

Prolegomena to a Distributional Model of Sentence Comprehension

Alessandro Lenci

Università di Pisa
Dipartimento di Filologia Letteratura e Linguistica
and
Scuola Normale Superiore



COLING LAB

Computational Linguistics Laboratory

Productivity In Language Comprehension

Two facts about language comprehension

- i.) we have a potentially endless capacity to produce and understand **novel meaningful** sentences
 - ii.) *ceteris paribus*, **novel sentences** have a different cognitive status (i.e., they are processed differently) from **familiar sentences**
-
- (1)
 - a. The student reads the book in the library.
 - b. The surfer reads the papyrus in the forest.
 - c. * The stone reads the wave in the taxes.

Productivity In Language Comprehension

Two facts about language comprehension

- i.) we have a potentially endless capacity to produce and understand **novel meaningful** sentences
 - ii.) *ceteris paribus*, **novel sentences** have a different cognitive status (i.e., they are processed differently) from **familiar sentences**
-
- (1)
 - a. The student reads the book in the library.
 - b. The surfer reads the papyrus in the forest.
 - c. * The stone reads the wave in the taxes.

Productivity In Language Comprehension

Two facts about language comprehension

- i.) we have a potentially endless capacity to produce and understand **novel meaningful** sentences
 - ii.) *ceteris paribus*, **novel sentences** have a different cognitive status (i.e., they are processed differently) from **familiar sentences**
-
- (1)
 - a. The student reads the book in the library.
 - b. The surfer reads the papyrus in the forest.
 - c. * The stone reads the wave in the taxes.

Productivity In Language Comprehension

Two facts about language comprehension

- i.) we have a potentially endless capacity to produce and understand **novel meaningful** sentences
 - ii.) *ceteris paribus*, **novel sentences** have a different cognitive status (i.e., they are processed differently) from **familiar sentences**
-
- (1)
 - a. The student reads the book in the library.
 - b. The surfer reads the papyrus in the forest.
 - c. * The stone reads the wave in the taxes.

Productivity In Language Comprehension

Two facts about language comprehension

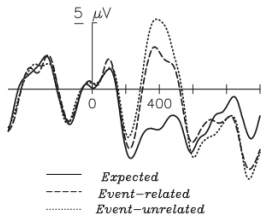
- i.) we have a potentially endless capacity to produce and understand **novel meaningful** sentences
 - ii.) *ceteris paribus*, **novel sentences** have a different cognitive status (i.e., they are processed differently) from **familiar sentences**
-
- (1)
 - a. The student reads the book in the library.
 - b. The surfer reads the papyrus in the forest.
 - c. * The stone reads the wave in the taxes.

Productivity In Language Comprehension

Two facts about language comprehension

- i.) we have a potentially endless capacity to produce and understand **novel meaningful** sentences
 - ii.) *ceteris paribus*, **novel sentences** have a different cognitive status (i.e., they are processed differently) from **familiar sentences**
-
- (1)
 - a. The student reads the book in the library.
 - b. The surfer reads the papyrus in the forest.
 - c. * The stone reads the wave in the taxes.

Sentence Comprehension in the Brain



Sentences including possible but unexpected (**novel**) combinations of lexical items, evoke stronger N400 components in the ERP waveform than sentences with expected (**non-novel**) combinations (Baggio and Hagoort 201)

- anomaly > unexpected > expected

What Does N400 Mean?

Pre-activation

The more the incoming word is congruent with the **expectations pre-activated by the context**, the smaller the amplitude of the N400 evoked by that word (Federmeier and Kutas 1999, Van Berkum et al. 2005)

Unification

N400 reflects the effort needed to combine the meaning of a word with the current contextual meaning

- **semantic unification** refers to the integration of word meaning into an unfolding representation of the preceding context, is a constructive process which results in a representation that is not already available in memory (Hagoort et al 2009, Baggio and Hagoort 2011)
- N400 is a processing consequence of the **cognitive cost to build compositional semantic representations** (Baggio, Van Lambalgen, Hagoort 2012)

What Does N400 Mean?

