Underlying Form

Richard K. Larson (Stony Brook Univ.)

Question: What kind of object is a natural language?

For centuries natural language has been viewed primarily as an **artifact** or **created object**, an aspect of human culture much like tool-making or agriculture:

- discovered or invented by humans
- taught by humans to their children
- something that could be tinkered with and perhaps even improved (Esperanto, Basic English, Interlingua,…)

Midway through the last century that picture began to change. Science now views human language as a **natural object**, a unique aspect of human biology much like upright bipedal posture.

- evolved in the human lineage
- "grows" spontaneously in human children (in the proper environment)
- invariant across the human species (barring pathology or injury)

This change of perspective is important. There is evidence that the distinction between **artifacts** and **natural objects** is intrinsically significant for us, detectable in children in the course of conceptual development (Keil 1986).

Young children seem to analyze objects using characteristic surface features & properties: "what you see is what you get". They later learn that not all objects are the same.

- Artifacts are basically whatever we make them to be, as a matter of convention.
- Natural objects have their own defining properties, which may be concealed by surface appearances, and need to be discovered.

When we change our view of an object as artifact vs. natural, we change our reactions to it.

- With an artifact, we can ask: When/How/Why was it made?
  We cannot ask: What is it really? What are its essential properties & structure?
- With a natural object, we can.

1.0 Attributing Structure

In the course of exploring natural objects and trying to understand their properties it typically becomes necessary, at some point, to attribute **structure** to them.

1.1 Chemical Theory

Democritus (400 BCE): all material bodies are aggregates of **atoms** (ἄτομος), particles too small to be seen by the eye. Atoms are indivisible and cannot be analyzed further.

Kinds of atoms:	Stone	(dry and heavy)
	Water	(wet and heavy)
	Air	(cold and light)
	Fire	(hot and slippery)

Material objects consist of combinations of these atoms in different proportions:

Plants: stone atoms + water atoms (from the soil) + fire atoms (from the sun).
Metals: stone atoms + fire atoms

Dalton (1807): retained Democritus' basic picture, but changed its atomic inventory.

Kinds of atoms: Hydrogen, Nitrogen, Oxygen, Carbon, …

Material substances consist of combinations of these atoms in different proportions; we specify them with **empirical formulae**:

Water:	H_2O
Carbon monoxide:	CO
Carbon dioxide:	CO_2
Nitric oxide:	NO
Nitrous oxide:	N_2O
Nitrogen peroxide:	NO_2

Question: To specify a material substance, is it enough to give the atomic constituents in their relative proportions - its empirical formula?

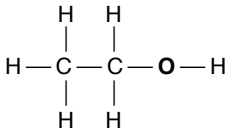
Answer: No! The empirical formula C_2H_6O corresponds to two distinct material substances. One is a liquid, and an intoxicant. The other is a gas!

C_2H_6O → volatile, flammable, colorless liquid (good to drink!)
C_2H_6O → flammable, colorless gas

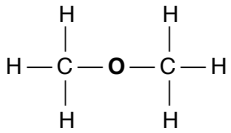
Question: How is this possible?

Answer: Substances have **structure**. The same atomic constituents in the same proportions can be structured in different ways.

Structural Diagrams:

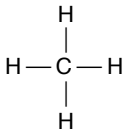


Ethyl Alcohol (C_2H_6O)

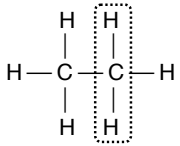


Dimethyl Ether (C_2H_6O)

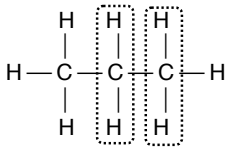
These structural diagrams clarify not only the nature of "ambiguous substances" (**isomers**) like ethyl alcohol and dimethyl ether, but also the regular relationships between whole families of substances:



Methane (CH_4)



Ethane(C_2H_6)



Propane(C_3H_8)

Summary: Chemistry started out with basic constituents and their proportions, and eventually came to structure.

1.2 Linguistic Theory

Dionysus Thrax (100 BCE): (Greek) sentences are composed of words belonging to particular grammatical categories (parts of speech) arranged in a certain linear order.

