Arjuna Mahenthiran
Molly Moran
Samantha Richards
Group 1

Annotation Specifications for Annotation of Metaphor

Background

For this project, you will be annotating instances of **verbal metaphor** in sentences from the Penn TreeBank Wall Street Journal Corpus. We will focus on *main verbs* (excluding copula and auxiliary verbs).

Definitions/Theoretical Motivations

What is Verbal Metaphor?

We define verbal metaphor as a *any non-literal usage of a verb.* A non-literal use of a verb, for present purposes, is a *local phenomenon,* arising from some sort of semantic 'mismatch' between the verb and one or more of its of its arguments, which falls outside of the core semantic domain for that verb.

This means that identifying a verbal metaphor is a two-part process. First, the participant(s) (or argument(s)) of the verb must be identified. Then, it must be determined which, if any, of these arguments is 'mismatched' with the verb. Consider the examples below:

(1) Mary whistled
(2) The wind whistled

We might think of the definition for the verb *whistle* as something like 'to intentionally produce a high-pitched sound with the lips or teeth'. This tells us that a literal usage of whistle requires an agentive, animate subject. Since *Mary* is the only participant (a subject) in sight, and she is presumably human, (1) can safely be classified as a literal instance.

In (2), however, we have a semantic mismatch. At minimum, *whistle* requires an agentive subject with a mouth to be used literally: *The wind* fails both of these requirements, as an abstract and inanimate noun. Thus we say that 'The wind' lies *outside of the domain of semantically canonical subject arguments for our verb.*

Example (2) shows a verbal metaphor caused by a mismatch between subject and verb. Recall, though, that verbs may occur with 0 to any number of arguments (though typically in English, no more than 3). Therefore,when a verb with multiple arguments is used in a nonliteral way, any of these arguments might be responsible for triggering the nonliteral semantic interpretation. The culprit might be the subject, direct object, any other type of argument, or a combination of these. The argument(s) responsible for a non-literal interpretation of a given verb is referred to as the *'locus of (semantic) domain incongruence'.*

What is it Not?

Verbal Metaphor is just *one* subtype of metaphor. Other types of metaphor - including adjectival metaphor and noun-noun metaphors - are outside the scope of this project and should *not* be annotated. Therefore, the following SHOULD BE IGNORED:

1. "Her classroom IS a zoo".
    a. Syntactically, in instances like this, "is" along with its conjugations acts to modify a noun, and therefore acts as a noun-noun metaphor, which is outside the scope of this project.
2. "She watched his PRINCELY nap from afar."

a. While the noun “nap” cannot be described literally as “princely”, and therefore suggests figurative language, this example of metaphor involves the pairing of an adjective with a noun, which is outside the scope of this project.

Annotation Scheme

The present annotation scheme will consist of three components:

1. A simple binary annotation of each *verb*, specified for “type” as either as **literal or nonliteral**
    - **Individual tokens** are tagged as verbs
    - For phrasal verbs, choose the main verb - leave particles out.
        - Example: for phrasal verbs like “step up” or “throw out”, *step* and *throw* should be annotated individually.
    - For verb phrases with auxiliaries, tag only the main verb
        - Example: In “He will crumble…”, only the main verb “crumble” should be tagged
2. For nonliteral verbs, the domain incongruent argument(s) will be annotated with the **Arg** tag, specified for “type” as **subject**, **object** or **other.**
    - The *entire noun phrase* should be given the **Arg** tag, not just the head noun.
    - If no such mismatch exists, but the verb does not appear to be used literally, it is likely a case of *figurative language* that does not meet our requirements for a verbal metaphor as presently defined. In this case, *do not tag the verb at all.*
3. Finally, a linking tag between the incongruent Arg and its corresponding Verb of type nonliteral - called **Mismatch.**

