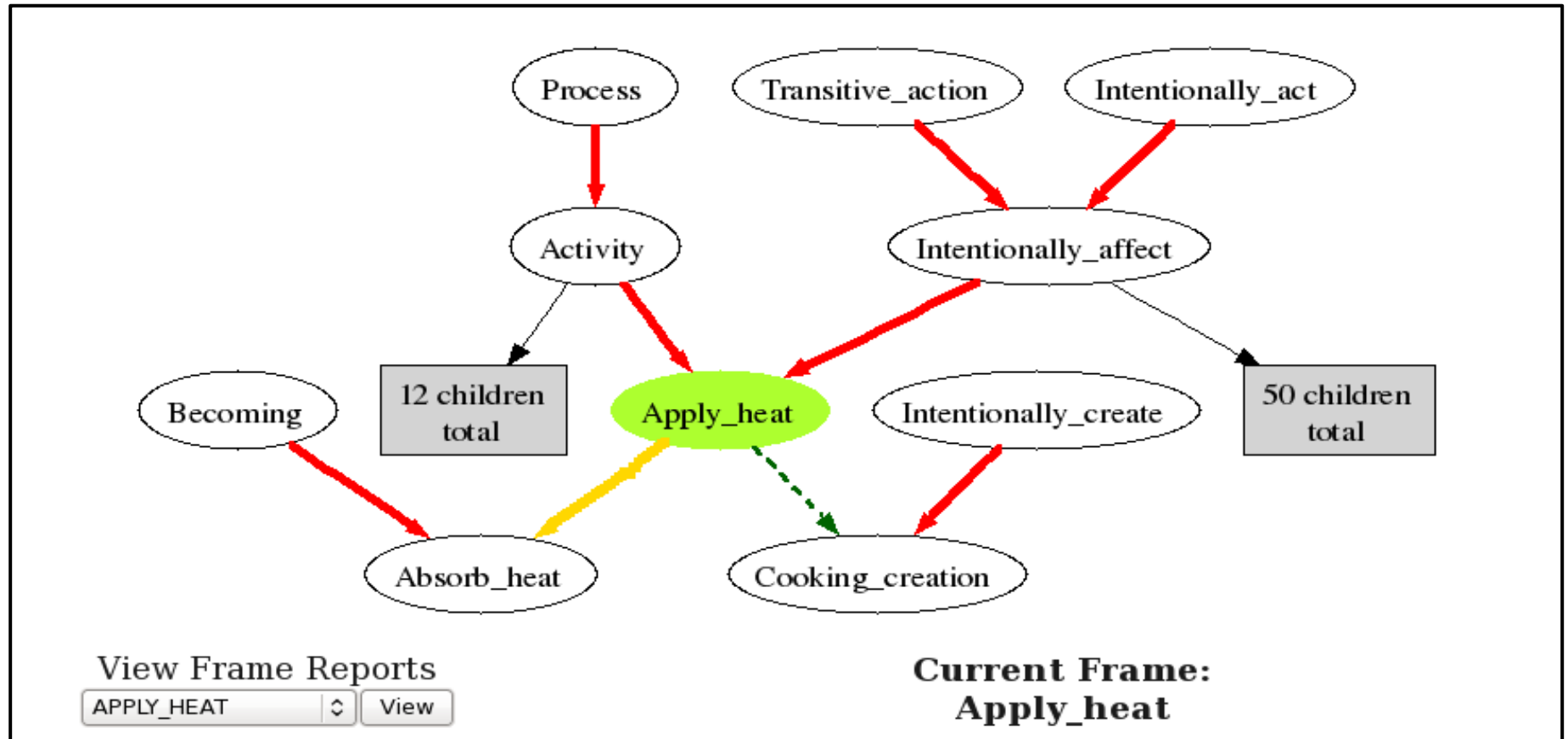


# Inheritance

- Relationship between a more general frame, the **parent** frame, and a more specific one, the **child**
- 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`