Pre-activation

The more the incoming word is congruent with the **expectations pre-activated by the context**, the smaller the amplitude of the N400 evoked by that word (Federmeier and Kutas 1999, Van Berkum et al. 2005)

Unification

N400 reflects the effort needed to combine the meaning of a word with the current contextual meaning

- **semantic unification** refers to the integration of word meaning into an unfolding representation of the preceding context, is a constructive process which results in a representation that is not already available in memory (Hagoort et al 2009, Baggio and Hagoort 2011)
- N400 is a processing consequence of the **cognitive cost to build compositional semantic representations** (Baggio, Van Lambalgen, Hagoort 2012)

A Balance between Storage and Composition

- Productivity entails that **not everything** can be stored in semantic memory (i.e. it is an argument for semantic compositionality)
- ERP data suggest that there is a large amount of **stored knowledge** in semantic memory about **event contingencies and concept combinations** (cf. also Culicover and Jackendoff 2005)
- This knowledge is activated by linguistic items during processing and affects language processing
- Combinations that are more **“distant” from the stored ones** (e.g., **novel combinations**) require more cognitive effort to be interpreted

A Balance between Storage and Composition

- Productivity entails that **not everything** can be stored in semantic memory (i.e. it is an argument for semantic compositionality)
- ERP data suggest that there is a large amount of **stored knowledge** in semantic memory about **event contingencies and concept combinations** (cf. also Culicover and Jackendoff 2005)
- This knowledge is activated by linguistic items during processing and affects language processing
- Combinations that are more **"distant" from the stored ones** (e.g., **novel combinations**) require more cognitive effort to be interpreted



A Balance between Storage and Composition

- Productivity entails that **not everything** can be stored in semantic memory (i.e. it is an argument for semantic compositionality)
- ERP data suggest that there is a large amount of **stored knowledge** in semantic memory about **event contingencies and concept combinations** (cf. also Culicover and Jackendoff 2005)
- This knowledge is activated by linguistic items during processing and affects language processing
- Combinations that are more **"distant" from the stored ones** (e.g., **novel combinations**) require more cognitive effort to be interpreted

A Balance between Storage and Composition

- Productivity entails that **not everything** can be stored in semantic memory (i.e. it is an argument for semantic compositionality)
- ERP data suggest that there is a large amount of **stored knowledge** in semantic memory about **event contingencies and concept combinations** (cf. also Culicover and Jackendoff 2005)
- This knowledge is activated by linguistic items during processing and affects language processing
- Combinations that are more **“distant” from the stored ones** (e.g., **novel combinations**) require more cognitive effort to be interpreted

The Goals of this Research

- i.) Explain how the **distributional semantic representation** of a sentence and discourse is built incrementally
- ii.) Associate with each distributional semantic representation a **composition cost**, to model the cognitive effort necessary to build it



The Goals of this Research

- i.) Explain how the **distributional semantic representation** of a sentence and discourse is built incrementally
- ii.) Associate with each distributional semantic representation a **composition cost**, to model the cognitive effort necessary to build it



Memory, Unification and Control (MUC) Model

Hagoort(2005, 2013, 2016)

Memory The Memory component refers to the linguistic knowledge that is stored in long-term memory

- Memory contains **unification-ready structures** corresponding to **constructions** (Goldberg 1995, 2006)
- constructions are represented by sets of **constraints** pertaining to the various levels of linguistic representation
- each constraint specifies how a given construction can combine with other constructions

Unification The Unification component refers to the assembly in working memory of pieces stored in memory into larger structures

- unification is a constraint-based process, which attempts at solving the constraints defining the constructions.
- unification operations take place in parallel at all the representation levels (cf. the Parallel Architecture in Jackendoff 1997, 2002).

Memory, Unification and Control (MUC) Model

Hagoort(2005, 2013, 2016)

Memory The Memory component refers to the linguistic knowledge that is stored in long-term memory

- Memory contains **unification-ready structures** corresponding to **constructions** (Goldberg 1995, 2006)
- constructions are represented by sets of **constraints** pertaining to the various levels of linguistic representation
- each constraint specifies how a given construction can combine with other constructions

Unification The Unification component refers to the assembly in working memory of pieces stored in memory into larger structures

- unification is a constraint-based process, which attempts at solving the constraints defining the constructions.
- unification operations take place in parallel at all the representation levels (cf. the Parallel Architecture in Jackendoff 1997, 2002).

