The annotations

- In this version I include some annotations to cover both things that I said during the talk and things that I realized later weren’t very clear.
- Why did I call this approach "extremist"? I realize now that it’s not as extremist as Ray Jackendoff’s position, in which even memorized poems, familiar phrases from TV commercials, and the like, all had to get included in the lexical knowledge of a speaker. In my case I wanted to include a wide range of grammatical constructions as well as listable word configurations.

What is a MWE?

Any linguistic expression involving more than one word that requires an interpreter – human or machine – to have more than the abilities of an "Innocent Speaker-Hearer".

Innocent Speaker-Hearer

The ISH knows
- individual simple lexical units,
- the basic head-to-dependent grammatical relations,
- the basic head-to-dependent semantic relations as determined by the semantic frame of the governing lexical unit,
- regular and specific rules for realizing these,
- strategies for building a semantic structure out of all this.

That’s all it knows.
ISH's knowledge is about
– unitary words and
– word-to-word relations.
That can be represented in dependency diagrams in
– which each node is a word and
– each word-to-word link, i.e., each branch,
  • stands for one of the basic grammatical
    relations and
  • is capable of bearing a frame-based semantic
    relation to the governor.

The graphs
• There are other, perhaps more standard, ways of showing
dependency graphs. A common one is to show the
dependency relations with simple points and then to show
the linearization with lines pointing down to a text.

His parents gave me a copy of that fascinating book about frogs.

Basic syntactic relations
• Complementation
  – some governors require particular kinds of
    dependents
• Specification
• Modification

(there are others)
The blue lines

- Each blue line in the graphs which follow identify the kind of dependency relation named in the slide's heading.

Complementation

His parents gave me a copy of that fascinating book about frogs.

Actually, *copy of* should be treated as a MWE.

Specification

His parents gave me a copy of that fascinating book about frogs.

Actually *his* can also be thought of as satisfying a frame requirement of the relational noun *parents*.

Modification

His parents gave me a copy of that fascinating book about frogs.
So ...

The study of MWEs can begin with an examination of meaning units of the language that do not lend themselves to such a simple treatment.

Where the ISH idealization fails

1. Some apparent MWEs are best analyzed as single words, occupying one node.
2. Some MWEs can be represented as dependency subgraphs (not "just" word strings, or collocate sets).
3. Some MWEs are the product of "non-core" semi-independent mini-grammars (e.g., personal names, number words or the phrases we use for telling clock time).

Some motivation

• The highlighted passage in the next slide - a repetition of the preceding slide - is in reference to work on multiword entities that has not been sensitive to grammatical relations. Early on the search for simple contiguous strings of words was seen as inadequate (many MWEs can be interrupted) and attention was given rather to collocations, words that co-occur in statistically significant ways. In such work the emphasis was on choosing a "span" within which significant collocates could be reliably found, in general recognizing that they at least have to be in the same sentence. This paper emphasizes the specific kinds of grammatical relations that collocates have to have with each other to be linguistically significant.
Theoretical Issues

- Headedness
- Relation to the rest of the grammar
- How many of them are there? (Jackendoff; Mel’cuk)
- How do we count them?
- Language acquisition
- Language typology
- Role in parole
- How are they represented?

1. “Runs”

There are things that look like MWEs (that are written as sequences of words), but they have no internal variation and may just as well be thought of as long words with spaces in them.

Examples
- used to, let alone, of course, all of a sudden, first off
- Some are easily mislearned:
  - by and large > by in large
  - to all intents and purposes > to all intensive purposes
  - an arm and a leg > a nominal egg

"Runs"

- By "runs" is meant strings of words that have no interruptable substructure - they are just run through from beginning to end. Nobody would interrupt a phrase like "of course" ("of y’know course") or "all of a sudden" ("all of, well, a sudden") or, in general, any of the other expressions listed here.
- The fact that many complex runs get reanalyzed means that they are not learned by building up structures out of their parts, but as wholes; but, that said, they nevertheless have to be given some kind of analysis, if only because English simply doesn’t have many really long words, and also because they need some kind of syntactic structuring for us to know how to apply stress rules to them.