# FrameGrapher

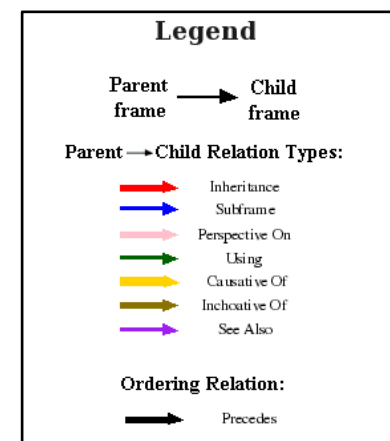
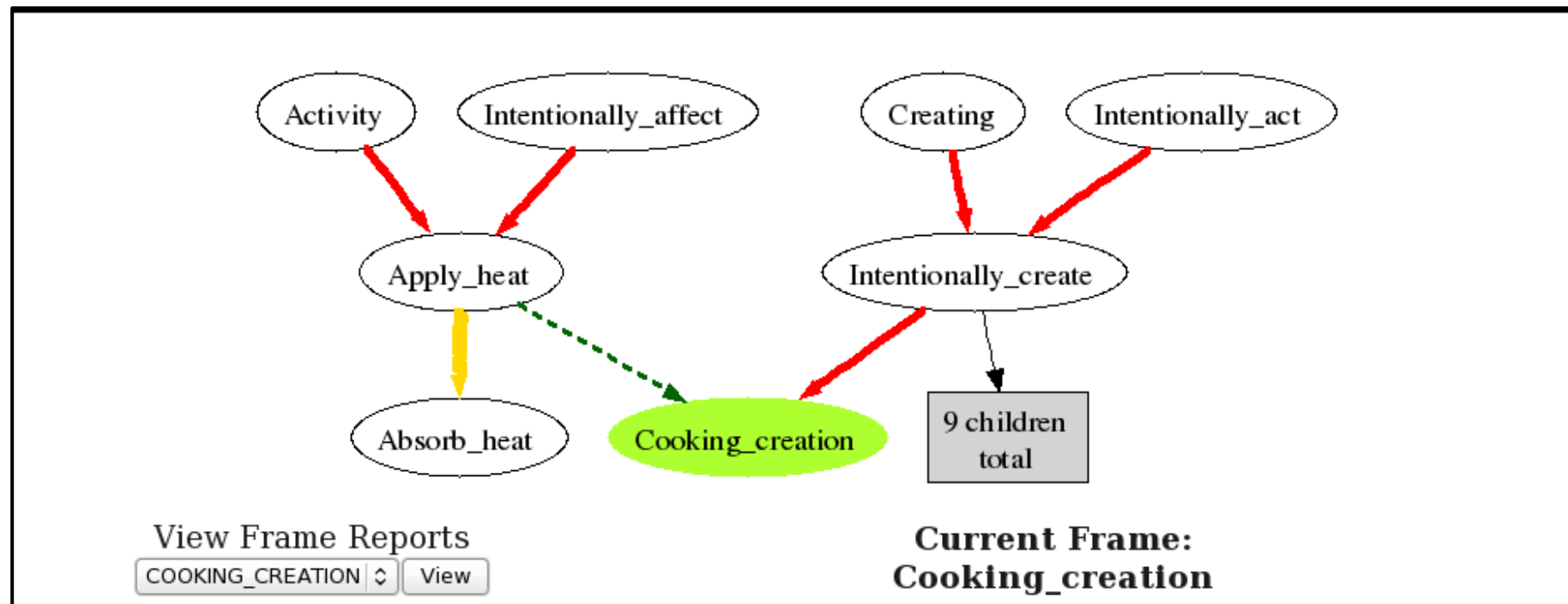