Generalized Event Knowledge (GEK)

- Long-term semantic memory stores **generalized knowledge about events and their participants**
- GEK derives from **first-hand experience** and from **linguistic experience** (e.g., from linguistic descriptions of events)
- Language provides multiple cues that can be used to focus and activate various aspects of events (McRae and Matsuki 2009: 1419)

“the specific choice of verb can be used to bring to mind somewhat different scenarios, such as *eating* versus *dining*. In terms of the possible entities that participate in such events, knowing that a *waitress* is involved, for example, invokes a certain type of eating event. The phrase *hamburgers and hot dogs* produces a different type of scenario than does *turkey and stuffing*, including perhaps information about location and time of year. Instrument nouns can cue certain types of eating, as in *eating with a fork* versus *eating with a stick*. Finally, event nouns like *breakfast* or location nouns like *cafeteria* cue specific types of eating scenarios.”



Words as Cues to Events

Elman (2009, 2011, 2014)

- Words are **cues** to activate information stored in long-term memory

“But suppose one views words not as elements in a data structure that must be retrieved from memory, but rather as stimuli that alter mental states (which arise from processing prior words) in lawful ways. In this view, words are not mental objects that reside in a mental lexicon. They are operators on mental states. From this perspective, words do not *have* meaning; they are rather *cues* to meaning” (Elman 2014: 129)
- Words are **cues to GEK**

“Events play a major role in organizing our experience. Event knowledge is used to drive inference and access memory, and it affects the categories we construct. An event may be defined as a set of participants, activities, and outcomes that are bound together by causal interrelatedness” (Elman 2014: 129)



Words as Cues to Events

Elman (2009, 2011, 2014)

- Words are **cues** to activate information stored in long-term memory

“But suppose one views words not as elements in a data structure that must be retrieved from memory, but rather as stimuli that alter mental states (which arise from processing prior words) in lawful ways. In this view, words are not mental objects that reside in a mental lexicon. They are operators on mental states. From this perspective, words do not *have* meaning; they are rather *cues* to meaning” (Elman 2014: 129)
- Words are **cues to GEK**

“Events play a major role in organizing our experience. Event knowledge is used to drive inference and access memory, and it affects the categories we construct. An event may be defined as a set of participants, activities, and outcomes that are bound together by causal interrelatedness” (Elman 2014: 129)

GEK Activation by Words in Isolation

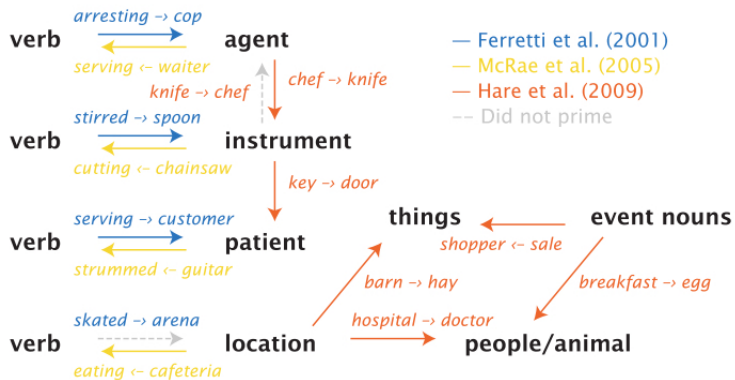


Figure: A summary of experiments that have investigated event-based priming.
 Source: McRae and Matsuki (2009)



Dynamic GEK Composition

- Abstract constructions may also activate GEK (cf. Johnson and Goldberg (2013) data about priming from jabberwocky sentences)
- Words can rapidly combine in sentences to cue specific concepts to narrow down the type of event that is being referred to
- When an argument is combined with a verb, it **modulates the set of events** to which the verb and the sentence refer



Dynamic GEK Composition

- Abstract constructions may also activate GEK (cf. Johnson and Goldberg (2013) data about priming from jabberwocky sentences)
- Words can rapidly combine in sentences to cue specific concepts to narrow down the type of event that is being referred to
- When an argument is combined with a verb, it **modulates the set of events** to which the verb and the sentence refer



Dynamic GEK Composition

- Abstract constructions may also activate GEK (cf. Johnson and Goldberg (2013) data about priming from jabberwocky sentences)
- Words can rapidly combine in sentences to cue specific concepts to narrow down the type of event that is being referred to
- When an argument is combined with a verb, it **modulates the set of events** to which the verb and the sentence refer