But,

I won’t talk about certain other kinds of semantically penetrable "long words” that identify very specific entities or concepts; e.g.,
- names of organizations
  The American Society for the Prevention of Cruelty to Animals. (ASPCA)
- names of officially designated crimes
  assaulting a federal officer with a lethal weapon
Some MWEs can be analyzed in terms of sets of nodes connected to each other through syntactic head-dependent relations.

- We can think of the red parts of these diagrams as the lexical units, and the black part as the subcategorization requirements.

The linearization of the elements of this kind of MWE may be constrained, or it may be subject to ordinary syntactic principles.
2. Dependency Subgraphs

Some MWEs can be analyzed in terms of sets of nodes bearing direct or indirect syntactic head-dependent relations to each other.

The linearization of the elements of this kind of MWE may be constrained, or it may be subject to ordinary syntactic principles.

There is no need to see the elements of such a subgraph as making up a "single constituent" in a corresponding phrase-structure representation.

Particle Verbs - Intransitive

- **Verb > particle** is the lexical unit.
- Exx: wake up, go away, sit down, shut up,
- Interruptible: Shut the hell up!

Particle Verbs - Transitive

- **Verb > particle** is the lexical unit.
- Exx: take off ('remove'), take out ('date'),
- Interruptible: Take your shoes off. I took her out once.
In the Old Days ...

About half a century ago it was generally believed that in Deep Structure, phrases like pick up, take off, etc., started out as single constituents, and a Particle Movement Transformation allowed the extraction of the particle so that it could follow the direct object.

\[ \text{[take off] [your shoes]} \rightarrow \text{[take] [your shoes] [off]} \]

A dependency subgraph can recognize the unity of the two-word block without worrying about phrasal constituency.

Prepositional Verbs - Intransitive

Verb > preposition is the lexical unit.
Exx:
- look for ('seek'), object to ('oppose'), look into ('investigate')

Interruptible:
- I looked long and hard for the perfect wife. We objected strenuously to her proposal.

Comment:
- Some PPs are omissible, some aren't. look (for), look into

Omissibility

- It may seem paradoxical that the real form of the lexical unit is something like look for but that the for-phrase can be omitted. This can be more straightforwardly represented in the Construction Grammar framework - but that's another story.
Prepositional Verbs - Transitive

Verb > preposition is the lexical unit.

Exx:
talk into (‘persuade’), rid of
Comment: PP is sometimes omissible:
The judge cleared me (of all charges).
They tried to talk me *(into quitting my job).
Who will rid me *(of this meddlesome priest)?

Particle-&-Preposition Verbs

Verb > {part, prep} is the lexical unit.

Exx:
put up with (‘tolerate’), look up to (‘respect’), break in on (‘interrupt’)
Not generally interruptible, I think (haven’t checked corpus data).
(But: This is the sort of errant nonsense up with which I shall not put.
Winston Churchill)

The Churchill quote

• That's supposed to be what Churchill once said in response to an editorial suggestion that he rewrite a sentence that ended with a preposition. It doesn't really stand as an exception to the generalization that V+particle+preposition lexical units aren't interruptible.

V+N+P Verbs

Verb > /N. prep/ is the lexical unit.

Exx:
take advantage of (‘exploit’), take part in (‘participate in’), take charge of
Comments:
N can be modified; N can be passive subject:
Considerable advantage was taken of this opportunity.
Pseudo-passive:
They were cruelly taken advantage of.
N does not take a determiner.
V+N+P Verbs

Verb \(/N, prep/\) is the lexical unit.

Exx:
- take advantage of (‘exploit’),
- take part in (‘participate in’),
- take charge of

Comments:
- N can be modified; N can be passive subject.
- Considerable advantage was taken of this opportunity.
- Pseudo-passive:
  They were cruelly taken advantage of.
- N does not take a determiner.

These are getting close to support verb constructions.