Extant Tags

Verb

Required “Type” attribute: Literal | Nonliteral

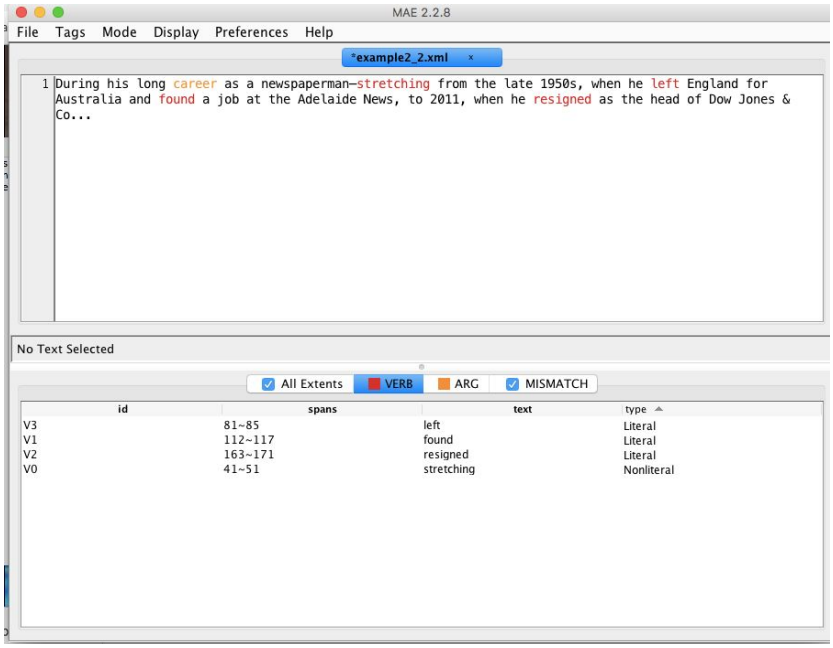
Arg

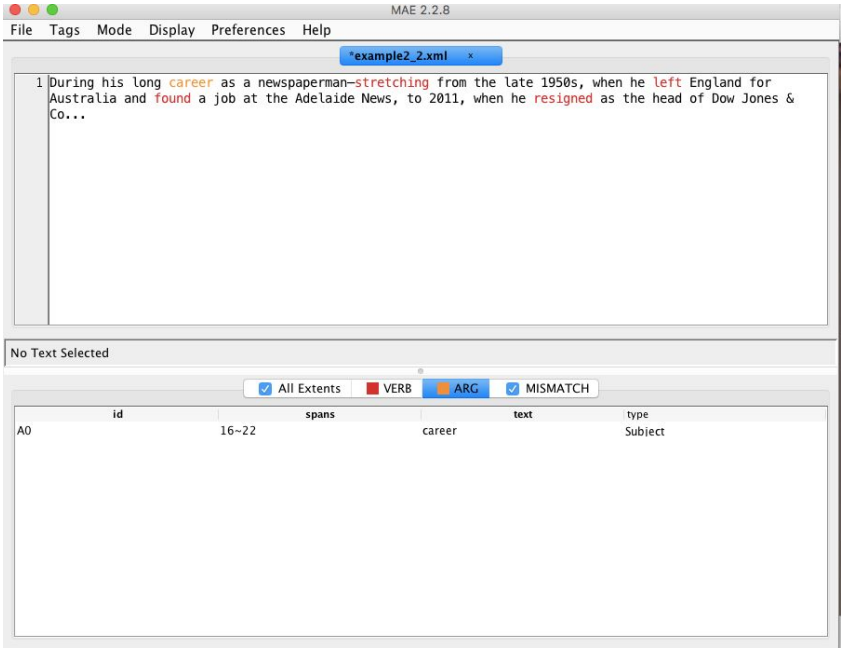
Required “Type” attribute: Subject | Object | Other