Sentence Comprehension as Event Representation

- Situation models (Zwaan and Radvansky 1998) propose that language is a set of instructions used to create a mental representation of an event or situation that is described in language
 - storage and retrieval of previously experienced situations plays a central role in this construction process (Johns and Jones 2015)
- The goal of the comprehender is to identify the event or situation the speakers wants to convey, and this is **the event that best explains the linguistic cues used in the sentence** (Kuperberg 2016)
 - stored information in long-term memory (GEK) is activated by linguistic cues and integrated (**unified**) into the available situation model, which is then dynamically updated



Sentence Comprehension as Event Representation

- Situation models (Zwaan and Radvansky 1998) propose that language is a set of instructions used to create a mental representation of an event or situation that is described in language
 - storage and retrieval of previously experienced situations plays a central role in this construction process (Johns and Jones 2015)
- The goal of the comprehender is to identify the event or situation the speakers wants to convey, and this is **the event that best explains the linguistic cues used in the sentence** (Kuperberg 2016)
 - stored information in long-term memory (GEK) is activated by linguistic cues and integrated (**unified**) into the available situation model, which is then dynamically updated

Recognize, Generalize and Adapt

Kleinschmidt and Jaeger (2015)

- **Cue-to-category** mapping: *How different cues are mapped on some underlying category, which is the used to make further inferences?*



- Three steps model to solve this problem:
 - a) recognize the familiar
 - b) generalize to the similar
 - c) adapt to the novel

Recognize, Generalize and Adapt

Kleinschmidt and Jaeger (2015)

- **Cue-to-category** mapping: *How different cues are mapped on some underlying category, which is the used to make further inferences?*



- Three steps model to solve this problem:
 - a) **recognize the familiar**
 - b) **generalize to the similar**
 - c) **adapt to the novel**

The Elements of GEK

- GEK is knowledge about events and situations that we have experienced under different modalities, **including linguistic input**
 - we use the term “event” in a very broad sense, encompassing both dynamic and static situations or eventualities (Dowty 1979, Rothstein 2004)
 - e.g., the information that student read books and that books have pages are both parts of GEK
- GEK is highly structured, and organized under various levels of complexity, granularity, and schematicity
 - information about **fully-specified micro-events** (e.g., students read books, surfers surf in the sea, etc.)
 - **schematic events** with entities that co-occur in the same situation, abstracting away from the specific events linking them (e.g, surfers, boards, waves, and wax tend to co-occur in the situations)
 - information about **complex scenarios**, much like scripts, frames or narrative schemas, which include various sub-events and complex temporal and causal relations about them (e.g., the surfing scenario includes events such as bringing a surf board, diving in the sea, swimming, etc.)

GEK and Distributional Semantics

- We represent events in GEK with **attribute-value matrices** (AVM) specifying their participants and roles
- We refer with GEK_{DS} to the subset of GEK that can be derived from co-occurrences in the linguistic input
- Events in GEK_{DS} contain information directly extracted from parsed sentences in corpora
 - attributes are **syntactic dependencies** (e.g. SUBJ, COMP-IN, etc.), as a surface approximation of deeper semantic roles
 - values are **distributional vectors** of dependent lexemes
 - “out-of-context” distributional vector encodings of lexical items

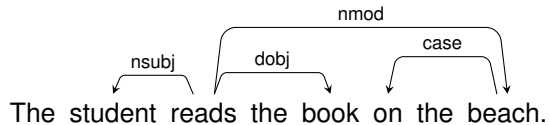
GEK and Distributional Semantics

- We represent events in GEK with **attribute-value matrices** (AVM) specifying their participants and roles
- We refer with **GEK_{DS}** to the subset of GEK that can be derived from co-occurrences in the linguistic input
- Events in GEK_{DS} contain information directly extracted from parsed sentences in corpora
 - attributes are **syntactic dependencies** (e.g. SUBJ, COMP-IN, etc.), as a surface approximation of deeper semantic roles
 - values are **distributional vectors** of dependent lexemes
 - “out-of-context” distributional vector encodings of lexical items