Support Constructions

Other Parts of Speech

Adjectives can have prepositional and clausal complements:
- fond of cats; interested in math; similar to mud

Nouns can have prepositional and causal complements:
- top of the tower; friend to the poor; journey into the jungle; copy of the book

The red boxes

- The red boxes in the next set of slides are supposed to indicate that these are the words that evoke the semantic frame expressed by the entire multiword. The support verb, in other words, just comes along as a way of using the noun as a verb.
Support Verbs with Subject N

- **Verb > N** is the lexical unit, N is semantic head, V is support verb.
- **Exx:**
  - The wind is blowing, the fire is burning, the rain is falling, a riot occurred; an accident happened.
  - Comment: The frame is evoked by the noun. The support verb is selected by the noun.
  - Compare "the fire is burning" with "the house is burning".
- **Note linearization:** Since these are intransitive, the N is (or heads) the subject NP and the verb is the predicate.

Support Verbs with Object N

- **Verb > N** is the lexical unit, N is semantic head, V is support verb. N has its own valence.
- **Exx:**
  - We had an argument with the kids. ("we argued with the kids")
  - I made the decision to leave. ("I decided to leave")
  - Comment: The frame is evoked by the noun. The SV is selected by the noun, which also brings in its own complement structure.
- **Comment:** The N doesn’t have to be deverbal: wage war, commit a crime.

Ditransitive Support Verbs

- **Verb > N** is the lexical unit, N is semantic head, V is support verb. X and Y are each participants in N’s frame.
- **Exx:**
  - She gave me a kiss. ("she kissed me")
  - I paid him a bribe. ("I bribed him")
  - They gave me good advice. ("they advised me well")
- **Comment:** The N doesn’t have to be deverbal: wage war, commit a crime.
SVs can resolve polysemy.

Polysemous event nouns can take different support verbs:
- ('quarrel') have an argument
- ('reason') make an argument
- ('rest') take a break
- ('flight') make a break

A common test of SVs:

One frequent proposed characteristic of support verbs is that their nominal object can’t really be interrogated - meaning that the verb in question isn’t functioning as a self-standing verb. The following are not natural conversations:
- What did you heave? - A sigh.
- What have you made? - A decision to go home.
- What did you have? - A fight with my brother.
- What did you wreak? - Vengeance on my enemies.
- What did you lodge? - A complaint.

Interchangeable with Verbs

She heaved a sigh.
(He sighed.)

We made the decision to give up.
(We decided to give up.)

I took a bath.
(I bathed.)

He suffered a relapse.
(He relapsed.)

Let’s say a prayer.
(Let’s pray.)

Profiling Different Participants

Agent of event
- perform an operation
- inflict injury
- exact/whelm vengeance
- launch an attack
- give instructions
- submit an application
- ask a question

Undergoer of event
- undergo an operation
- sustain injury
- have a setback
- suffer a defeat
- undergo an operation
- receive a rebuke
- get advice
Beyond "light verbs"

Simple cases: the verb has essential no meaning except to reveal that its subject is necessarily a participant in the event named by the noun.
  - a. active role
  - b. passive role

More nuanced cases: the verb contributes information about register, attitude, aktionsart, or the like.

More extended cases: the verb identifies its subject as a participant in the larger scenario associated with the event named by the verb.

Examples

Simple, active:
  – he made a complaint

Nuanced:
  – he registered a complaint

Simple, active:
  – she gave an exam

Simple, passive:
  – he took/sat the exam

The "extended" cases

- The examples given in red belong to an extended view of the frame indicated by the noun. That is, the purpose of an examination is to evaluate someone's competence in some area; decisions on whether that competence has been achieved are part of the larger scenario and often have special verbs (pass vs. fail here); the purpose of a promise is a commitment to one person on the part of another to act in a certain way: expressing success or failure in acting on the promise also requires separate verb (keep vs. break in this case).
Examples

Simple, active:
– she gave an exam

Simple, passive:
– he took/sat the exam

Extended:
– he passed/failed the exam

Examples

Simple, active:
– she made a promise

Examples

Simple, active:
– she made a promise

Extended:
– she kept/broke her promise

Unintended omissions.

• All of the examples I’ve given are ones in which all of the arguments of a governing verb have to be participants in the event named by the noun. However, there are also causative support verbs in which the object is a participant in the frame, but the subject is external to it. (Examples?)