Parts of speech:	noun	("inflected for case, signifying a person, or thing")
	verb	("uninflected for case, but inflected for tense, person and number, signifying an activity or process performed or undergone")
	article	("inflected for case; pre- or postposed to nouns")
	adverb	("lacks inflection; qualifying or added to the verb")
	pronoun	("replaces a noun; marked for person")
	participal	("shares feartures of nouns and verbs")

	preposition	("is placed before other words in composition and in syntax")
	conjunction	("binds together the discourse and fills gaps in its interpretation"

Order:	N V N	("Subject – Verb – Object")
	N V N P N	("Subject – Verb – Object" – "Preposition – Object")
	P N	("Preposition – Object")

This information specifies many of essential properties of an expression, typically including its interpretation:

John saw Mary	(John = seer, Mary = seen)
Mary saw John	(Mary = seer, John = seen)
Mary gave Fido to John	(Mary = giver, Fido = given, John = recipient)
Mary gave John to Fido	(Mary = giver, John = given, Fido = recipient)

Thrax's views (like those of Democritus) prevailed in grammatical tradition for many centuries. To record a language, we write down:

A **dictionary**, specifying its "word atoms" and their classification into categories.

A **grammar**, specifying the patterns among its categories (together with whatever idiosyncrasies of inflection must be stated)

In this conception, **words and their linear arrangement** function for linguistics much like **atoms and their proportions** function for chemistry, and in many cases that seems to be enough.

Question: To specify a linguistic expression and its properties, is it indeed enough to give the atomic constituents (words) in their linear order?

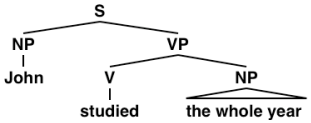
Answer: No! A single sequence like *John studied the whole year* can correspond to two quite different thoughts:

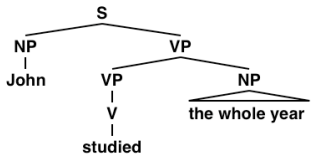
John studied the whole year	→	'John examined the entire year period'
John studied the whole year	→	'John studied for 12 months'

Question: How is this possible?

Answer: Expressions have **structure**. The same words in the same order can be structured in different ways.

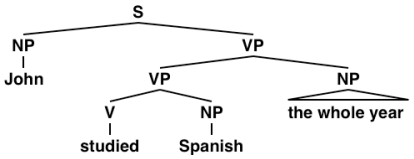
Structural Diagrams:



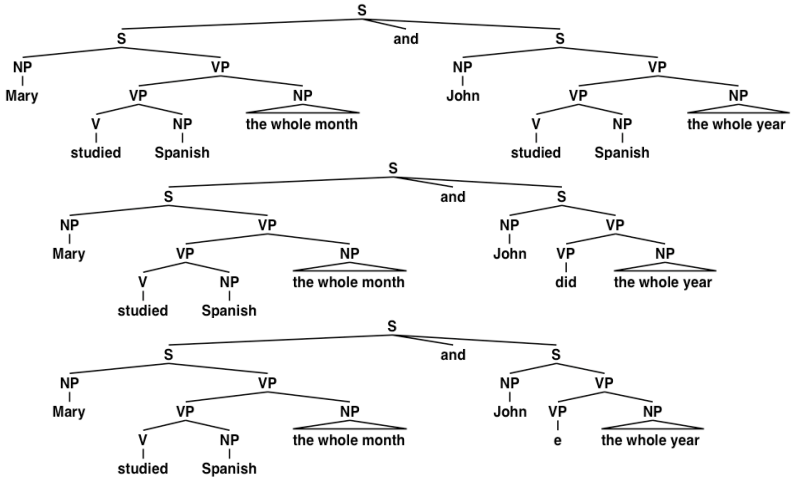


Again these structural diagrams clarify not only the nature of **ambiguous sentences**, but also the regular relationships between whole families of sentences:

Mary studied Spanish the whole year



*Mary studied Spanish the whole month and John **studied Spanish** the whole year*
*Mary studied Spanish the whole month and John **did** the whole year*
Mary studied Spanish the whole month and John __ the whole year



Summary: Linguistics began with basic elements and their linear ordering. It too eventually came to structure.

2.0 Deep Atoms?

The Chemistry – Linguistics analogy is close and suggestive, but it is <u>not</u> exact.

Chemistry succeeded in isolating a small number of universal atomic elements (118) as the basis of all chemical structure.