GEK and Distributional Semantics

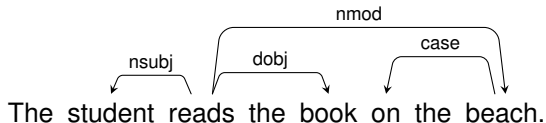
- We represent events in GEK with **attribute-value matrices** (AVM) specifying their participants and roles
- We refer with **GEK_{DS}** to the subset of GEK that can be derived from co-occurrences in the linguistic input
- Events in **GEK_{DS}** contain information directly extracted from parsed sentences in corpora
 - attributes are **syntactic dependencies** (e.g. SUBJ, COMP-IN, etc.), as a surface approximation of deeper semantic roles
 - values are **distributional vectors** of dependent lexemes
 - “out-of-context” distributional vector encodings of lexical items

GEK and Distributional Semantics



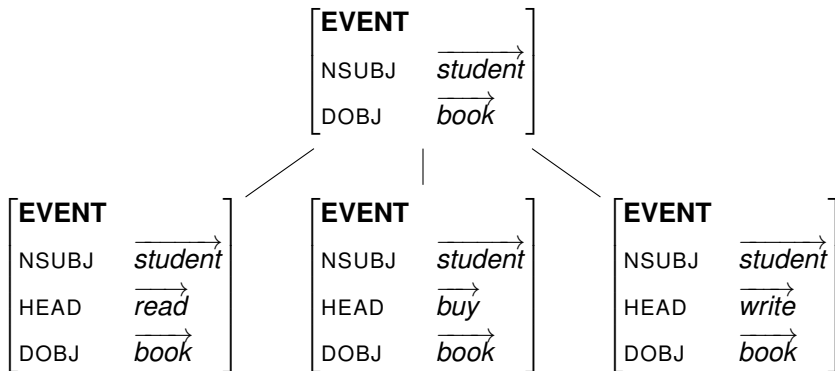
EVENT	
NSUBJ	$\xrightarrow{\text{student}}$
HEAD	$\xrightarrow{\text{read}}$
DOBJ	$\xrightarrow{\text{book}}$
NMOD-ON	$\xrightarrow{\text{beach}}$

GEK and Distributional Semantics



EVENT	
NSUBJ	$\xrightarrow{\quad}$ <i>student</i>
HEAD	$\xrightarrow{\quad}$ <i>read</i>
DOBJ	$\xrightarrow{\quad}$ <i>book</i>
NMOD-ON	$\xrightarrow{\quad}$ <i>beach</i>

Events in GEK are Hierarchically Structured



Lexical Items as Cues to GEK

- The lexicon is a repository of **constructions** stored in long-term memory
- Each construction Cxn is defined by a FORM and a content (SEM), represented with AVMs as in Sign-Based Construction Grammar (Sag 2012, Michaelis 2013)
- SEM is formed by two types of information:
 - a set of **events** stored in GEK and cued by the construction
 - a set of **semantic neighbors** (NEI) of the construction

$$\left[\begin{array}{l} \text{FORM} \quad \textit{student} \\ \text{SEM} \quad \left[\begin{array}{l} \text{GEK} \quad \langle e_1, \sigma_1 \rangle, \dots, \langle e_n, \sigma_n \rangle \\ \text{NEI} \quad \langle n_1, s_1 \rangle, \dots, \langle n_n, s_n \rangle \end{array} \right] \end{array} \right]$$

Lexical Items as Cues to GEK

- The lexicon is a repository of **constructions** stored in long-term memory
- Each construction C_{xn} is defined by a FORM and a content (SEM), represented with AVMs as in Sign-Based Construction Grammar (Sag 2012, Michaelis 2013)
- SEM is formed by two types of information:
 - a set of **events** stored in GEK and cued by the construction
 - a set of **semantic neighbors** (NEI) of the construction

$$\left[\begin{array}{l} \text{FORM} \\ \text{SEM} \end{array} \begin{array}{l} \textit{student} \\ \left[\begin{array}{l} \text{GEK} \\ \text{NEI} \end{array} \begin{array}{l} \langle e_1, \sigma_1 \rangle, \dots, \langle e_n, \sigma_n \rangle \\ \langle n_1, s_1 \rangle, \dots, \langle n_n, s_n \rangle \end{array} \right] \end{array} \right]$$

