Annotation Guidelines for Word Sense Structure

1 OVERVIEW

Words have multiple senses (definitions). For example, the word *evening* has three senses:

Senses of *evening*	
evening$_1$	[*eve*, *even*, *eventide*] the latter part of the day (the period of decreasing daylight from late afternoon until nightfall), e.g. "he enjoyed the evening light across the lake"
evening$_2$	a later concluding time period, e.g. "it was the evening of the Roman Empire"
evening$_3$	the early part of night (from dinner until bedtime) spent in a special way, e.g. "an evening at the opera"

The senses of a word have an underlying structure. Our annotation procedure is used to identify the relationships between a word's senses, including which senses are metaphorical. Annotators are shown all of the (noun) senses of a word. They should first read through the senses, and label any that they did not previously know [§2]. The procedure is then as follows:

1. Identify the "**core**" sense(s) of the word, and label it/them. [§3.1]

2. Identify which senses (if any) are "**associated**" with a core sense, and label them. [§3.2]
   For each associated sense:

   - Identify which core sense it connects to, and label this.

3. Identify which senses (if any) have "**metaphorical**" similarities with another sense, and label them. [§3.3]
   For each metaphorical sense:

   - Identify which core <u>or</u> associated sense it connects to, and label this.
   - Identify (and record) features which are preserved and changed in the metaphorical transformation.

Once this procedure is complete, all of a word's senses should be labelled. Occasionally however, a word is unsuitable for this annotation task. In these cases, annotators have three additional operations which they can use to properly treat it:

- In rare cases, a sense might combine a metaphorical sense with a non-metaphorical (core or associated) sense. In these cases, annotators will need to "**split**" the sense in two. [§4.1]

- In rare cases, a sense which is important for the annotation is missing. In these cases, annotators will need to add a "**virtual sense**". [§4.2]

- Our procedure requires that each associated senses be connected to a core sense, and each metaphorical sense be connected to either a core or associated senses. However, in rare cases, an associated sense might need to connect to a sense that is not a core sense, or a metaphorical sense might need to connect to another metaphorical sense. In these cases, annotators have the option to mark a sense as a "**conduit**", to make this possible. [§4.3]

An example of complete annotation, for the word *evening*, is shown overleaf.

Senses of *evening*		Label	Features
evening$_1$	[*eve*, *even*, *eventide*] the latter part of the day (the period of decreasing daylight from late afternoon until nightfall), e.g. "he enjoyed the evening light across the lake"	Core	is the end of the day
evening$_2$	a later concluding time period, e.g. "it was the evening of the Roman Empire"	Metaphorical (from *evening*$_1$)	is the end of any time period
evening$_3$	the early part of night (from dinner until bedtime) spent in a special way, e.g. "an evening at the opera"	Associated (from *evening*$_1$)	

We can also visualise this annotation in graphic form (below). As a convention, metaphors are depicted as vertical extensions, while associations are horizontal. Note that in this notation the features are not shown.
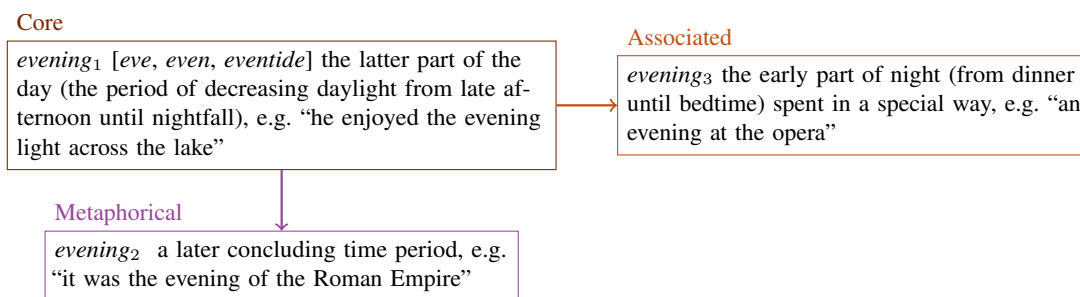
2 READING SENSES AND INDICATING FAMILIARITY

Before labelling a word, annotators should first read through the senses, to understand what each one refers to. Senses have an index (e.g. *leaf*$_1$), and a definition. The definition usually (but not always) begins with a number of synonyms (italicised, in square brackets), followed by a description of the meaning, and ending sometimes in one or more example usages.