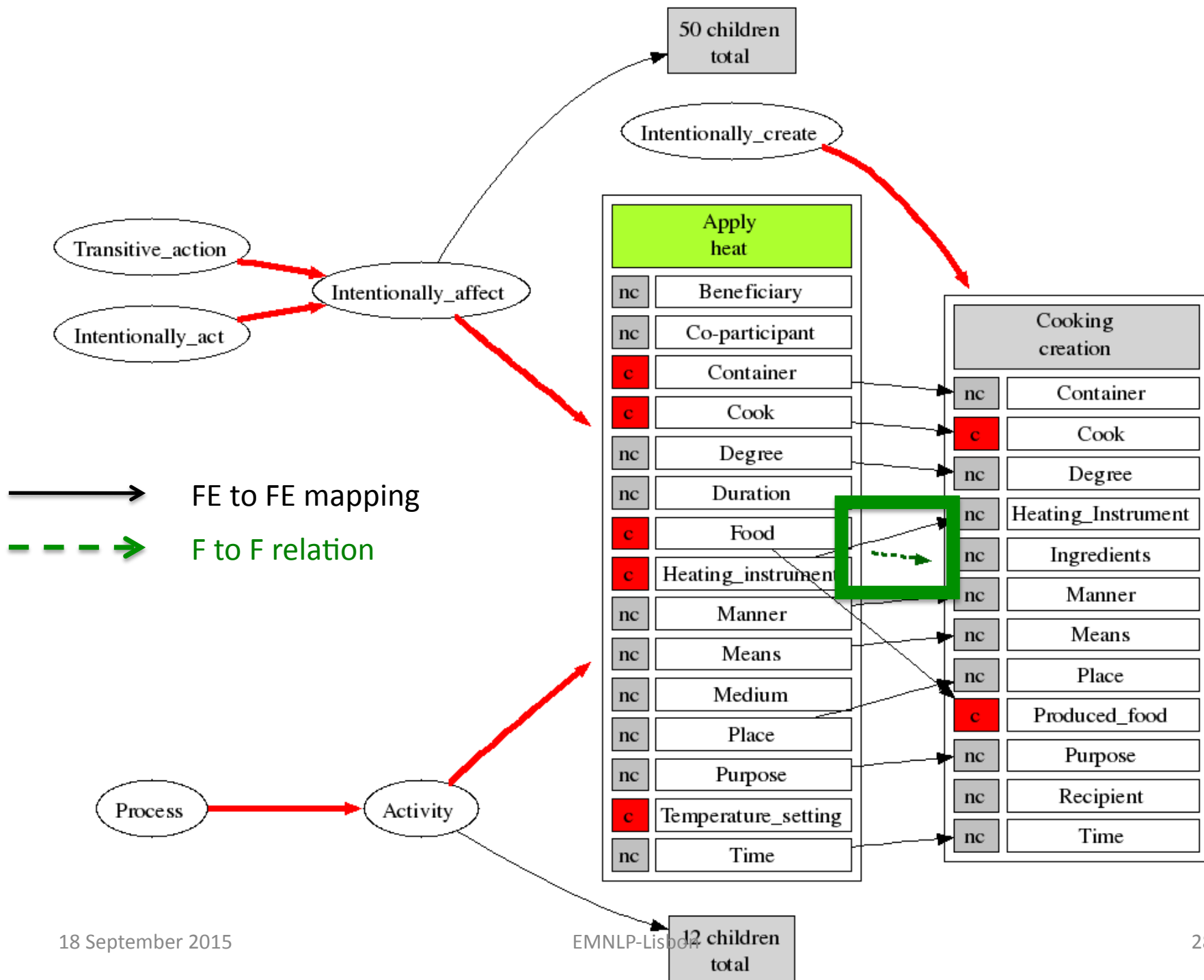
# Using (weak inheritance)

