

Last update: December 4, 2008

COMMUNICATION AND LANGUAGE

CMSC 421: CHAPTER 22

Outline

- ◇ Communication
- ◇ Grammar
- ◇ Syntactic analysis
- ◇ Problems

Communication

“Classical” view (pre-1953):

language consists of sentences that are true/false (cf. logic)

“Modern” view (post-1953):

language is a form of action

Wittgenstein (1953) *Philosophical Investigations*

Austin (1962) *How to Do Things with Words*

Searle (1969) *Speech Acts*

Why?

Communication

“Classical” view (pre-1953):

language consists of sentences that are true/false (cf. logic)

“Modern” view (post-1953):

language is a form of action

Wittgenstein (1953) *Philosophical Investigations*

Austin (1962) *How to Do Things with Words*

Searle (1969) *Speech Acts*

Why? **To change the actions of other agents**

Speech acts

SITUATION

Speaker → **Utterance** → **Hearer**

Speech acts are attempts to achieve the speaker's goals:

Inform	"There's a pit in front of you."
Query	"Can you see the gold?"
Request	"Could you open my file for me?"
Command	"Pick it up."
Promise	"I'll share the gold with you."
Acknowledge	"OK"

Planning a speech act requires knowledge of

- Situation
- Semantic and syntactic conventions
- Hearer's goals, knowledge base, and rationality

Stages in communication (informing)

Intention	S wants to inform H that P
Generation	S selects words W to express P in context C
Synthesis	S utters words W
Perception	H perceives W' in context C'
Analysis	H infers possible meanings P_1, \dots, P_n
Disambiguation	H infers intended meaning P_i
Incorporation	H incorporates P_i into KB

How could this go wrong?

Stages in communication (informing)

Intention	S wants to inform H that P
Generation	S selects words W to express P in context C
Synthesis	S utters words W
Perception	H perceives W' in context C'
Analysis	H infers possible meanings P_1, \dots, P_n
Disambiguation	H infers intended meaning P_i
Incorporation	H incorporates P_i into KB

How could this go wrong?

- Insincerity (S doesn't believe P)
- Speech wreck ignition failure
- Ambiguous utterance
- Differing understanding of current context ($C \neq C'$)

Grammar

Vervet monkeys, antelopes etc. use isolated symbols for sentences

⇒ restricted set of communicable propositions, no **generative capacity**
(Chomsky (1957): *Syntactic Structures*)

Grammar specifies the compositional structure of complex messages
e.g., text (linear), speech (linear), music (multi-dimensional)

A **formal language** is a set of **strings** of **terminal symbols**

Each string in the language can be analyzed/generated by the grammar

The grammar is a set of **rewrite rules**, e.g.,

$$S \rightarrow NP VP$$

$$Article \rightarrow \mathbf{the} \mid \mathbf{a} \mid \mathbf{an} \mid \dots$$

Here S is the **sentence** symbol, NP and VP are **nonterminals**

Grammar types

Regular: *nonterminal* \rightarrow *terminal*[*nonterminal*]

$$S \rightarrow aS$$

$$S \rightarrow \Lambda$$

Context-free: *nonterminal* \rightarrow *anything*

$$S \rightarrow aSb$$

Context-sensitive: more nonterminals on right-hand side

$$ASB \rightarrow AAaBB$$

Recursively enumerable: no constraints

Related to Post systems and Kleene systems of rewrite rules

Natural languages probably context-free, parsable in real time

Wumpus lexicon

Noun → *stench* | *breeze* | *glitter* | *nothing*
| *wumpus* | *pit* | *pits* | *gold* | *east* | ...

Verb → *is* | *see* | *smell* | *shoot* | *feel* | *stinks*
| *go* | *grab* | *carry* | *kill* | *turn* | ...

Adjective → *right* | *left* | *east* | *south* | *back* | *smelly* | ...

Adverb → *here* | *there* | *nearby* | *ahead*
| *right* | *left* | *east* | *south* | *back* | ...

Pronoun → *me* | *you* | *I* | *it* | ...

Name → *John* | *Mary* | *CollegePark* | *UMD* | ...

Article → *the* | *a* | *an* | ...

Preposition → *to* | *in* | *on* | *near* | ...

Conjunction → *and* | *or* | *but* | ...