Chemical Atoms: Hydrogen, Helium, Lithium, …, Ununoctium

What are the corresponding atoms for linguistics?

Linguistic Atoms: ??

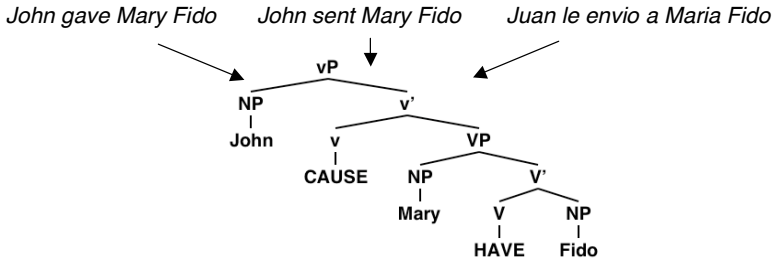
Categories? Linguistics has isolated a small number of (possibly universal) <u>categories</u> of words (nouns, verbs, etc.). But categories aren't sufficient to determine the properties of a sentence, e.g., its meaning.

John saw Mary
John kissed Mary

These involve the same categories (parts of speech) occurring in the same order. But they don't mean the same thing. Categories cannot be atoms.

Words? Words and the structure in which they occur <u>do</u> seem to determine the properties of an expression. But, the set of words of any single language is vast, and any two languages will have their own separate stocks of words. Individual words of individual languages cannot be atoms.

IDEA: Perhaps individual sentences of individual languages with their language-particular words are just "impure" expressions of a deeper linguistic substance.



On this view (ignoring names) true linguistic "atoms" would be a small number of universal abstract elements like *CAUSE* and *HAVE*, and perhaps also *BECOME*, *DO*, *BE*,

AT, *GO*, *TO*, etc. The English and Spanish examples would all share a deep, atomic structure as expressing "caused-possession".

This idea has exerted an enormous attraction on Linguistics – the idea of a deep chemistry of linguistic forms.

3.0 Blaming, Sending, Giving and Dragging

The "deep atoms" view allows apparently different sentences, even coming from different languages, to have the same underlying form. Conversely, superficially similar sentences may have different underlying forms. Consider:

(1) a. Job blamed God for his troubles.
b. Job blamed his troubles on God.

Differences here seem minor: just word order & preposition choice. BUT thinking further…..

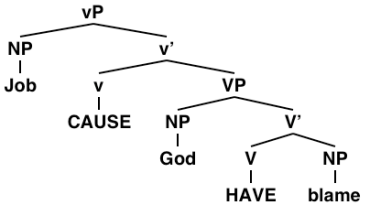
The first resembles *give*-type sentences:

(2) a. Job thanked God for his blessings.
b. Job gave God thanks for his blessings.
c. Job gave thanks to God for his blessings.

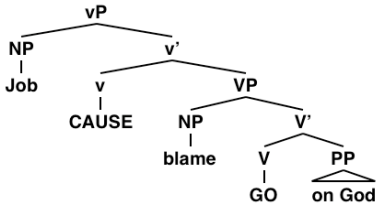
The second resembles *put*-type sentences:

(3) a. Job put his troubles on God.
b. Job put the blame for his troubles on God.

These would correspond to two quite different underlying forms:



Job blamed God for his troubles.
CAUSED-POSSESSION



Job blamed his troubles on God.
CAUSED-MOTION/LOCATION

Differences between other superfically similar sentences can be understood in a similar way:

(4) a. John sent **Fido to Mary**. OBLIQUE-FORM (USING A PrepPhrase)
b. John sent **Mary Fido**. "DOUBLE-OBJECT" FORM

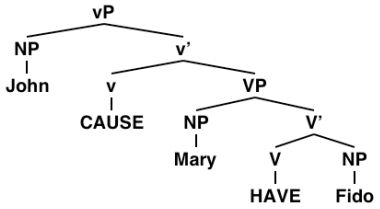
These look similar, and appear to have the same meaning, but …:

(5) a. John sent a letter to **Madrid**.
b. #John sent **Madrid** a letter. ('Sounds like Madrid is a person!)

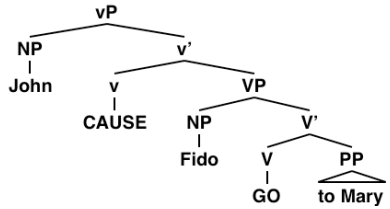
(6) a. John sent Fido **halfway** to Mary.
b. *John sent Mary Fido **halfway**.

