

# The Syntax of Idioms

LIN 205



April 16, 2018

## THE PURPOSE OF SYNTAX?

- ▶ Since there's no upper bound on the number of potential sentences, we can't be simply listing them in our brains
- ▶ We have no problem interpreting sentences that we have never been encountered before
- ▶ This is straightforwardly explained if we assume that English is compositional, i.e. it obeys *the principle of compositionality*

## PRINCIPLE OF COMPOSITIONALITY

- Definition: The meaning of a complex expression depends on the meanings of the parts, and how the parts combine

$$(6 / 2) + 1 = 4$$

$$6 / (2 + 1) = 2$$

I shot (an elephant) in my pajamas.

I shot (an elephant in my pajamas).

- So the lexicon provides you with “ingredients”, syntax gives you a “recipe” for arriving at the meaning of complex expressions by specifying how smaller elements are allowed to combine

## LEXICON

- ▶ Assumed to consist of *lexemes*: any irreducibly meaningful unit, such as the word **dog**, or the suffix **-ness**
- ▶ If an association between form and meaning is unpredictable, it has to be stored in the lexicon

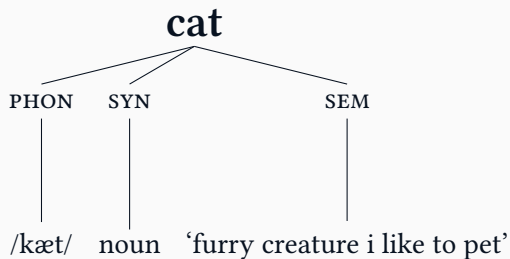
There is nothing doglike about the sequence [dɒg]

There is nothing catlike about the sequence [kæt]

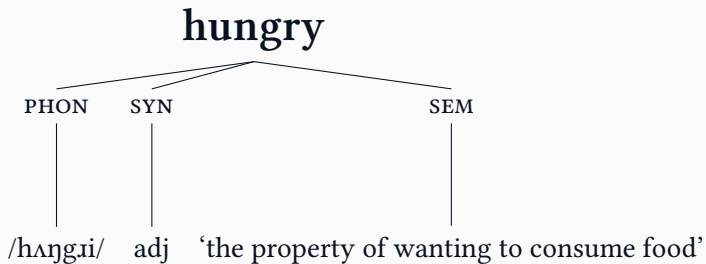
## RECAP

- ▶ We have to distinguish between what speakers *know*, and what they are able to figure out on the basis of what they know
- ▶ If you know what **hungry** means, and you know what **cat** means, then you can figure out **hungry cat** means
- ▶ i.e. you *do not* have to store the meaning of **hungry cat** in your lexicon. You can build these sorts of meanings on the fly.

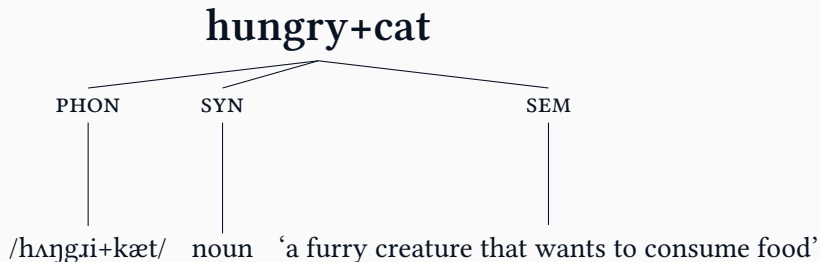
# LEXICAL ENTRIES



## LEXICAL ENTRIES



# COMBINATION



- Our grammar allows us to compute phon, syn, and sem

SYNTACTIC RULE:  $N \rightarrow \text{Adj } N$

SEMANTIC CORRELATE: A noun that has the property denoted by Adj



## MOVING ON TO IDIOMS

So what's the problem with idioms?

- ▶ It's raining cats and dogs.
- ▶ Stop beating a dead horse.
- ▶ Break a leg!
- ▶ Stop throwing him under the bus.
- ▶ Don't get so bent out of shape.

# IDIOMS

- ▶ A tentative definition of **IDIOM**: a multi-word expression whose meaning is unpredictable given the meanings of its parts
- ▶ Idioms appear to be phrases composed of multiple words, but somehow they don't seem to obey the principle of compositionality
- ▶ Maybe they're just listed in the lexicon – we could call them “words-with-spaces”

**Prediction:** since they're rigid like words, you can't manipulate them syntactically

## PHRASES ARE FLEXIBLE, WORDS ARE 'FROZEN'

Given a standard English sentence, it's possible to shuffle the units around and manipulate the structure without changing its meaning

1. A dog bit John.
2. John was bitten by a dog.

The same holds true for phrases, e.g. NP

3. the angry guy you saw yesterday
4. the guy you saw yesterday who was angry

But not for words:

5. un-break-able
6. #able-break-un
7. im-patient
8. #un-patient

## FROZENNESS / RIGIDITY

Consider **kick the bucket**. What happens when we substitute a different determiner?