Digit → **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9**

Divided into **closed** and **open** classes

Wumpus lexicon

Noun → *stench* | *breeze* | *glitter* | *nothing*
| *wumpus* | *pit* | *pits* | *gold* | *east* | ...

Verb → *is* | *see* | *smell* | *shoot* | *feel* | *stinks*
| *go* | *grab* | *carry* | *kill* | *turn* | ...

Adjective → *right* | *left* | *east* | *south* | *back* | *smelly* | ...

Adverb → *here* | *there* | *nearby* | *ahead*
| *right* | *left* | *east* | *south* | *back* | ...

Pronoun → *me* | *you* | *I* | *it* | *they* | *y'all* ...

Name → *John* | *Mary* | *CollegePark* | *UMD* | ...

Article → *the* | *a* | *an* | ...

Preposition → *to* | *in* | *on* | *near* | ...

Conjunction → *and* | *or* | *but* | ...

Digit → **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9**

Divided into **closed** and **open** classes

Wumpus grammar

$S \rightarrow NP VP$	I + feel a breeze
$S Conjunction S$	I feel a breeze + and + I smell a wumpus
$NP \rightarrow Pronoun$	I
$Noun$	pits
$Article Noun$	the + wumpus
$Digit Digit$	3 4
$NP PP$	the wumpus + to the east
$NP RelClause$	the wumpus + that is smelly
$VP \rightarrow Verb$	stinks
$VP NP$	feel + a breeze
$VP Adjective$	is + smelly
$VP PP$	turn + to the east
$VP Adverb$	go + ahead
$PP \rightarrow Preposition NP$	to + the east
$RelClause \rightarrow \mathbf{that} VP$	that + is smelly