Linking Tags

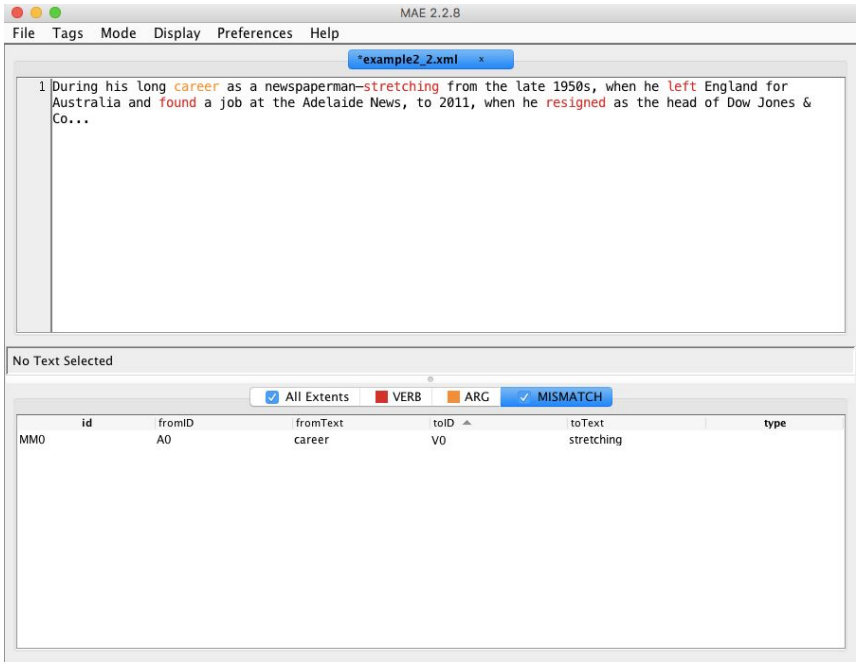
Mismatch

(Arg → Verb (type: Nonliteral))


MAE 2.2.8
File Tags Mode Display Preferences Help
*example2_2.xml
1 During his long career as a newspaperman—stretching from the late 1950s, when he left England for Australia and found a job at the Adelaide News, to 2011, when he resigned as the head of Dow Jones & Co...
No Text Selected
All Extents VERB ARG MISMATCH
id spans text type
V3 81~85 left Literal
V1 112~117 found Literal
V2 163~171 resigned Literal
V0 41~51 stretching Nonliteral


MAE 2.2.8
File Tags Mode Display Preferences Help
*example2_2.xml
1 During his long career as a newspaperman—stretching from the late 1950s, when he left England for Australia and found a job at the Adelaide News, to 2011, when he resigned as the head of Dow Jones & Co...
No Text Selected
All Extents VERB ARG MISMATCH
id spans text type
A0 16~22 career Subject

Arjuna Mahenthiran
Molly Moran
Samantha Richards
Group 1



Part 1. Literal versus Nonliteral Classification

For *each verb* in the dataset, you will determine whether the usage is: (a) literal or (b) nonliteral. Literal verbs will be tagged with the general extant tag "Verb", and further subcategorized by setting the "type" attribute of the verb tag to either "Literal" or "Nonliteral".

Decision procedure

When trying to determine whether a verb's usage is literal or nonliteral (metaphorical), ask the following questions:

1. Is the current usage of the verb the most core, canonical, or concrete that you can think of?
2. If the answer is no to 1: what is the source of the distinction between the current usage of the verb and its most literal possible use? Does it spring from a mapping between two different but comparable domains?
3. If the answer is yes to 2: Does the verb relate in a literal or nonliteral way to its arguments? Are there differences in the kinds of arguments that can be used with the verb in its current nonliteral usage, that would not be available or possible when used with the verb in its literal sense?
    a. Some differences might be:
        i. Concreteness: Does this verb in its core usage only take concrete arguments?
            - Example: 'He smoothed out the problem', where 'smoothed' canonically takes a concrete/physical object
        ii. Animacy: Does this verb in its core usage only take animate arguments?

- Example: 'The wind whistled', where whistled canonically takes an animate/agentive subject

iii. Domain: Does this verb in its core usage take arguments from a specific semantic domain?
- Example: "We hope to cure poverty", where 'cure' canonically takes an object from the biomedical/disease domain

Examples
Here are a few examples of paired literal and nonliteral uses of the same verbs, selected from the WSJ corpus, most of which should represent obvious distinctions between literal and nonliteral usages:

Absorb:
Literal: "The yellow beta carotene pigment **absorbs** blue -LRB- not yellow -RRB- laser light"
Nonliteral:"brokerage firms were **absorbed** by larger , diversified financial services companies"

Cool:
Literal: "other gases might cause a lot of clouds to form, blocking sunlight and **cooling** things off"
Nonliteral: "...there have been some signs that the economy has begun to **cool off**..."