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

Cooking\_creation *uses* Apply\_heat

# FrameGrapher





# 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

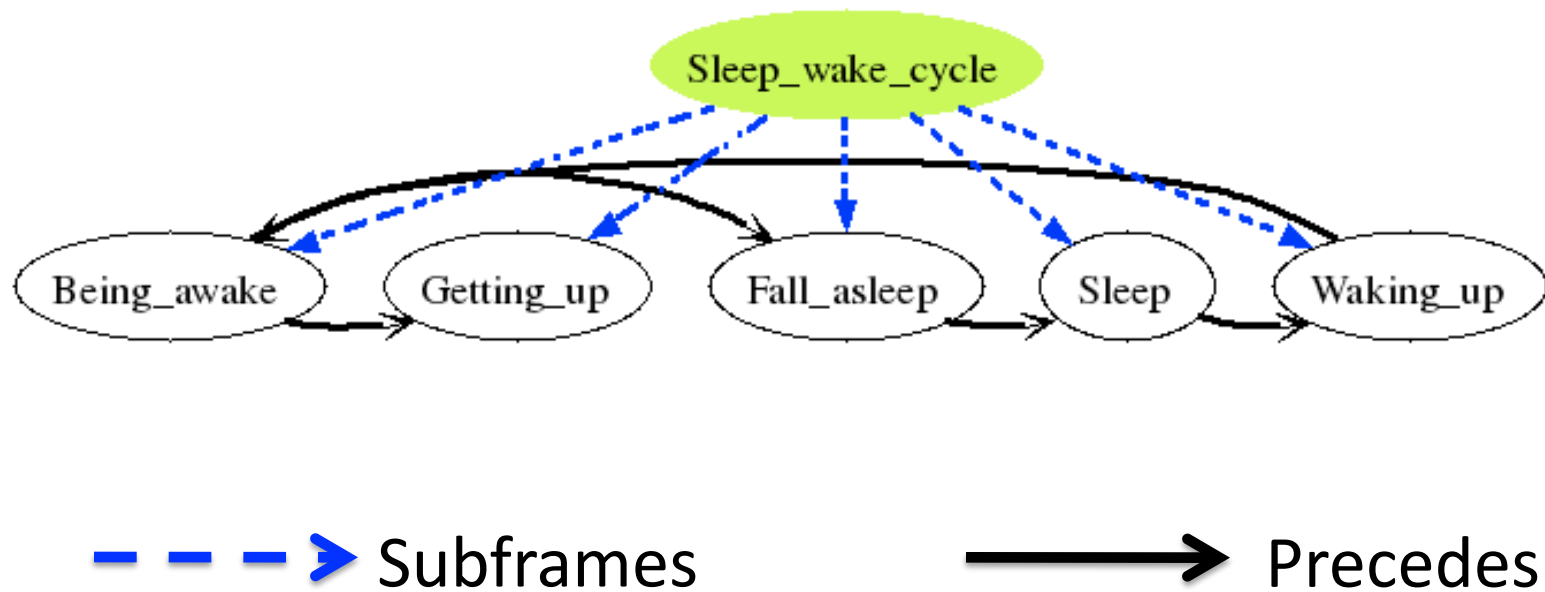
Hiring is a **subframe** of Employer\_scenario

# 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

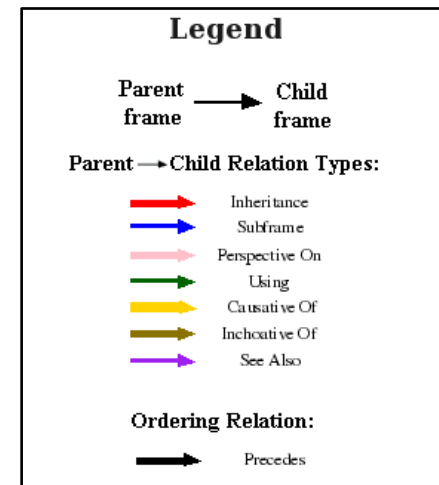
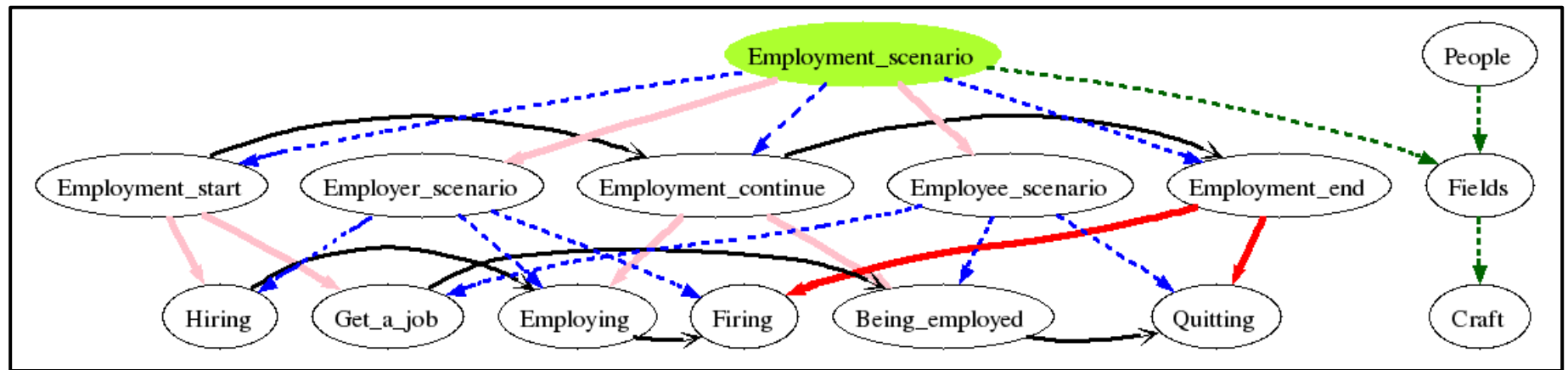
Being\_awake **precedes** Falling\_asleep