Parse trees

Exhibit the grammatical structure of a sentence

I **shoot** **the** **wumpus**

Parse trees

Exhibit the grammatical structure of a sentence

Pronoun

I

Verb

shoot

Article

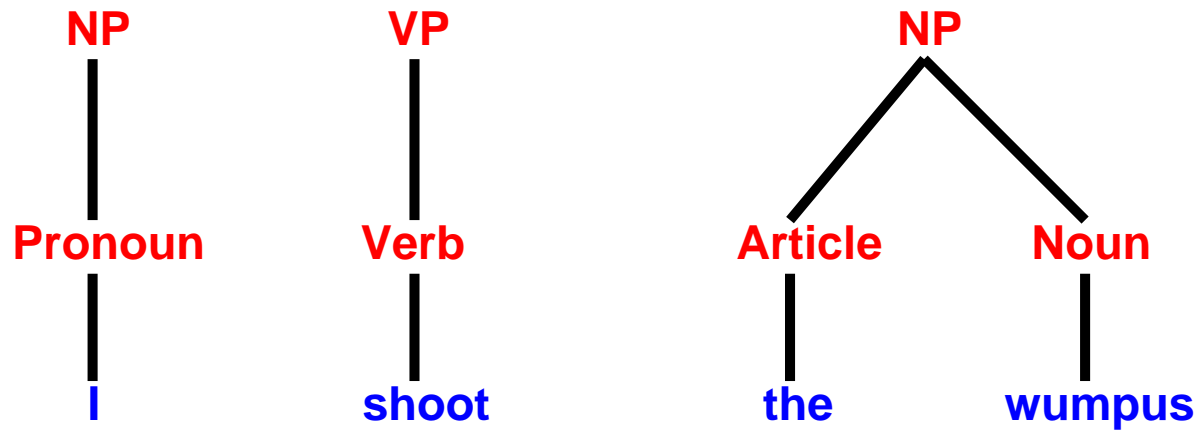
the

Noun

wumpus

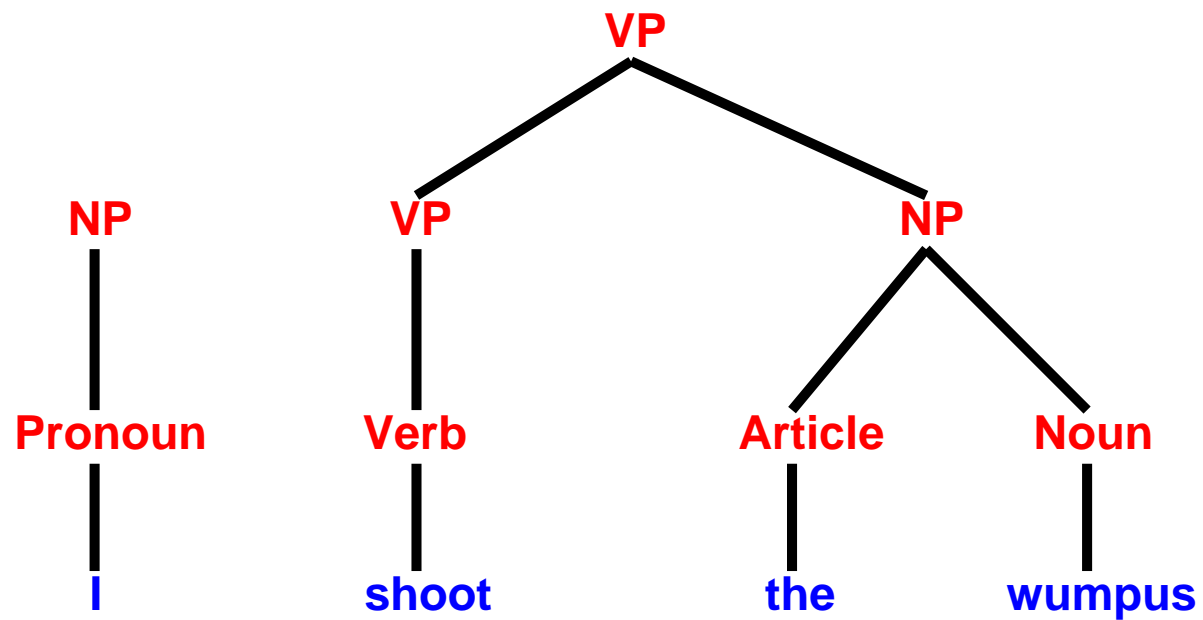
Parse trees

Exhibit the grammatical structure of a sentence



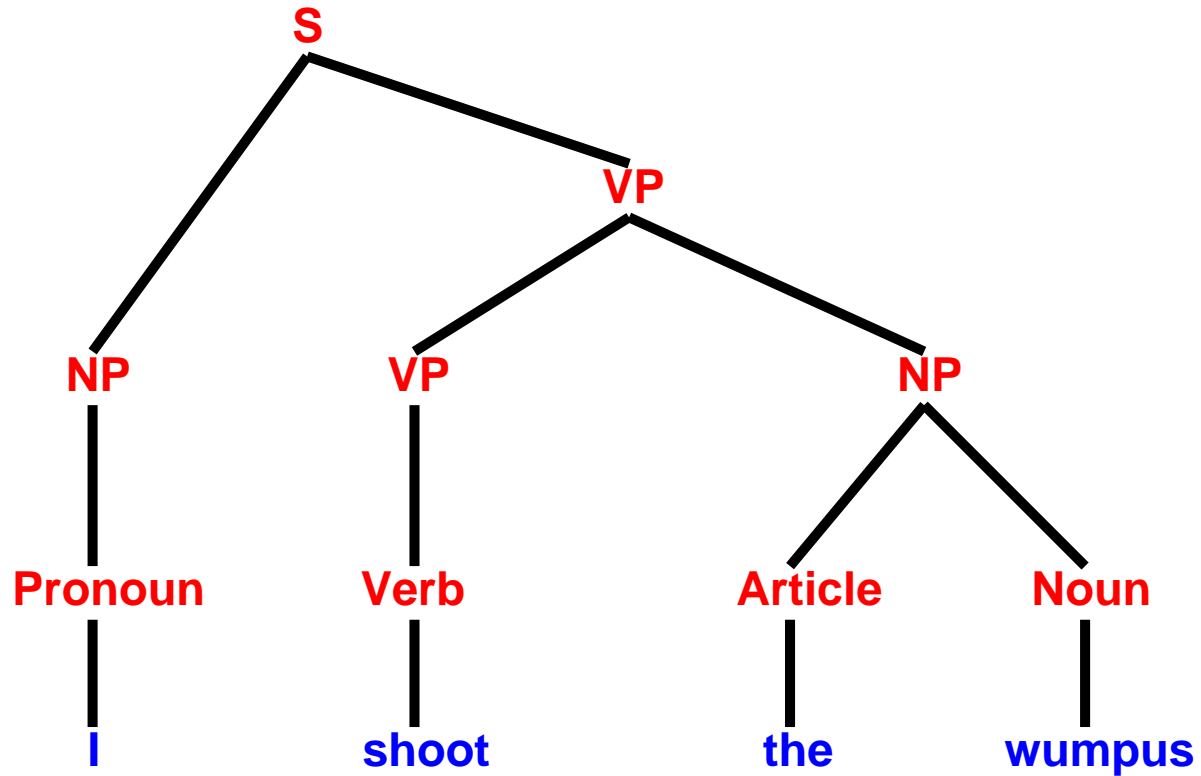
Parse trees

Exhibit the grammatical structure of a sentence



Parse trees

Exhibit the grammatical structure of a sentence



Context-free parsing

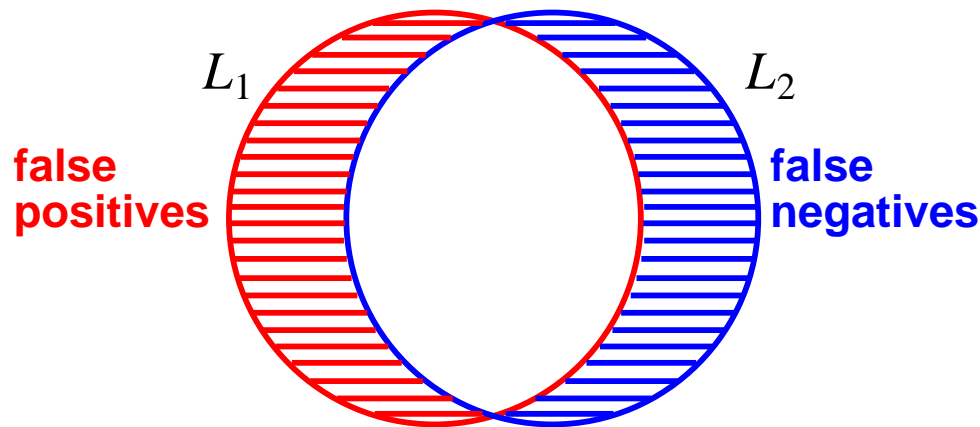
Bottom-up parsing works by replacing any substring that matches RHS of a rule with the rule's LHS

Efficient algorithms (e.g., chart parsing, Section 22.3) $O(n^3)$ for context-free, run at several thousand words/sec for real grammars

Context-free parsing \equiv Boolean matrix multiplication (Lee, 2002)
 \Rightarrow unlikely to find faster practical algorithms

Grammaticality judgements

Formal language L_1 may differ from natural language L_2



Adjusting L_1 to agree with L_2 is a learning problem!

- * the gold grab the wumpus
- * I smell the wumpus the gold
- I give the wumpus the gold
- * I donate the wumpus the gold

Intersubjective agreement somewhat reliable, independent of semantics
Real grammars 10–500 pages, insufficient even for “proper” English