Rain:
Literal: In addition , Mr. Sieck said , NASA would cancel the test if it **rained**
Nonliteral: It's **raining** proposals

Rest:
Literal: Efforts to improve on-board bedding are aimed not at passengers but at pilots and cabin staff , who are supposed to get a chance to **rest** after working a certain number of hours
Nonliteral: Much of this argument **rests** on definitions of control and influence

Smooth:
Literal: For the next eight hours , she will take the pegs one by one and place them in a machine that **smooths** their uneven edges
Nonliteral: Since that time , Kodak has been trying to **smooth** relations with its neighbors , to little avail

Edge Cases

When to annotate literal versus nonliteral may not always be clear. Below are a handful of 'edge cases' representing verbs whose metaphorical status may be perceived as ambiguous, followed with a short description intended to clarify the appropriate label and its reasoning.

Case 1: "Lightning hasn't struck twice: The best of his last three films was the documentary"

Gold: Literal

While the use of the entire clause is nonliteral, we are interested in *local* usage of the verb

'struck'--that is, does it relate in a literal or nonliteral way to its arguments?

Case 2: U.S. Stinger missiles, machine guns and mangled human limbs came **raining** down on what was left

Gold: Nonliteral

(With apologies for the gory example) Although 'rain' here takes a concrete argument, 'human limbs', the verb 'rain' in its literal sense refers to climactic precipitation *only* (as noted already above). This extension to anything that falls piecemeal from above is a metaphorical one.

Case 3: Blithely free of guilt about her licentiousness , Stella makes **sleeping** around with strangers sound like something thoughtful people can do

Gold: Literal

Although this use of language is figurative, this use of 'sleep' is an instance of *metonymy*, not metaphor. Although Stella's action is not one of becoming unconscious for the purposes of resting, neither is the verb's interaction with its arguments ('Stella' and 'strangers') domain-incongruent in any way--the arguments are perfectly compatible with a literal (excluding the preposition 'around') reading of 'sleep'. Since the task is to annotate metaphor, and non metonymy, by default this instance should be annotated as literal.

Part 2. Arguments (Locating Semantic Incongruence)

Our schema identifies three potential argument locations relative to the verb that can lead to a mismatch: subject, direct object, and "other". The subject refers to the canonical subject of the sentence, as commonly thought of in English grammar, and is frequently, though not inherently, the agent of the verb. The direct object refers to the object of the verb, and is necessitated by a transitive verb. Noun phrases in the "other" category are often embedded within freely movable, concatenative prepositional phrases. Finally, in instances where there are mismatches between the verb and arguments in multiple locations, the expectation is that each mismatching argument will be tagged.

Decision Procedure

After determining that the verb is being used in a non-literal manner, use the following questions to determine which noun phrase arguments are causing the mismatch:

1. What is the literal meaning of the verb in the sentence?
2. Is the subject of the sentence one that can be used with the literal meaning of the verb in another real-world context? If not, determine which noun phrase in the subject is contributing the most to the mismatch and mark it.
3. Does the verb take a direct object? Can the literal meaning of the verb be performed on the direct object in another real-world context? If not, determine which noun phrase in the direct object is contributing the most to the mismatch and mark it.
4. If neither the subject or the object is causing the mismatch, identify and check all additional participants of the verb. Does the sentence provide some information (e.g. manner, location, etc., benefaction) about the verb? If so, can it modify the

literal meaning of the verb in another real-world context? If not, determine which noun phrase in the clause is contributing the most to the mismatch and mark it.