(7) **Where** did John send Fido?
a. John sent Fido **to MARY**.
b. #John sent **MARY** Fido.

Once again it's attractive to associate (4a,b) with different underlying forms:



John sent Mary Fido
CAUSED-POSSESSION



John sent Fido to Mary
CAUSED-MOTION/LOCATION

Basic Question: How do surface sentences get matched up with underlying forms? What is the "mapping"?

Idea: Sentence form determines the mapping! One meaning-one underlying form.

DOUBLE-OBJECT FORM ➔ CAUSED-POSSESSION
OBLIQUE-FORM ➔ CAUSED-MOTION/LOCATION

(8) a. John sent/threw/mailed Mary Fido. ➔ Mary HAVE Fido
b. John sent/threw/mailed Fido to Mary. ➔ Fido GO-TO Mary

(9) a. John dragged/conveyed/lost/donated Fido to Mary.
b. *John dragged/conveyed/lost/donated Mary Fido.

Problem! What about verbs like *give (lend, sell, hand*, etc.)?

(10) a. John gave Mary Fido. DOUBLE-OBJECT FORM
b. John gave Fido to Mary. OBLIQUE-FORM

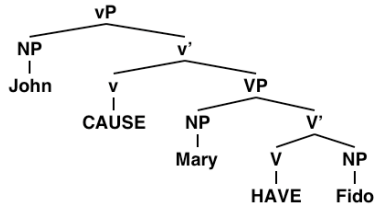
(11) a. #John gave a letter to Madrid. Both sound strange!
b. #John gave Madrid a letter.

(12) a. *John gave Peter halfway to Mary. Both are ungrammatical!
b. *John gave Mary Peter halfway.

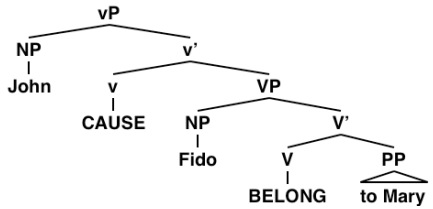
(13) *Where did John give Fido? Not possible!

Give (lend, sell, hand, etc.) behaves like a caused possession verb in both its double object and its oblique variant!

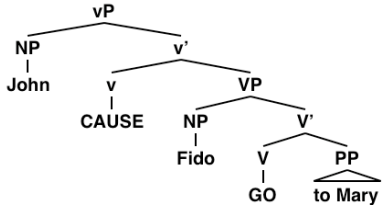
This suggests we must have two underlying forms for caused possession, not one. And oblique form must be able to encode **both** notions!



John gave Mary Fido
CAUSED-POSSESSION



John gave Fido to Mary
CAUSED-POSSESSION



John sent/dragged Fido to Mary
CAUSED-MOTION/LOCATION

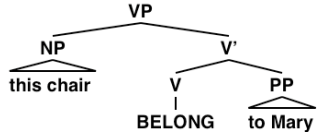
On reflection this may not be so surprising. The English verb *belong* has two meanings. It has a possession meaning: 'Object or entity X is the property of Y':

(14) a. This chair belongs **to Mary**.
b. A: **To whom** does this chair belong?
B: It belongs **to Mary**/***in there**.
c. The book/fault/responsibility/credit belongs to John
d. This piece belongs to that puzzle.

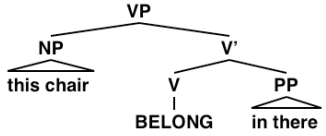
Belong also has a location meaning: 'Object or entity X is suitably, customarily or properly situated at location Y':

(15) a. This chair belongs **in the living room**/**near that one**/**there**.
b. A: **Where** does this chair belong?
B: It belongs **in there**/***to Mary**.
c. John belongs in prison.
d. Memories belong in the past.
e. This information belongs in the public domain.

In both cases we have oblique constructions (V- PP):



This chair belongs to Mary
POSSESSION



This chair belongs in there
LOCATION

Spanish shows something similar. (16a) is ambiguous between a CAUSED-LOCATION meaning, and an ASSEMBLY meaning involving part-whole relations. (16b) does not have both meaning since frogs cannot be parts of tables.