1. John kicked the bucket.
2. John kicked **a** bucket.
3. John kicked **that** bucket.
4. John kicked **some** bucket.
5. John kicked **many** buckets.

John is dead 😞  
 $\Delta$ det, John is alive 😊  
 $\Delta$ det, John is alive 😊  
 $\Delta$ det, John is alive 😊  
 $\Delta$ det, John is alive 😊

## FROZENNESS / RIGIDITY

What happens when we substitute a different noun or verb?

- |  |                           |
|--|---------------------------|
| 1. John kicked the bucket.             | +idiomatic                |
| 2. John kicked the <b>pail</b> .       | $\Delta$ noun, -idiomatic |
| 3. John kicked the <b>can</b> .        | $\Delta$ noun, -idiomatic |
| 4. John kicked the <b>tub</b> .        | $\Delta$ noun, -idiomatic |
| 5. John <b>punted</b> the bucket.      | $\Delta$ verb, -idiomatic |
| 6. John <b>booted</b> the bucket.      | $\Delta$ verb, -idiomatic |
| 7. John <b>drop-kicked</b> the bucket. | $\Delta$ verb, -idiomatic |

# SYNTACTIC MANIPULATIONS

What happens when we passivize or topicalize?

1. ACTIVE: John kicked the bucket.



2. PASSIVE: The bucket was kicked by John.

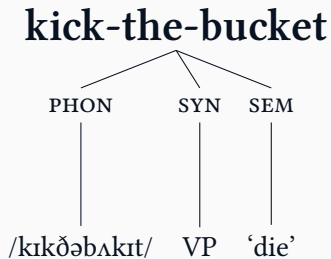
**-idiomatic**

3. TOPICALIZED: The bucket, John kicked.

**-idiomatic**

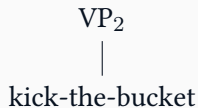
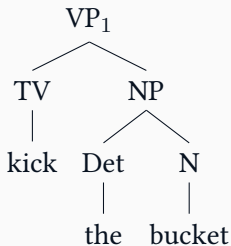
If we substitute any of the parts or shuffle them around, we lose the idiomatic meaning. Seems to be frozen, just like an ordinary word.

## LEXICAL ENTRY



## TREES FOR BOTH READINGS

1. kick the bucket (literal)
2. kick-the-bucket (idiomatic)





## SUBSTITUTION

But let's try another idiom: **spill the beans**.

- |                                       |                           |
|---------------------------------------|---------------------------|
| 1. Eric spilled the beans.            | +idiomatic                |
| 2. #Eric <b>dropped</b> the beans.    | $\Delta$ verb, -idiomatic |
| 3. #Eric spilled the <b>legumes</b> . | $\Delta$ noun, -idiomatic |

But it seems that choice of determiner is variable:

4. The call rests with Johnson, who is not about to spill **any** beans.
5. We're going to try to give him a chance to spill **some** beans.
6. Don't look at me, I didn't spill **a single** bean.

## SYNTACTIC MANIPULATIONS

We can even passivize or topicalize:

- |  |                   |
|--|-------------------|
| 1. ACTIVE: Eric spilled the beans.                   | <b>+idiomatic</b> |
| 2. PASSIVE: The beans got spilled yesterday.         | <b>+idiomatic</b> |
| 3. TOPICALIZED: Those beans, Eric would never spill. | <b>+idiomatic</b> |

## TO PULL STRINGS

What about **to pull strings**?

- |  |                           |
|--|---------------------------|
| 1. Eric pulled the strings for us.         | +idiomatic                |
| 2. #Eric <b>tugged</b> the strings for us. | $\Delta$ verb, -idiomatic |
| 3. #Eric pulled the <b>CORDS</b> for us.   | $\Delta$ noun, -idiomatic |
| 4. Eric pulled <b>SOME</b> strings for us. | $\Delta$ det, +idiomatic  |

## DETERMINER VARIATION

Determiner variation appears to be pretty robust here:

- |   |                             |
|---|-----------------------------|
| 1. Eric could pull <b>a few</b> strings for us.                               | $\Delta$ det, +idiomatic    |
| 2. Eric pulled <b>some</b> strings on our behalf.                             | $\Delta$ det, +idiomatic    |
| 3. Eric managed to pull <b>those</b> strings for us.                          | $\Delta$ det, +idiomatic    |
| 4. Eric pulled <b>all of the</b> strings he could.                            | $\Delta$ det, +idiomatic    |
| 5. I'm not sure Eric could pull<br><b>enough</b> strings to get the job done. | $\Delta$ det, +idiomatic    |
| 6. Maybe Eric could pull $\emptyset$ strings<br>to cut him loose.             | $\emptyset$ det, +idiomatic |