Indicating Familiarity Annotators should label any sense which they did not previously know as "unknown". They should also label a word as "unknown" if they do not know the word at all. By default, it will be assumed that they know the word and all of its senses.

Unclear Senses Sometimes, a sense's definition might be unclear, and the sense could refer to different things depending on how it is read. For example, consider this sense of attack:

$$attack_5 \ [attempt] \text{ the act of attacking}$$

This could refer to either a physical attack, or a verbal attack. To understand the definition, it is possible to look recursively at the definitions of words in the definition of the sense you are considering. In this case, that would mean to look at the definition of *attacking*. In the accompanying annotation tool this can be done by hovering the mouse over words in the definition.

3 THREE POSSIBLE LABELS FOR A SENSE

There are three labels that can be given to senses. The labels are "core", "associated", and "metaphorical". Depending on the label given to a sense, annotators will need to supply extra information. This is all described in the following sections.

3.1 Core Senses

A core sense of a word is a "basic" sense of a word. For example, in the *evening* example (above), the core sense refers to the end of the day (*evening*$_1$). A core sense of a word is usually

- the default noun sense that is evoked when the word is heard out of context,

- related to the physical world (but not always),

- historically older (but not always),

- the most likely sense to be referenced using the word on its own, as opposed to in a multi-word phrase.

Every word has at least one core sense. It is common for a word to have exactly one. A word can have multiple core senses, but this only occurs when it has senses that are completely unrelated. All other senses of a word are related to a core sense, either by association (§3.2) or by metaphorical similarity (§3.3).

Selecting the Best Core Sense   Sometimes, there are senses which are closely related, each of which could be a candidate core sense. For groups of related sense like this, only one should be labelled as a core sense. It is up to the annotator to determine which they think is the best. For example, the word *shower* refers to both a showering fixture (*shower*$_1$) and the act of showering (*shower*$_2$). In this case, the latter is the core sense, since it is most likely noun sense to be evoked by the word *shower* without any additional context:

Senses of *shower*		Label
shower$_1$	a plumbing fixture that sprays water over you, e.g. "they installed a shower in the bathroom"	
shower$_2$	[*shower bath*] washing yourself by standing upright under water sprayed from a nozzle, e.g. "he took a shower after the game"	Core

In cases like these, the core sense is usually the sense which is the most likely sense to be referenced using the word on its own, as opposed to in a multi-word phrase. For example, the word *apple* has two senses: it can refer to either a fruit (*apple*$_1$) or a tree which bears that fruit (*apple*$_2$). Most speakers would probably be more likely to say e.g. "I ate the *apple*" than "I ate the *apple fruit*". Inversely, they would probably be less likely to say "I watered the *apple*" than "I watered the *apple tree*". Because of this, the fruit is the core sense:

Senses of *shower*		Label
apple$_1$	fruit with red or yellow or green skin and sweet to tart crisp whitish flesh	Core
apple$_2$	[*orchard apple tree*, *Malus pumila*] native Eurasian tree widely cultivated in many varieties for its firm rounded edible fruits	

While it is possible for a word to have multiple core senses, this only occurs when the same word is used for multiple totally unrelated senses (see below).

Deprioritise History   Sometimes it is quite easy to predict which of a word's senses came first historically. However, annotators should label according to what they believe most people in today's language community would think of as the core sense of a word. This often goes against the real history of a word. For example, the word *train* refers to both a locomotive that runs on rail tracks (*train*$_1$), and a procession of wagons or camels (*train*$_3$). The latter likely precedes the former, but the former is the core sense, because it is what most people today would think of.

Senses of *train*		Label
train$_1$	[*railroad train*] public transport provided by a line of railway cars coupled together and drawn by a locomotive, e.g. "express trains don't stop at Princeton Junction"	Core
train$_3$	[*caravan*, *wagon train*] a procession (of wagons or mules or camels) traveling together in single file, e.g. "they joined the wagon train for safety"	

Multiple Core Senses   Words can have multiple core senses. This is known as homonymy, and it only happens when a word has senses that are completely unrelated. For example, the word *bridge* refers to a crossing over a river ($bridge_1$), a card game ($bridge_5$), and a part of a ship ($bridge_9$). These three senses are completely unrelated to each other, and therefore are all core senses:

Senses of *bridge*		Label
$bridge_1$	[*span*] a structure that allows people or vehicles to cross an obstacle such as a river or canal or railway etc.	Core
$bridge_5$	any of various card games based on whist for four players	Core

In graphical form, they will be represented as disconnected nodes:

Core

$bridge_1$ [*span*] a structure that allows people or vehicles to cross an obstacle such as a river or canal or railway etc.