• There are verbs whose subjects control the passive (Undergoer) participant in the event named by the noun. To deserve consideration, merit review; etc.
For the full story, and then some, see...


Support Verbs with Adjective

- Verb > A is the lexical unit, A is semantic head, V is support verb, A may have its own complements (e.g., rid of).
- Exx: be + any predicate adjective; go crazy, turn red, get naked
- Comment: The unit rid of seems to occur only with a SV.

Support Prepositions

- Prep > N is the lexical unit, N is semantic head, V is support verb. N has its own valence.
- Exx: at risk, in danger, on fire, under scrutiny, under arrest
- Some are modifiable: at considerable risk, in grave danger, under careful scrutiny
- Comment: The P>N structure functions as an adjective; the N can have its own complements. at considerable risk to his health, in danger of collapse

More Complex Cases

- Verb > P > N is the lexical unit, N is semantic head, V is support verb, N is generally not expandable.
- Exx: take into account, take under consideration, have in (one’s) possession
Support Verbs with PP

Verb > Prep > N is the lexical unit, N is semantic head, V is support verb. With possession there are two alignments of the arguments:

Possessor - Possessed
I came into possession of these documents.

Possessed - Possessor
These documents have come into my possession.

Transparent Nouns

N of N

N > of > N is the lexical unit. The second N is semantic head for purposes of external selection.

Comment: sometimes the N > of > N is "transparent" to the pieces of an MWE; and sometimes the N > of > N is itself an MWE, especially in the case of aggregates and unitizers:

- a case of the flu
- a round of golf
- a herd of cattle
- a flock of geese
- a school of fish
- a pinch of salt
- a pod of whales

Those two types

- The left and right examples in the preceding slide are supposed to indicate two different kinds of situations: in the one on the left we have uses of transparent nouns where there are no restrictions on the kinds of nouns that can occur in dependent position; in the one on the left we have transparent nouns where the dependent nouns have preferred governors. These tend to be aggregates (school of fish, flock of geese) or units (stick of gum, slice of bread).
Types of transparent nouns

1. Aggregates
   - bunch, group, collection, herd, school, flock
2. Quantities
   - flood, number, scores, storm
3. Types
   - breed, class, ilk, kind, type, sort
4. Portions and Parts
   - half, segment, top, bottom, part
5. Unitizers
   - glass, bottle, box, serving
6. Evaluations
   - gem, idiot, prince

"Transparency"

- The term, I think, is from Naomi Sager.
- The idea is that the dependent noun - the second noun in a N+of+N phrase - has certain selectional relations to the external context of the phrase; the contextual pattern can be noticed by "seeing through" the transparent noun.
- The first example on the next page does not involve an MWE, but we can at least notice that the locative prepositions in and on go with nouns of particular semantic types (in with a volume and on with a surface, for example), and this is independent of whether there's an expression like part of between the preposition and its preferred collocate.

"Transparent" to what?

- Relation between locative preposition and object:
  - on the shelf; on this part of the shelf
  - in the room; in this part of the room
- Relation between verb and typical collocating object
  - play golf; play a round of golf
  - eat fish; eat this type of fish
- Relation between possessor and kin-term
  - my wife; my gem of a wife
  - her husband; her jerk of a husband

"Transparent" to what?

- Relation between typical subject and verb:
  - water would quench my thirst; a bit of water would quench my thirst
- Relation between typical modifier and noun head:
  - a fine mess; now that's a fine sort of mess
  - nice hot coffee; a nice hot cup of coffee
Compounds

The compounds

- There is a world of things to say about compounds. Here I was only interested in cases where there's no regular analysis between the two pieces, cases where there is something regular to say about the relationship between the two parts yet the result names a pre-existing category, and cases in which the compound consists in having the first noun satisfy a frame requirement in the second noun.
- I thought that light year was a very strange compound, and I still think there's nothing else quite like it in English; but in the last few days I found Fußminuten and Autostunden, as measure of distance in German, and they do seem to join with light year (Lichtjahr) in establishing a pattern.
A-N Compounds

N > A is the lexical unit; listed compounds have the dependent in red; the syntactic head is the frame evoker, the dependent is either a frame element or a "quale".