# 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
  - Das et al. 2010. Probabilistic Frame Semantic Parsing.
  - Chen et al. NAACL-HLT 2010. SEMAFOR
  - Das et al. 2014. *Computational Linguistics*, 40.1:9-56
  - 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 heirarchy 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
    - flexibility
    - lexical proliferation
  - Relevant Issues
    - Idiomaticity
    - Compositionality
    - Productivity

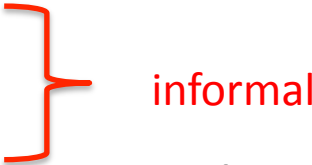
# Baldwin and Kim (2010): Idiomaticity of MWEs

- 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”)

# 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
    - Spanish: Comisión **de la** Cultura (...**of the**...)
    - French: Commission **de la** Culture (...**of the**...)
    - Italian: Commissione **per la** Cultura (...**for the**...)
- paraphrasable with one word
  - take advantage of → exploit
  - blow the whistle on → report
- proverbiality: describe/explain recurrent situation of social interest
  - piss off = annoy
  - drop off = fall asleep
- prosody: related to semantic idiomaticity
  - sóft spot (vs. soft spót)

# MWEs in NLP

- Workshops:
  - 11<sup>th</sup> Workshop on MWEs (2015 NAACAL/HLT)
  - 12<sup>th</sup> 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

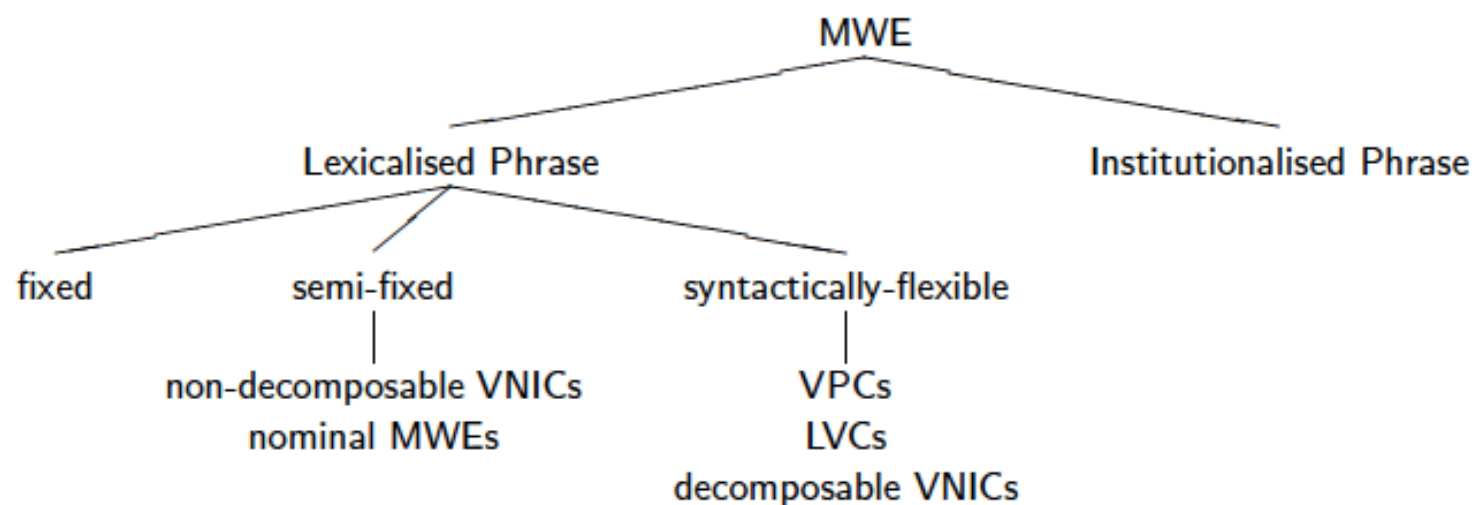
# Types of MWEs

## Baldwin & Kim (2010)

- identify MWEs in form-al terms (nominal, verbal, prepositional)
- classify MWEs based on their syntactic and semantic propoerties, distinguishing between *lexicalized* MWEs and *institutionalized* MWEs