(16) a. María puso las patas **en** la mesa
María put the legs on the table
'Mary placed the legs upon the table' (CAUSED-LOCATION)
'Mary attached the table's legs' (ASSEMBLY)
b. María puso **las ranas** en la mesa
María put the frogs on the table
'Mary placed the frogs upon the table' (CAUSED-LOCATION)
'#Mary attached the table's frogs' (ASSEMBLY)

In Spanish we can substitute dative *a* → locative *en.* Now we only get the posession meaning, as with *belong* in English!

(17) a. María le puso las patas **a** la mesa
María Cl-3Dat put the legs to the table
'Mary attached the table's legs' (ASSEMBLY meaning only!)
b. #María puso las ranas **a** la mesa
María put the frogs to the table
'#Mary attached the table's frogs' (ASSEMBLY meaning only!)

Consider also European Portuguese. EP has no DOUBLE-OBJECT FORM, only OBLIQUE FORM. But it uses different prepositions to distinguish CAUSED-POSSESSION/CAUSED-MOTION/LOCATION:

(18) a. O João enviou uma carta **à** Maria/**para** a Maria
the John sent a letter to Maria
'John sent a letter to Maria'
b. O João enviou uma carta ***à** Lisboa/**para** a Lisboa
the John sent a letter to Lisbon
'John sent a letter to Lisbon'
c. O João deu um livro **à** Maria/***para** a Maria
the John gave a book to Maria
'John gave a book to Maria'
d. O João empurraram/ arrastaram a mesa ***à** Maria/**para** perto da Maria
the John pushed/dragged a table to Maria/to the vicinity of Maria
'John pushed/dragged a table to Maria'

OKAY, IT'S TIME TO WORRY:

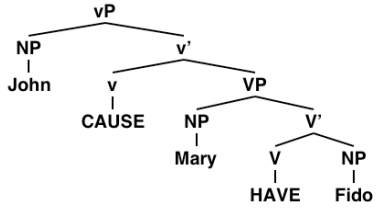
Our chemical model is in trouble. In chemistry, the same atoms in the same proportions can yield different substances (C_2H_6O). This can be true because structure can be different. The same result holds in linguistics: the same words in the same order can yield different expressions. Again, this is because of structure.

BUT in chemistry, different atoms in different structures never yield the same substance. Substances have **unambiguous composition**. There is only one way to make water, only one way to make alcohol, just one way to make nitrous oxide, etc.

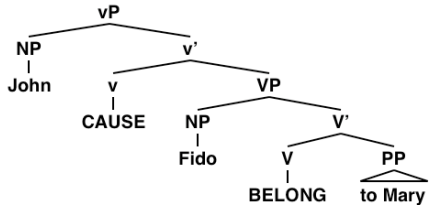
In linguistics, different atoms in different structures seem to yield the same "linguistic substance"– the same interpretation. There are (apparently) two ways to make possession expressions: DOUBLE-OBJECT FORM with *HAVE* and OBLIQUE FORM with *BELONG*.

QUESTIONS:

1. What is the relation between the two caused possession structures? Are these truly independent forms?



John gave Mary Fido
CAUSED-POSSESSION



John gave Fido to Mary
CAUSED-POSSESSION

2. What ties DOUBLE-OBJECT FORM exclusively to CAUSED-POSSESSION? If OBLIQUE FORM can encode both POSSESSION and LOCATION, why can't DOUBLE-OBJECT FORM do so as well? Why can't we have the double object counterparts of locatives?

4.0 Into the Wider (and Wilder) World

Bantu **applicative constructions** exhibit a radical generalization of the English double object form. (19)-(21) are from Chichewa, and closely resemble their English glosses:

(19) a. **APPLIED OBJ** **BASE OBJ**
Mbidzi zi-na-perek-**er**-a nkhandwe msampha
zebras SP-PST-hand-**APP**-ASP fox trap
'The zebras handed the fox the trap'
b. Mbidzi zi-na-perek-a msampha **kwa** nkhandwe
zebras SP-PST-hand-ASP trap **to** fox
'The zebras handed the trap to the fox'

(20) Kalulu a-na-gul-**ir**-a mbidzi nsapato
hare SP-PST-buy-**APP**-ASP zebras shoes
'The hare bought the zebras shoes'

(21) Mavuto a-na-umb-**ir**-a mfumu mtsuko
Mavuto SP-PST-mold-**APP**-ASP chief waterpot
'Mavuto molded the chief the waterpot'

The resemblance is not superficial. The first object in an English double object structure can be passivized, but not relativized (22a,b); the second object shows the opposite behavior (23a,b):

(22) a. John was bought __ those shoes.
b. *The shoes were bought John __.