## SYNTACTIC MANIPULATIONS

Once again, idiomatic meaning is retained when passivizing or topicalizing

- |  |                   |
|--|-------------------|
| 1. ACTIVE: Eric pulled some strings for us.        | <b>+idiomatic</b> |
| 2. PASSIVE: Some strings were pulled for us.       | <b>+idiomatic</b> |
| 3. TOPICALIZED: Those strings, Eric pulled for us. | <b>+idiomatic</b> |

## OPEN SLOTS

Notice that the most flexible idioms seem to have open slots for their determiners:

1. spill **X** beans
2. pull **X** strings
3. yank **X**'s chain (requires possessive determiner)
4. pull **X**'s leg (requires possessive determiner)

## IDIOMS

We have to conclude that some idioms do have internal syntactic structure, and thus cannot be fixed lexical entries.

**QUESTION:** how can we account for the fact that some idioms are more syntactically flexible than others?

**POTENTIAL ANSWER:** Maybe we were wrong when we said that *all* idioms disobey the principle of compositionality

## A RETURN TO COMPOSITIONALITY

Consider the literal paraphrase of **spill the beans**

**spill the beans**  $\longrightarrow$  **reveal the secret**

We can metaphorically map each of the elements of the idiom onto the elements of its literal paraphrase

spill	the	beans
reveal	the	secret



## A RETURN TO COMPOSITIONALITY

Consider the literal paraphrase of **pull the strings**

**pull the strings**  $\longrightarrow$  **influence the situation**

Again, a portion of the literal meaning is distributed to each token of the idiom

pull	the	strings
influence	the	situation

## ONE MORE EXAMPLE

Consider the literal paraphrase of **turn the tables**

**turn the tables** → **reverse the outcome**

Again, since the structure of the idiom mirrors the structure of the paraphrase, we should expect syntactic variability, e.g.

1. How early did they decide to **turn the tables** on Littlefinger?
2. They can't understand how **the tables have been turned** on them so swiftly.
3. Well, this morning, **the tables are being turned** and it is the credit card companies themselves on the hot seat in front of Congress.

## A RETURN TO COMPOSITIONALITY

What's the problem with **kick the bucket**?

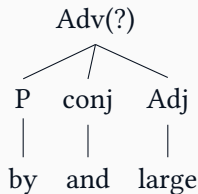
**kick the bucket** → **die**

kick	the	bucket
die	???	???

We can't make a similar mapping. It's not clear how the literal meaning could be distributed over the parts of the idiom, so there is a *mismatch* between the structure of the idiom and the structure of the paraphrase.

## BY AND LARGE

Consider **by and large**: an idiom consisting of a preposition, a, conjunction, and an adjective, but it functions as an adverb, and means ‘generally’



- This clearly doesn't mesh with our  $X \rightarrow X \text{ conj } X$  rule.

## COMPLETELY FIXED PHRASES

We could address this by adding a (somewhat ridiculous) new rule to our mini-grammar:

$\text{Adv} \rightarrow \text{P conj Adj}$

But this will massively overgenerate:

1. #from and yellow
2. #on but strong
3. #in or tasty

- It is very unlikely that there is some general syntactic rule which could assemble 'by and large.' It is completely idiosyncratic, and completely fixed.

## THREE-WAY CATEGORIZATION OF IDIOMS

We can lump idioms into three distinct groups:

1. FIXED IDIOMS:

- ▶ by and large
- ▶ all of a sudden

2. SEMI-FIXED IDIOMS:

- ▶ kick the bucket
- ▶ shoot the breeze

3. FLEXIBLE IDIOMS:

- ▶ pull X strings
- ▶ spill X beans

## THREE-WAY CATEGORIZATION OF IDIOMS

Idioms	Morphologically Alternating	Compositional Semantics	Syntactic Flexibility
Fixed	—	—	—
Semi-fixed	+	—	—
Flexible	+	+	+

## THREE-WAY CATEGORIZATION OF IDIOMS

### 1. FIXED IDIOMS:

- ▶ can disobey ordinary syntactic rules, and are not semantically compositional
- ▶ in all likelihood, they are listed in the lexicon

### 2. SEMI-FIXED IDIOMS:

- ▶ are constructed according to ordinary syntax rules, but are not semantically compositional
- ▶ hybrid creatures that are partially word-like and partially phrase-like

### 3. FLEXIBLE IDIOMS:

- ▶ obey ordinary syntactic rules, *are* semantically compositional, and thus syntactically variable
- ▶ basically ordinary phrases: the only catch is that the components are assigned non-literal meanings