# Baldwin & Kim (2010): MWE Classification



# 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

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

# 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

# Extragrammatical Constructions

- Exclusively identified and characterized by lexical form, don't have canonical syntax
- Examples
  - Ed doesn't eat fish, *let alone* sea urchin.
    - 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
- Heid (2008) Multi-layered annotation of MWE parts
  - $[[by_{\text{prep.}} \text{ and}_{\text{conj.}} \text{ large}_{\text{adj.}}]]_{\text{adv}}$
- Schneider (2014)
  - formal representation of shallow token groupings into “strong” MWEs (noncompositional expressions and proper names included) and “weak” collocations



# Representational Issues: Creating Standards

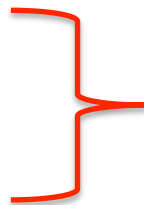
- International Standard for Language Engineering
  - Calzolari, Lenci, and Zampolli (2001)
    - includes proposals for the representation of support verbs and noun-noun compounds cross-linguistically
- Cross-lingual Multi-word Expression Lexicons for Language Technology (XMELLT)
  - N. Ide (Vassar) 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



discrepancies 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 [*held*<sub>Supp</sub>] [exhaustive **DISCUSSIONS**]  
with the Foreign Minister via Skype



# FrameNet Treatment of Support Verb Constructions

- Adjective **evokes** the frame
  - get happy                      Emotion\_directed
  - turn blue                      Color
- Analyze Support Vs in terms of **evoked** frame

Jasper's face [*turned*<sup>Supp</sup> [ *a dark*<sub>Color.Descriptor</sub> ] ]  
**BLUE**<sub>Color.Color</sub>] 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  $N_1$   
 $N_1$  = transparent N and syntactic head  
determines integration of semantics  
 $N_2$  = semantic head

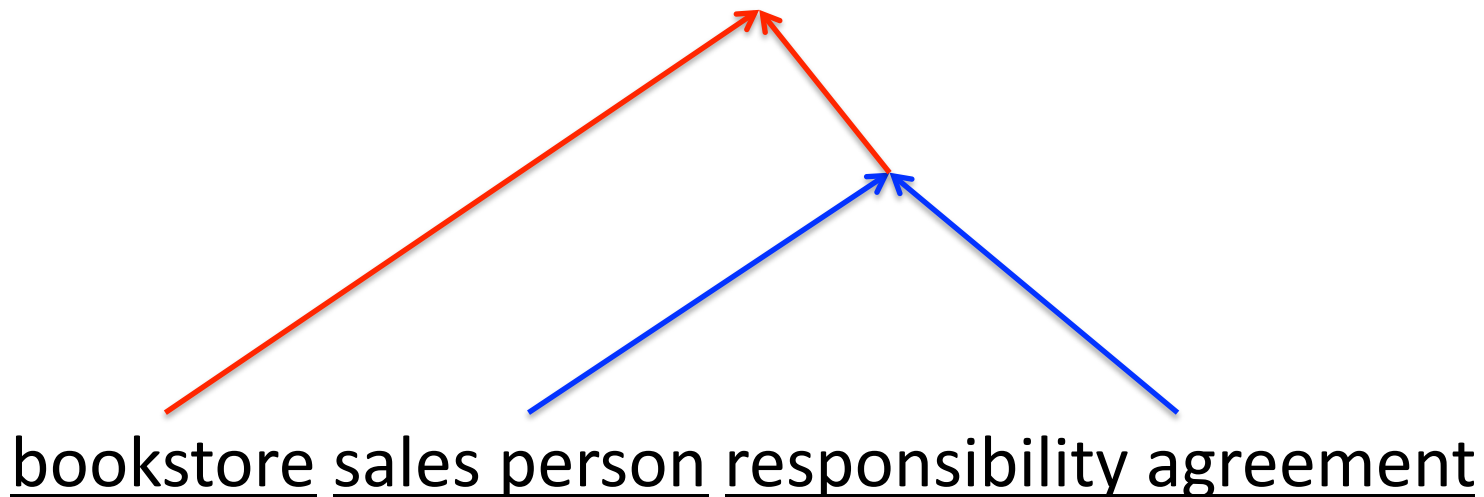
...[a piece<sub>Part\_Piece.Piece</sub> [of cake<sub>Part\_Piece.Whole</sub>]]

# 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
  - N<sub>1</sub> of N<sub>2</sub> Construction
    - N<sub>2</sub> is semantic head
    - round of golf

# Compound Nouns

- lexicalized compounds
  - picture frame, bookstore
    - w/o regard to typographical convention
- productive compounds



# 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<sub>Containers.Use</sub>] [BOTTLE<sub>Containers.Container</sub>]]  
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 the language.