(23) a. [What John bought Mary __] was those shoes.
b. *[Who John bought __ those shoes] was Mary.

Applicatives show the same pattern. The applied object can be passivized (24a), but not relativized (25a). The base object shows the opposite behavior (24b)/(25b):

(24) a. Mbidzi zi-na-gul-**ir**-idw-a nsapato (ndi kalulu)
zebras SP-PST-buy-**APP**-PASS-ASP shoes by hare
'The zebras were bought shoes (by the hare)'
b. *Nsapato zi-na-gul-**ir**-idw-a mbidzi (ndi kalulu)
shoes SP-PST-buy-**APP**-PASS-ASP zebras by hare
'The shoes were bought for zebras (by the hare)'

(25) a. *Iyi ndiyo mfumu imene ndi-ku-ganiz-a Mavuto a-na-umb-**ir**-a mtsuko
this is chief which 1sS-PRES-think-ASP Mavuto SP-PST-mold-**APP**-ASP waterpot
'This is the chief that I think Mavuto molded the waterpot for'
b. Uwu ndiyo mtsuko umene ndi-ku-ganiz-a Mavuto a-na-umb-**ir**-a mfumu
this is waterpot which 1sS-PRES-think-ASP Mavuto SP-PST-mold-**APP**-ASP chief
'This is the waterpot that I think Mavuto molded for the chief'

Applicatives are important in showing two things: a **semantic range** beyond English (and other European) languages; a **special morphology**.

Semantic Range. English double object structures express <u>only</u> CAUSED-POSSESSION (26)-(27). Applicatives show much broader possibilities (28)-(32).

(26) a. John put the key on the table.
b. *John put the table the key.

(27) a. John baked the cake for Mary.
i. 'John baked the cake for Mary to have.'
ii. 'John baked the cake as a favor to Mary/at her request."
iii. 'John baked the cake in Mary's place, as a substitute for her. '
b. John sang Mary the cake.
i. 'John baked the cake for Mary to have.'
ii. #'John baked the cake as a favor to Mary/at her request."
iii. #'John baked the cake in Mary's place, as a substitute for her. '

(28) **Benefactive/Source/Malefactive**
a. ni θáy -**ɔɬc** -θ -ámʔš -əs ʔə k^{w}θə nə-snéx^{w}əɬ
AUX fix -**APP** -TR -1O -3A OBL Det 1POS-canoe
'He fixed my canoe for me' (Lit: 'He fixed me my canoe')
b. Kambuku a-na-b-**er**-a mkango njinga
leopard SP-PST-steal-**APP**-ASP lion bicycle
'The leopard stole the bicycle from/on the lion'

(29) **Instrumental**
a. Fisi a-na-dul-a chingwe **ndi** mpeni
hyena SP-PST-cut-ASP rope **with** knife
'the hyena cut the rope with the knife'
b. Fisi a-na-dul-**ir**-a mpeni chingwe
hyena SP-PST-cut-**APP**-ASP knife rope
'the hyena cut the rope with the knife'

(30) **Locative**
a. Umwaana y-a-taa-ye igitabo **mu** maazi
child SP-PST-throw-ASP book **in** water
'The child threw the book into the water'
b. Umwaana y-a-taa-ye-**mo** amaazi igitabo
child SP-PST-throw-ASP-**APP** water book
'The child threw the book into the water'

(31) **Manner**
a. Umugabo a-ra-som-a ibaruwa **n'** -iibyiishiimo
man SP-PRES-read-APP-ASP letter **with** joy
'The man is reading a letter with joy'
b. Umugabo a-ra-som-**an**-a ibaruwa ibyiishiimo
man SP-PRES-read-**APP**-ASP letter joy
'The man is reading a letter with joy'

(32) **Reason**
Nsima iyi ndi-ku-dy-**er**-a njala
cornmeal this 1sS-PRES-eat-**APP**-ASP hunger
'I am eating this cornmeal form/because of hunger'

Lesson: <u>Nothing</u> ties double object form exclusively to caused-possession – nothing universal. When we move into a wider linguistic domain, the tie we see in English and other European languages breaks down.