Logical grammars

BNF notation for grammars too restrictive:

- difficult to add “side conditions” (number agreement, etc.)
- difficult to connect syntax to semantics

Idea: express grammar rules as logic

$X \rightarrow YZ$ becomes $Y(s_1) \wedge Z(s_2) \Rightarrow X(\text{Append}(s_1, s_2))$

$X \rightarrow \mathit{word}$ becomes $X([\mathit{word}])$

$X \rightarrow Y \mid Z$ becomes $Y(s) \Rightarrow X(s) \quad Z(s) \Rightarrow X(s)$

Here, $X(s)$ means that string s **can be interpreted** as an X

Logical grammars, continued

Now it's easy to augment the rules

$$NP(s_1) \wedge Human(Ref(s_1)) \wedge VP(s_2) \\ \Rightarrow NP(Append(s_1, ["who"], s_2))$$

$$NP(s_1) \wedge Number(s_1, n) \wedge VP(s_2) \wedge Number(s_2, n) \\ \Rightarrow S(Append(s_1, s_2))$$

Parsing is reduced to logical inference:

$$ASK(KB, S(["I" "am" "a" "wumpus"]))$$

(Can add extra arguments to return the parse structure, semantics)

Generation simply requires a query with uninstantiated variables:

$$ASK(KB, S(x))$$

If we add arguments to nonterminals to construct sentence semantics, NLP generation can be done from a given logical sentence:

$$ASK(KB, S(x, At(Robot, [1, 1])))$$

Syntax in NLP

Most view syntactic structure as an essential step towards meaning;

“Mary hit John” \neq “John hit Mary”

Can you figure out the meaning of the following?