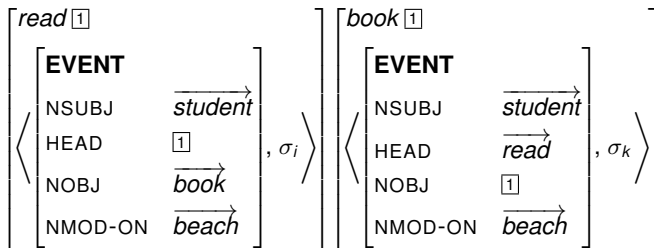
Lexical Items as Cues to GEK

- The lexicon is a repository of **constructions** stored in long-term memory
- Each construction C_{xn} is defined by a FORM and a content (SEM), represented with AVMs as in Sign-Based Construction Grammar (Sag 2012, Michaelis 2013)
- SEM is formed by two types of information:
 - a set of **events** stored in GEK and cued by the construction
 - a set of **semantic neighbors** (NEI) of the construction

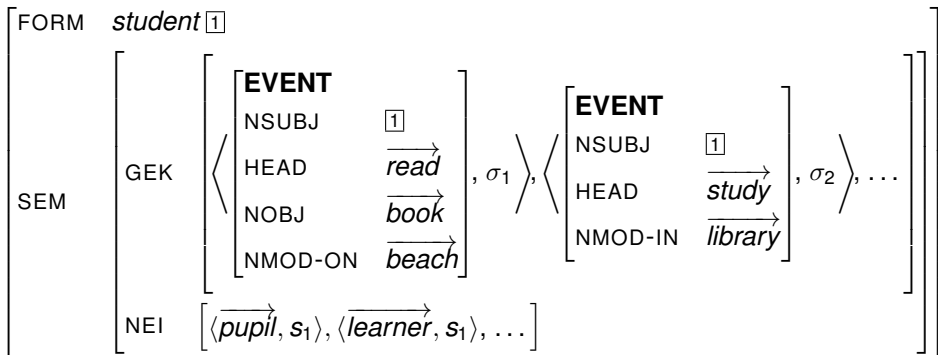
$$\left[\begin{array}{l} \text{FORM} \\ \text{SEM} \end{array} \begin{array}{l} \textit{student} \\ \left[\begin{array}{l} \text{GEK} \\ \text{NEI} \end{array} \begin{array}{l} \langle e_1, \sigma_1 \rangle, \dots, \langle e_n, \sigma_n \rangle \\ \langle n_1, s_1 \rangle, \dots, \langle n_n, s_n \rangle \end{array} \right] \end{array} \right]$$

Lexical Items as Cues to GEK

- The GEK component in SEM is a set of pairs $\langle e, \sigma \rangle$, such that:
 - e is an event stored in GEK_{DS}
 - σ is a score expressing the salience of the event with respect to the construction that cues it (e.g., $P(e|Cxn)$)
- Each event in GEK_{DS} may be cued by several lexical items, as part of their semantic content



Lexical Items as Cues to GEK





Sentences and Events

- Sentences are **partial descriptions of events**
 - several details of events are left unspecified by the sentences describing them
 - **implicit aspects** can be (probabilistically) **recovered** or **inferred** thanks to our GEK

- *John surfed yesterday*
 - John used a board
 - John was in the ocean or the sea
 - There were waves
 - There was wind
 - John wore a swimsuit or a wetsuit
 - ...



Sentences and Events

- Sentences are **partial descriptions of events**
 - several details of events are left unspecified by the sentences describing them
 - **implicit aspects** can be (probabilistically) **recovered** or **inferred** thanks to our GEK

- *John surfed yesterday*
 - John used a board
 - John was in the ocean or the sea
 - There were waves
 - There was wind
 - John wore a swimsuit or a wetsuit
 - ...



A Model for Sentence Comprehension

- The comprehension of a sentence is modeled as an incremental process of recovering (reconstructing) the mostly likely event it describes
- The construction of a sentence semantic representations proceeds incrementally with three-step compositional operations:
 - **retrieval** and activation of the GEK associated with lexical items in long-term memory
 - **integration** of the current GEK with the one of the preceding linguistic input
 - **adaptation** (accommodation) to integrate elements for which we can not retrieve any available GEK