Core

$bridge_5$ any of various card games based on whist for four players
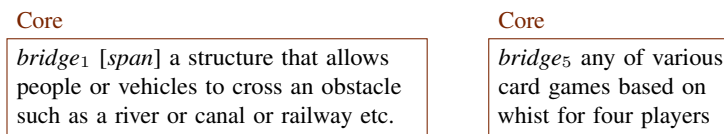
When determining whether two senses are related or not, once again the history of these senses is of secondary importance: what matters primarily is whether the average speaker of English would perceive two senses as related or not.

3.2   Associated Senses

Words often have several senses which are associated to a core sense. An associated sense is a sense which is closely related to a core sense, but not "similar" to it. These associated senses should be labelled as such, and connected to the core sense which they are associated with. Consider the word *neck*. The core sense of *neck* refers to the part of the body connecting the head to the torso ($neck_1$). It has two associated senses: meat from that part of the body ($neck_3$), and a hole in clothing for that part of the body ($neck_5$):
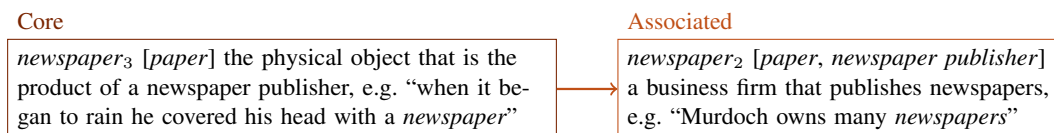
Senses of *neck*		Label
$neck_1$	[*cervix*] the part of an organism (human or animal) that connects the head to the rest of the body, e.g. "he admired her long graceful neck"; "the horse won by a neck"	Core
$neck_2$	a narrow elongated projecting strip of land	
$neck_3$	a cut of meat from the neck of an animal	Associated (from $neck_1$)
$neck_4$	a narrow part of an artifact that resembles a neck in position or form, e.g. "the banjo had a long neck"; "the bottle had a wide neck"	
$neck_5$	[*neck opening*] an opening in a garment for the neck of the wearer; a part of the garment near the wearer's neck	Associated (from $neck_1$)

Note however, that there are two additional senses which are unlabelled. Senses $neck_2$ (long strip of land) and $neck_4$ (narrow part of object) are metaphorically similar to $neck_1$, and so they should not be labelled as associated (these will be treated in §3.3). Senses $neck_3$ and $neck_5$, on the other hand, are not "like" $neck_1$, but they are related, so they are associations.

   There are several different kinds of association. These are systematic associations (§3.2.1), metonymic associations (§3.2.2), restrictive association (§3.2.3), and ad-hoc associations (§3.2.4). Annotators do not need to distinguish between these different kinds. They are detailed in the following sections.
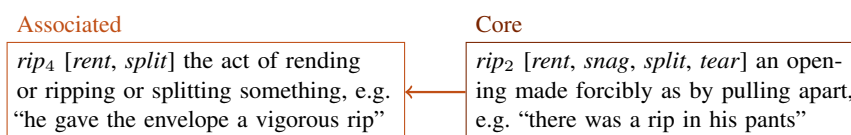
3.2.1 Systematic Associations

Systematic associations are predictable, and affect many words. You would expect new words in a language to develop these kinds of associations. One example is **product–producer alternations**. For example, the word *newspaper* can refer to both a physical newspaper (*newspaper$_3$*), and the company that produces that product (*newspaper$_2$*):

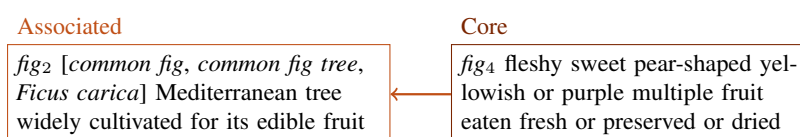| Core	Associated
newspaper$_3$ [*paper*] the physical object that is the product of a newspaper publisher, e.g. "when it began to rain he covered his head with a *newspaper*"	*newspaper$_2$* [*paper*, *newspaper publisher*] a business firm that publishes newspapers, e.g. "Murdoch owns many *newspapers*"