“And since I was not informed—as a matter of fact, since I did not know that there were excess funds until we, ourselves, in that checkup after the whole thing blew up, and that was, if you’ll remember, that was the incident in which the attorney general came to me and told me that he had seen a memo that indicated that there were no more funds.”

–President Ronald Reagan, April 28, 1987

<http://www.reagan.utexas.edu/archives/speeches/1987/042887e.htm>

Syntax in NLP

Most view syntactic structure as an essential step towards meaning;

“Mary hit John” \neq “John hit Mary”

Can you figure out the meaning of the following?

“And **since** I was not informed—as a matter of fact, **since** I did not know that there were excess funds **until** we, ourselves, in that checkup after the whole thing blew up, and that was, if you’ll remember, that was the incident in which the attorney general came to me and told me that he had seen a memo that indicated that there were no more funds.”

–President Ronald Reagan, April 28, 1987

<http://www.reagan.utexas.edu/archives/speeches/1987/042887e.htm>

since I was not informed, *I did what?*

until we, ourselves, *did what?*

Syntax in NLP

Most view syntactic structure as an essential step towards meaning;

“Mary hit John” \neq “John hit Mary”

Can you figure out the meaning of each of the following?

“The first one won one one-dollar bill.”

“John, where Mary had had ‘had,’ had had ‘had had.’”

“Wouldn’t the sentence ‘I want to put a hyphen between the words Fish and And and And and Chips in my Fish-And-Chips sign’ have been clearer if quotation marks had been placed before Fish, and between Fish and and, and and and And, and And and and, and and and And, and And and and, and and and Chips, as well as after Chips?”

Syntax in NLP

Most view syntactic structure as an essential step towards meaning;

“Mary hit John” \neq “John hit Mary”

Can you figure out the meaning of each of the following?

“The first one won one one-dollar bill.”

“John, where Mary had had ‘had,’ had had ‘had had.’”

“Wouldn’t the sentence ‘I want to put a hyphen between the words
“Fish” and And and And and Chips in my Fish-And-Chips
sign’ have been clearer if quotation marks had been placed

before Fish, and **between Fish and and**,

and and and And, and And and and,

and and and And, and And and and,

and and and Chips, as well as after Chips?”

Syntax in NLP

Most view syntactic structure as an essential step towards meaning;

“Mary hit John” \neq “John hit Mary”

Can you figure out the meaning of each of the following?

“The first one won one one-dollar bill.”

“John, where Mary had had ‘had,’ had had ‘had had.’”

“Wouldn’t the sentence ‘I want to put a hyphen between the words “Fish” and “And” and And and Chips in my Fish-And-Chips sign’ have been clearer if quotation marks had been placed before Fish, andbetween Fish and and, and **and and And**, and **And and and**, and and and And, and And and and, and and and Chips, as well as after Chips?”

Syntax in NLP

Most view syntactic structure as an essential step towards meaning;

“Mary hit John” \neq “John hit Mary”

Can you figure out the meaning of each of the following?

“The first one won one one-dollar bill.”

“John, where Mary had had ‘had,’ had had ‘had had.’”

“Wouldn’t the sentence ‘I want to put a hyphen between the words “Fish” and “And” and “And” and Chips in my Fish-And-Chips sign’ have been clearer if quotation marks had been placed before Fish, andbetween Fish and and, and and and And, and And and and, and **and and And**, and **And and and**, and and and Chips, as well as after Chips?”

Syntax in NLP

Most view syntactic structure as an essential step towards meaning;

“Mary hit John” \neq “John hit Mary”

Can you figure out the meaning of each of the following?

“The first one won one one-dollar bill.”

“John, where Mary had had ‘had,’ had had ‘had had.’”

“Wouldn’t the sentence ‘I want to put a hyphen between the words “Fish” and “And” and “And” and “Chips” in my Fish-And-Chips sign’ have been clearer if quotation marks had been placed before Fish, and between Fish and and, and and and And, and And and and, and and and And, and And and and, and **and and Chips**, as well as **after Chips**?”