Examples

Subject:

- "My common sense **flew** out the door…"
    - The whole noun phrase "common sense" should be tagged as opposed to just the noun "sense"
- "Love shrank and **withered** , and all that remained was a quickening of the pulse , a rise in blood pressure and stomach contractions chasing each other in a horrible loneliness ; the loneliness of the wave absorbed by the sand , which never returns to the sea ./."

Direct Object:

- "The statesmen finally **smoothed** out the project by deciding…"
- He said GE 's strategy of **targeting** specific , lucrative niches in the semiconductor business…"

Other:

- "People who are willing to **step** into the marketplace…"
- "The president is **moving** towards a decision"

Multiple

- "...the market **stumbled** on news of an unexpected rise in U.K. interest rates" (subject & other)
- In early adulthood , she **stuck with** the Democratic Party (subject & object)
- "One look at her pudgy face , shot to make her seem like a powder-brained Kewpie doll , **vaporizes** whatever credibility Jamie 's anguish might have had" (subject & object

Edge Cases

Figurative Language vs. Verbal Metaphor

"He and others also figure Midwesterners are more likely than other citizens to step up and buy houses."

Gold: Literal

This is a case of figurative language that does not fall under our definition of non-literal/metaphorical verb usage. The verb *step* (and the verbal phrase *step up)* is perfectly capable of taking a subject like *citizens* in a literal context. The entire verb phrase is being used figuratively in context of the entire sentence: that is, the sentence means to convey something other than literal stepping. The local usage of the verb, however - the relationship between the verb and its arguments - remains literal, because the argument *citizens* is congruent with *step's* semantic domain.

Idioms

Similarly, you may encounter idioms while you annotate and be unsure how to tag them.

"The old man finally kicked the bucket."

Gold: Literal

Although *kick the bucket* is a great example of figurative language, notice that both the subject - *man* - and object - *bucket* - fall within the core semantic domain for *kick.* Thus this is not an instance of verbal metaphor under our definition, because the verb in question is taking arguments that satisfy a the conditions for a literal usage.

No Clear Argument:

"The Memphis, Tenn., retiree, her blouse bedecked with sequined cards and dice, has just received invitations to two nearby slot tournaments, along with vouchers for $200, all courtesy of Harrah's Entertainment Inc. "Harrah's is savvy," says Ms. Maranees, who admits that once in the casino door, she is bound to spend much more than what Harrah's has given her. **That** is exactly what the Las Vegas-based company is **banking on.**"

Gold: No tag

In this case, we have a referential pronoun as an argument for the phrasal verb *banking on*. From context, we know that *that* refers to 'she is bound to spend much more than what Harrah's has given her.', making this a nonliteral usage of the verb *banking.* However, there is no clear local, explicit argument. In this case, do not tag the verb.

Notice that both examples above involve phrasal verbs. Often, phrasal verbs will prove difficult to classify as literal or non-literal because they carry a different meaning from the main verb involved in them. When in doubt, do not tag the phrasal verb.

Metaphorical Sub-Phrases

Sometimes, a *verb phrase* may take on an overall non-literal meaning, but the mismatch may not involve the verb at all. Consider the below examples:

"A group of history buffs in this border city is intent on seeing that Pancho Villa **rests in peace** rather than in pieces."

Gold: Literal

If we consider the core usage of *rest,* the verb requires, at minimum, a concrete entity with physicality. Pancho Villa, as the subject in this case, satisfies those requirements for a literal

reading. A mismatch might be said to arise, however, with the phrase *in peace,* because resting cannot physically occur "in" an abstract idea. Indeed, *no verb* can occur "in" an abstract idea.

Since the phrase *in peace* is inherently metaphorical, the figurative use of language here is not centered on the verb *rest.* Rather, the idiom *in peace* is a metaphor of its own. Thus this does not constitute an instance of *verbal metaphor,* and should not be tagged as such.

Part 3: Mismatch Linking Tag

Once a nonliteral verb and its incongruent argument are identified, the last step is simple. A "Mismatch" linking tag should be added FROM EACH offending argument TO the nonliteral verb. In other words, some sentences will contain multiple "Mismatch" linking tags from multiple arguments to the same verb.