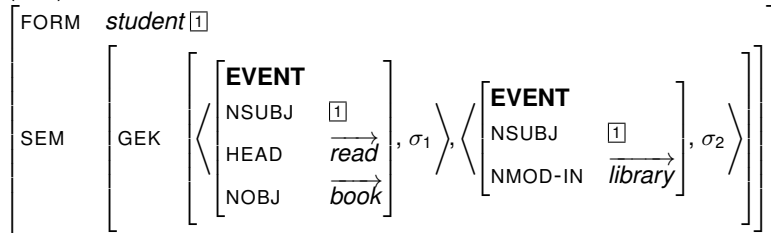


A Model for Sentence Comprehension

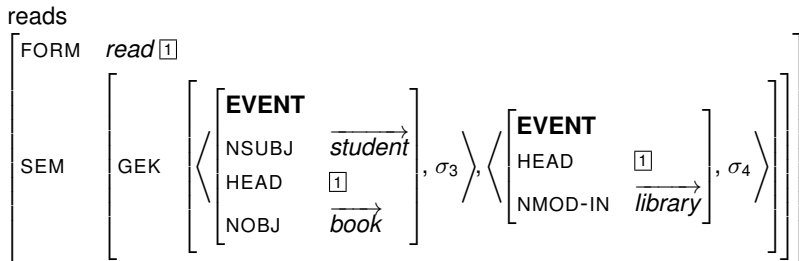
- The comprehension of a sentence is modeled as an incremental process of recovering (reconstructing) the mostly likely event it describes
- The construction of a sentence semantic representations proceeds incrementally with three-step compositional operations:
 - **retrieval** and activation of the GEK associated with lexical items in long-term memory
 - **integration** of the current GEK with the one of the preceding linguistic input
 - **adaptation** (accommodation) to integrate elements for which we can not retrieve any available GEK

Retrieval

(The) student



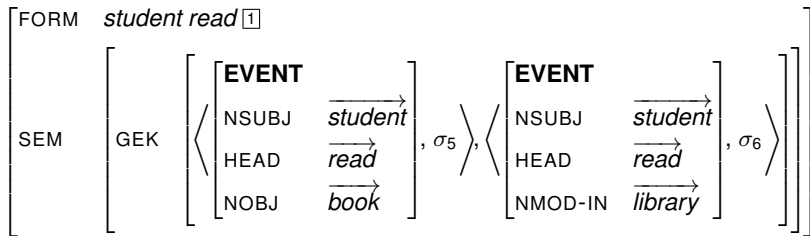
Retrieval



Integration

(The) student reads

↙ nsubj ↘



The Integration Function

- The integration function F is a compositional function that unifies the events AVMs of two lexical items and updates their scores:
 - $F(GEK_{w_1}, GEK_{w_2}) = GEK_{w_1, w_2}$
- F is actually formed by two functions F_e and F_σ :
 - 1 F_e **unifies** two event AVMs e_i and e_j , producing a new event AVM e_k :

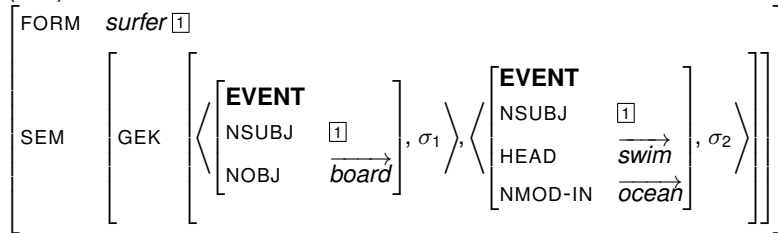
$$F_e(e_i, e_j) = e_k = e_i \sqcup e_j \quad (1)$$

- 2 F_σ **updates** the event weights of the successfully unified events, by combining the weights of e_i and e_j into a new weight assigned to e_k , e.g., by summation:

$$F_\sigma(\sigma_i, \sigma_j) = \sigma_k = \sigma_i + \sigma_j \quad (2)$$

Adaptation

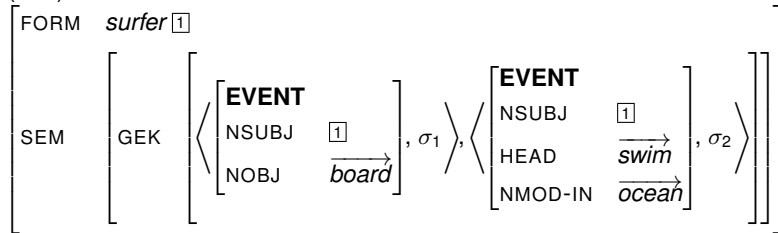
(The) surfer



- Assume that in GEK_{DS} there is **no event** of reading by a surfer
- How can we build the semantic representation of *The surfer reads*, given that we have never encountered such an event before?