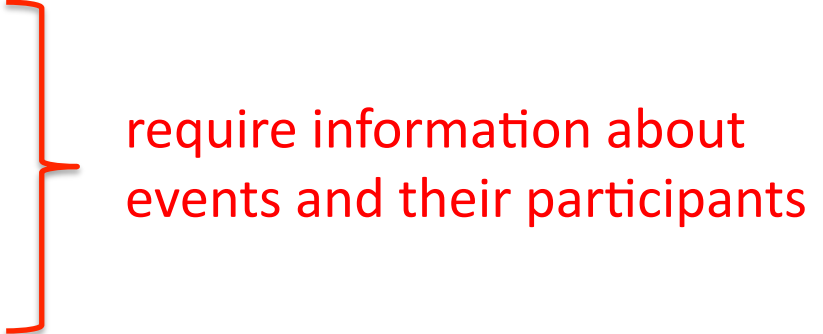
# Lexicon-Constructicon

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
- 
- require information about events and their participants

# 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

[a bit<sub>Part</sub>] [of a dirt nap<sub>Whole</sub>]

## Core Frame Elements

**Part**: identifies the part of the larger whole

**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 [*took*<sup>Supp</sup> { **A BIT** [of a **DIRT NAP**<sup>Target</sup> }]

[ ] Support Verb Construction

{ } Transparent Noun

NN Compound

# Example

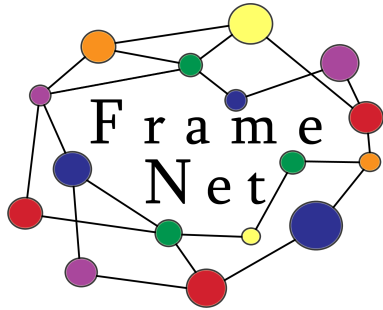
1. [Horatio<sub>Dead\_or\_alive</sub>.Protagonist.] [*took*<sup>Supp</sup>  
{a bit of a **DIRT NAP**<sub>Target</sub> }]

2. Horatio [*took*<sup>Supp</sup> {**A BIT**<sup>TARGET</sup> [[of a  
**dirt nap**<sub>Hedging</sub>.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!**



# Thanks!

Miriam R. L. Petruck  
[miriamp@icsi.berkeley.edu](mailto:miriamp@icsi.berkeley.edu)

# References

- T. Baldwin and Su Nam Kim. 2010. Multiword Expressions. In N. Indurkha and F. J. Damerau (eds.). *Handbook of Natural Language Processing*, 2<sup>nd</sup> Edition, London: Chapman & Hall/CRC, pp. 267-292.
- N. Calzolari, A. Lenci, and A. Zampolli. 2001. International Standards for Multilingual Resource Sharing: The ISLE Computational Lexicon Working Group. In *Proceedings of the ACL 2001 Workshop on Sharing Tools and Resources*, 15: 71-78.
- N. Calzolari, C. J. Fillmore, R. Grishman, N. Ide, A. Lenci, C. MacLeod, A. Zampolli. 2002. Towards Best Practice for Multiword Expressions in Computational Lexicons. In *Proceedings of the 3<sup>rd</sup> LREC*, Las Palmas, Spain, 1934-40.
- N. Chang, P. Paritosh, D. Huynh and C. Baker. 2015. Scaling Semantic Frame Annotation. In *Proceedings of LAW at NAACL-HLT*.

# References

- D. Chen, N. Schneider, D. Das and N. A. Smith 2010. SEMAFOR: Frame Argument Resolution with Log-linear Models. Proceedings of SemEval Workshop at ACL.
- D. Das, N. Schneider, D. Chen and N. A. Smith. 2010. Probabilistic Frame Semantic Parsing. Proceedings of NAACL-HLT.
- D. Das, D. Chen, A. F. T. Martins, N. Schneider and N. A. Smith. 2014. Frame- Semantic Parsing. *Computational Linguistics*, 40.1:9-56.
- D. Gildea and D. Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics* 28.3: 1-45.
- C. Fellbaum. In Press. The Treatment of Multi-word Units in Lexicography. *The Oxford Handbook of Lexicography*.