Syntax in NLP, continued

On the other hand, syntax by itself isn't sufficient to provide meaning.

Here's a sentence that's nonsensical but is grammatically correct:

Colorless green ideas sleep furiously.

–Noam Chomsky, 1957

Real language

Real human languages provide many problems for NLP:

- ◇ ambiguity
- ◇ anaphora
- ◇ indexicality
- ◇ vagueness
- ◇ discourse structure
- ◇ metonymy
- ◇ metaphor
- ◇ noncompositionality

Ambiguity

Squad helps dog bite victim

Ambiguity

Squad helps dog bite victim

Helicopter powered by human flies

Ambiguity

Squad helps dog bite victim

Helicopter powered by human flies

American pushes bottle up Germans

Ambiguity

Squad helps dog bite victim

Helicopter powered by human flies

American pushes bottle up Germans

I ate spaghetti with meatballs

Ambiguity

Squad helps dog bite victim

Helicopter powered by human flies

American pushes bottle up Germans

I ate spaghetti with meatballs
salad

Ambiguity

Squad helps dog bite victim

Helicopter powered by human flies

American pushes bottle up Germans

I ate spaghetti with meatballs

salad

abandon

Ambiguity

Squad helps dog bite victim

Helicopter powered by human flies

American pushes bottle up Germans

I ate spaghetti with meatballs

salad

abandon

a fork

Ambiguity

Squad helps dog bite victim

Helicopter powered by human flies

American pushes bottle up Germans

I ate spaghetti with meatballs

salad

abandon

a fork

a friend

Ambiguity

Squad helps dog bite victim

Helicopter powered by human flies

American pushes bottle up Germans

I ate spaghetti with meatballs

salad

abandon

a fork

a friend

Ambiguity can be lexical (polysemy), syntactic, semantic, referential

Anaphora

Using pronouns to refer back to entities already introduced in the text

After Mary proposed to John, **they** found a preacher and got married.

Anaphora

Using pronouns to refer back to entities already introduced in the text

After Mary proposed to John, **they** found a preacher and got married.

For the honeymoon, **they** went to Hawaii

Anaphora

Using pronouns to refer back to entities already introduced in the text

After Mary proposed to John, **they** found a preacher and got married.

For the honeymoon, **they** went to Hawaii

Mary saw a ring through the window and asked John for **it**

Anaphora

Using pronouns to refer back to entities already introduced in the text

After Mary proposed to John, **they** found a preacher and got married.

For the honeymoon, **they** went to Hawaii

Mary saw a ring through the window and asked John for **it**

Mary threw a rock at the window and broke **it**

Indexicality

Indexical sentences refer to utterance situation (place, time, S/H, etc.)

I am over **here**

Why did **you** do **that**?

Metonymy

Using one noun phrase to stand for another

I've read **Shakespeare**

Chrysler announced record profits

The **ham sandwich** on Table 4 wants another beer

Metaphor

“Non-literal” usage of words and phrases, often systematic:

I’ve tried killing the process but it won’t die. Its parent keeps it alive.

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

alligator shoes

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

alligator shoes

designer shoes

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

alligator shoes

designer shoes

brake shoes

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

alligator shoes

designer shoes

brake shoes

red book

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

alligator shoes

designer shoes

brake shoes

red book

red pen

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

alligator shoes

designer shoes

brake shoes

red book

red pen

red hair

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

alligator shoes

designer shoes

brake shoes

red book

red pen

red hair

red herring

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

alligator shoes

designer shoes

brake shoes

red book

red pen

red hair

red herring

small moon

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

alligator shoes

designer shoes

brake shoes

red book

red pen

red hair

red herring

small moon

large molecule

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

alligator shoes

designer shoes

brake shoes

red book

red pen

red hair

red herring

small moon

large molecule

mere child

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

alligator shoes

designer shoes

brake shoes

red book

red pen

red hair

red herring

small moon

large molecule

mere child

alleged murderer

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

alligator shoes

designer shoes

brake shoes

red book

red pen

red hair

red herring

small moon

large molecule

mere child

alleged murderer

real leather

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

basketball shoes

baby shoes

alligator shoes

designer shoes

brake shoes

red book

red pen

red hair

red herring

small moon

large molecule

mere child

alleged murderer

real leather

artificial grass