Adaptation

(The) surfer



- Assume that in GEK_{DS} there is **no event** of reading by a surfer
- How can we build the semantic representation of *The surfer reads*, given that we have never encountered such an event before?

Adaptation

- We assume that the event expressed by *The surfer reads* will be **similar** to other events of reading we have experienced and that can be retrieved from its GEK_{DS}
- We retrieve and unify with *surfer* the events cued by *read* whose subject is **most similar** to *surfer* (e.g. *swimmer*, etc.)
- Language productivity can be conceived as the capacity to adapt our GEK stored in semantic memory to novel situations: **Productivity is adaptation, and adaptation is by similarity**



Adaptation

- We assume that the event expressed by *The surfer reads* will be **similar** to other events of reading we have experienced and that can be retrieved from its GEK_{DS}
- We retrieve and unify $\xrightarrow{\quad}$ with surfer the events cued by *read* whose subject is **most similar** to surfer (e.g. *swimmer*, etc.)
- Language productivity can be conceived as the capacity to adapt our GEK stored in semantic memory to novel situations: **Productivity is adaptation, and adaptation is by similarity**

Adaptation

- We assume that the event expressed by *The surfer reads* will be **similar** to other events of reading we have experienced and that can be retrieved from its GEK_{DS}
- We retrieve and unify with $\overrightarrow{\text{surfer}}$ the events cued by *read* whose subject is **most similar** to $\overrightarrow{\text{surfer}}$ (e.g. *swimmer*, etc.)
- Language productivity can be conceived as the capacity to adapt our GEK stored in semantic memory to novel situations: **Productivity is adaptation**, and **adaptation is by similarity**

The Cost of Novelty

- The process of adaptation has a cost which can be proportionally to the distance between *surfer* and the typical subjects of the events of reading
 - the higher the similarity between *surfer* and the typical readers I have encountered, the smaller the adaptation cost
- **Thematic fit** is regarded to measure the congruence of a predicate with an argument
 - more typical arguments (e.g., *book*) of a predicate (e.g., *read*) have a higher thematic fit than atypical, less expected arguments (e.g., *papyrus*)
- We can also see thematic fit as the cost of the cognitive operation to **generalize and/or adapt** our stored GEK to new linguistic combinations that we have never encountered before, to build a coherent semantic representation

The Cost of Novelty

- The process of adaptation has a cost which can be proportionally to the distance between *surfer* and the typical subjects of the events of reading
 - the higher the similarity between *surfer* and the typical readers I have encountered, the smaller the adaptation cost
- **Thematic fit** is regarded to measure the congruence of a predicate with an argument
 - more typical arguments (e.g., *book*) of a predicate (e.g., *read*) have a higher thematic fit than atypical, less expected arguments (e.g., *papyrus*)
- We can also see thematic fit as the cost of the cognitive operation to **generalize and/or adapt** our stored GEK to new linguistic combinations that we have never encountered before, to build a coherent semantic representation

Ongoing Work and Near Future Plans

- Extracting GEK from parsed ukWaC, Wikipedia and BNC
- Formalizing the process of event unification for sentence composition
- Integrating the model in a **Dynamic Semantics** framework (cf. see next week workshop on *Referential Semantics One Step Further*)
- Defining the composition cost function, in terms of the cost needed to retrieve, unify, and adapt GEK
- Extending the unification model beyond sentences
- Modeling several psycholinguistic datasets for sentence comprehension (Bicknell et al. 2010, Metusalem et al. 2012, Warren et al. 2015)
- Preparing a new dataset to test the reconstruction and inference of implicit event information as an effect of sentence comprehension

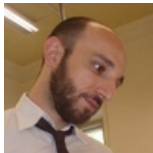
CoLing Lab

<http://colinglab.humnet.unipi.it/>

- This research is conducted in collaboration with:
 - **Emmanuele Chersoni**



- **Gianluca Lebani**



Grazie!!!
Thank You!!!
Danke!!!