# References

- C.J. Fillmore. 1975. An alternative to checklist theories of meaning. In *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, pp. 123-131.
- C.J. Fillmore. 1985. Frames and the semantics of understanding. *Quaderni di Semantica*, 6.2: 222-254.
- C.J. Fillmore. 2012. Encounters with Language. *Computational Linguistics* 38.4: 701-718.
- C. J. Fillmore and B.T.S. Atkins. 1992. Towards a Frame-based organization of the lexicon: the semantics of RISK and its neighbors. In *Frames, Fields, and Contrasts: New Essays in Semantics and Lexical Organization*, ed. by Adrienne Lehrer and Eva Kittay, 75-102. Hillsdale: Lawrence Erlbaum.

# References

- C.J. Fillmore and C. F. Baker. FrameNet's Contribution to Text Understanding. 2005. Presentation at Summer Symposium on Advanced Question Answering for Intelligence (AQUAINT), Boston, Massachusetts.
- Fillmore, C. J. and C. Baker. 2010. A Frames Approach to Semantic Analysis. *The Oxford Handbook of Linguistic Analysis*. In Heine, B. and H. Narrog (eds.), *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, pp. 791-816.
- C.J. Fillmore, C.F. Baker, and H. Sato. 2002. Seeing Arguments Through Transparent Structures. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas, Spain, pp. 787-91.
- C.J. Fillmore. and N. Ide. 2002. Unpublished NSF Grant Proposal to identify and represent MWEs in multiple languages.

# References

- C.J. Fillmore, R. Lee-Goldman, and R. Rhomieux. 2012. The FrameNet Constructicon. In Sag, I. A., and Hans C. Boas, (eds.), *Sign-Based Construction Grammar*. Stanford: CSLI Publications, pp. 283-322.
- C.J. Fillmore, S. Narayanan, and C. Baker. 2006. What Linguistics can contribute to Event Extraction, Workshop on Event Extraction, AAAI Boston.
- C.J. Fillmore and H. Sato. Transparency and Building Lexical Dependency Graphs. Proceedings of the 28<sup>th</sup> Annual Meeting of the Berkeley Linguistics Society. 87-99.
- T. Fontenelle, T. (ed.) 2003. Special Issue on FrameNet and Frame Semantics. *International Journal of Lexicography* 16.3:231-385.
- U. Heid. 2008. Computational phraseology: An overview. In S. Granger and F. Meunier (eds.) *Phraseology: An interdisciplinary perspective*, Amsterdam: John Benjamins, pp. 337–360.

# References

- K. M. Hermann, D. Das, J. Weston and K. Ganchev. 2014. Semantic Frame Identification with Distributed Word Representations. Proceedings of NAACL.
- P. Kay and L. A. Michaelis. Constructional Meaning and Compositionality. 2012. In C. Maienborn, K. von Stechow and P. Portner (eds.), Semantics: An International Handbook of Natural Language Meaning. Berlin: de Gruyter, pp. 2271-2296.
- M. R. L. Petrucci. 1996. Frame Semantics. Handbook of Pragmatics, 8 pp.
- M. R. L. Petrucci and G. de Melo. 2012. Precedes: A Semantic relation in FrameNet. Proceedings of the Workshop on Language Resources for Public Security Applications, 8th LREC Conference, Istanbul, pp.45-49.

# References

- J. Ruppenhofer, M. Ellsworth, M. R. L. Petruck, C. R. Johnson, and J. Scheffczyk. 2010. FrameNet II: Extended Theory and Practice. Web Publication The BOOK.
- I.A. Sag , T. Baldwin , F. Bond, A. Copestake , and D. Flickinger. 2002. Multiword Expressions: A Pain in the Neck for NLP. In Proceedings of the 3<sup>rd</sup> International Conference on Intelligent Text Processing and Computational Linguistics. (CICLing 2002). Berlin: Springer, pp. 1-15.
- N. Schneider. 2014. Lexical Semantic Analysis in Natural Language Text. Dissertation, Carnegie Mellon University.