Form. English double object verbs and oblique verbs look the same. But applicative languages show a difference: a **morpheme *APP*** that signals the double object form. This morpheme appears to be present, even when it is not spoken.

(33) a. Ngombe zi-na-tumiz-a mitolo ya udzu **kwa** mbuzi
cows SP-PRES-send-ASP bundles of grass **to** goats
'The cows send bundles of grass to the goats'
b. Ngombe zi-na-tumiz-**ir**-a mbuzi mitolo ya udzu
cows SP-PRES-send-**APP**-ASP goats bundles of grass
'The cows sent the goats bundles of grass'

(34)a. Joni a-na-pats-a nthochi **kwa** amai ake
John SP-PRES-give-ASP bananas **to** mother his
'John gave the bananas to his mother'
b. Joni a-na-pats-a amai ake nthochi **No *APP*!**
John SP-PRES-give-ASP mother his bananas
'John gave his mother the bananas'

It's attractive to view *APP* as a universal aspect of double object form. So English double objects must also have an *APP* morpheme, even if a silent one:

(35) Mary gave-**APP** John a present.

5.0 Where Are We?

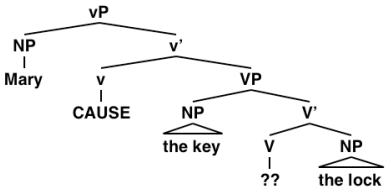
We began with the "one form-one meaning" idea:

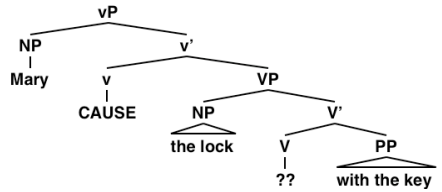
DOUBLE-OBJECT FORM ➔ CAUSED-POSSESSION
OBLIQUE-FORM ➔ CAUSED-MOTION/LOCATION

The idea weakened with *give* – two forms-one meaning. Now it seems to have broken down entirely. Natural language seems to be express the same linguistic substance (meaning) with either of two different forms:

	Applicative (DO)	←??→	**Oblique**
Form:	α V-**APP** β γ		α V γ [**P** β]
Senses:	CAUSED POSSESSION		CAUSED POSSESSION
	BENEFACTIVE/MALEFACTIVE/ SOURCE		BENEFACTIVE/MALEFACTIVE/ SOURCE
	INSTRUMENTAL		INSTRUMENTAL
	CAUSED MOTION/LOCATION		CAUSED MOTION/LOCATION
	MANNER		MANNER
	REASON		REASON

Worse yet, our "deep atom" picture itself is tottering. Consider "instrumental applicatives:





Mary opened-APP the key the lock *Mary opened the lock with the key*

What universal atoms should we put in the "??" positions? We need universal object-instrument relations comparable to *HAVE* and *GO*. What they are is very unclear!

What are the Options?

The "Two-Decompositions" Problem
- Show that the applicative-oblique decompositions don't really yield the same meaning (the same "linguistic substance"). This looks very hard.
- Analyze one of the two structures (Appl-Oblique) as derivative, for example, applicatives as deriving from obliques. This entails **one** basic decomposition.

The "Missing Atoms" Problem
- Look for a different kind of atom, a decomposition not based on abstract predicate atoms like *CAUSE*, *HAVE*, *BECOME*, *DO*, *BE*, *AT*, *GO*, *TO*
- E.g., atoms based on "roles" like *AGENT*, *PATIENT*, *GOAL*, *RECIPIENT*, etc.

Summary
- Chemistry, a brilliant success story in science, offers a potent metaphor for the analysis of other natural domains.
- View objects as composed of **universal atoms** arranged in a **structure**.
- Linguistics has some success with this metaphor: we've shown the need for structure. But what are the universal atoms?
- Deep conceptual atoms like *CAUSE*, *HAVE*, *BECOME*, *DO*, *BE*, *AT*, *GO*, *TO*?
- Promising, but only to a point. English *give* seems to have two atomic decompositions!
- Applicative languages show English *give* to be only "the tip of the iceberg". Applicative-oblique correspondences exhibit many kinds of semantic relations.
- A serious rethinking of basic assumptions seems to be in order.

Thank you!