Ready-made A+P compounds: hot news, friendly fire, blind alley, dead end

"Pertinative" adjectives

Pertinatives are adjectives whose senses are defined in (some) dictionaries with the phrase "of or pertaining to". Traditional term: relational adjectives. WordNet term: pertainyms.

They are not used predicatively in the same meaning. They aren't scalar, e.g., they don't get modified with very.

Pertinatives vs. Descriptives

judicial appointment
economic policy
educational practice
criminal law
linguistic society
Canadian government
national interest
these are MWEs

judicious appointment
economical housewife
educational experience
criminal behavior
ugly cat
amazing disclosure
bored child
these aren't

Mini-Grammars
3. Minigrammars

Some MWEs are generated by simple generative structures, usually finite state automata, for which dependency – or constituency – representations are not always relevant.
- Names
- Numbers
- Locations (addresses, coordinates)
- Time Expressions
- Kinterms
- Miscellaneous constructions

Personal Names

- Reverend Dr T. Allen Hampton-Smith III
- Components: titles, honorifics, given names, patronyms, family names, extensions, ...
- It would be best if names could be "sealed": handed over to the grammar with a category name and nothing else
- Sometimes personal names are penetrable. (Croatian)
  - Marko je Ivir zubar.
  - "Patrick is Hanks lexicographer.

English Kinterms

- grandfather, great grandfather, great great grandfather, etc.
- first cousin, second cousin, third cousin
- first cousin once removed, second cousin three times removed, etc.
- father-in-law, son-in-law, sister-in-law, etc.

siblings
cousins

second cousins

first cousins once removed

first cousins twice removed
Ordinary techniques of computational linguistics/corpus linguistics won't be able to recognize the constructional nature of some expressions.

Test case

another $600

This is just a bit of self-indulgence. In the full write-up of this paper I want to give a lot of attention to special grammatical constructions, and this unusual pattern with an indefinite determiner and a plural noun - with the requirement that there be both some kind of qualifier and a quantity (usually number) seemed strange, in the first place, and frequently hidden from view by the way we express numbered units of currency (e.g., by putting the "$" before the number while pronouncing it after the number).

<table>
<thead>
<tr>
<th>Indefinite article</th>
<th>Qualifier</th>
<th>Number</th>
<th>Plural Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>whopping</td>
<td>600</td>
<td>dollars</td>
</tr>
<tr>
<td>an</td>
<td>additional</td>
<td>10</td>
<td>pages</td>
</tr>
<tr>
<td>a</td>
<td>paltry</td>
<td>20</td>
<td>euros</td>
</tr>
<tr>
<td>a</td>
<td>respectable</td>
<td>6,000</td>
<td>francs</td>
</tr>
<tr>
<td>a</td>
<td>mere</td>
<td>-</td>
<td>pages</td>
</tr>
<tr>
<td>*a</td>
<td>-</td>
<td>12</td>
<td>pages</td>
</tr>
</tbody>
</table>

But now what do we do with "another $600"? We have to unpack it as "an other 600 $"!
Continuity Hypothesis

I assume the continuity of the lexicon and the constructicon.
Claim: many lexically-headed constructions can be analyzed as dependency subtrees.

Why this?

- Here the point is that in dealing with lexically headed special grammatical constructions, Kay and I have often found that they can be described in pretty much the same way as complex lexical items with contextual constraints that are pretty much like what you get with ordinary words. The full force of the "continuity hypothesis" is that from highly specific lexical items to maximally general grammatical patterns, there are always simple head-to-dependency relations of the kind mentioned in this paper.

Bottom Line

Lexical units can be represented as dependency subgraphs, specifying a semantic head, a syntactic head, required/preferred dependents.
Constraints on dependents can be specified lexically, sortally, morphosyntactically, and in terms of frame roles.
Dependents can be marked as "closed" (not open to modification) and/or "local" (not subject to extraction) and/or "omissible".
The lexical head of the construction bears information about contextual constraints: finiteness, inflection, polarity, etc.