In the case of *newspaper*, the sense describing the product (*newspaper$_3$*) is the core sense. However, there are cases where the core sense refers to a producer, and their product is an association (e.g. *Honda*). With systematic associations, the annotator needs to determine which is the core sense and which is the association. Below is a (non-exhaustive) list of more examples of regular association:
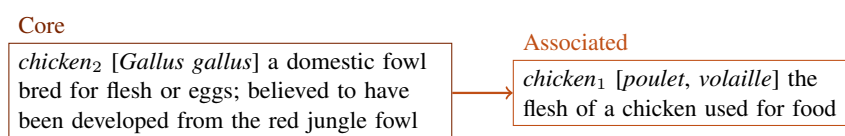
- **Process–result alternations**. Consider the word *rip*, which refers to both the act of creating such a rip in something (*rip$_4$*), and the physical rip itself (*rip$_2$*):
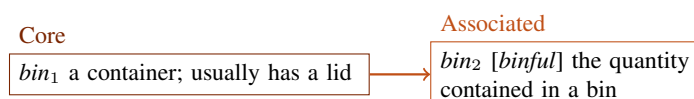
Associated	Core
rip$_4$ [*rent*, *split*] the act of rending or ripping or splitting something, e.g. "he gave the envelope a vigorous rip"	*rip$_2$* [*rent*, *snag*, *split*, *tear*] an opening made forcibly as by pulling apart, e.g. "there was a rip in his pants"

- **Plant–food alternations**. Consider the word *fig*, which refers to both a type of tree (*fig$_2$*), and the fruit which this tree bears (*fig$_4$*):

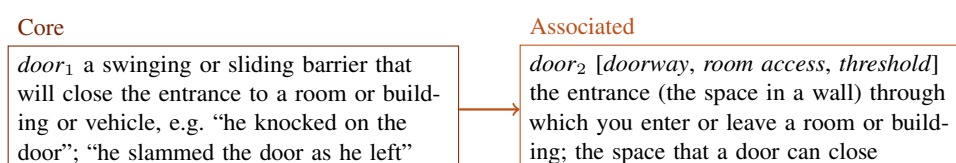| Associated	Core
fig$_2$ [*common fig*, *common fig tree*, *Ficus carica*] Mediterranean tree widely cultivated for its edible fruit	*fig$_4$* fleshy sweet pear-shaped yellowish or purple multiple fruit eaten fresh or preserved or dried

- **Count–mass alternations**. Consider the word *chicken*, which refers to both an animal in a farm (*chicken$_2$*), and the meat from this animal (*chicken$_1$*). The former is a countable noun (e.g. you can have "two chickens"), while the latter is a mass noun (e.g. you can have "some chicken"):

Core	Associated
chicken$_2$ [*Gallus gallus*] a domestic fowl bred for flesh or eggs; believed to have been developed from the red jungle fowl	*chicken$_1$* [*poulet*, *volaille*] the flesh of a chicken used for food

- **Container–containee alternations**. For example, the word *bin* refers to both a container for waste (*bin$_1$*), and the amount held by such a container (*bin$_2$*):

Core	Associated
bin$_1$ a container; usually has a lid	*bin$_2$* [*binful*] the quantity contained in a bin

- **Figure–ground reversals**. For example, the word *door* refers to both a rectangular piece of wood with a handle (*door$_1$*), and the hole which this object seals (*door$_2$*), which is created in its absence:

Core	Associated
door$_1$ a swinging or sliding barrier that will close the entrance to a room or building or vehicle, e.g. "he knocked on the door"; "he slammed the door as he left"	*door$_2$* [*doorway*, *room access*, *threshold*] the entrance (the space in a wall) through which you enter or leave a room or building; the space that a door can close

3.2.2 Metonymic Associations

Metonymy occurs when a concept is given the name of something associated with that concept. Metonymic senses should be labelled as associations. There are two common types of metonymy, which are described below.

Part–Whole Metonymy One common type of metonymy is when a word describing a part of an object is used to refer to the object itself. For example, the core sense of the word *wheel* is a circular object that is part of a vehicle (*wheel$_1$*). The word *wheel* also has a metonymic association, which refers to a whole bicycle (*wheel$_7$*)[1]:
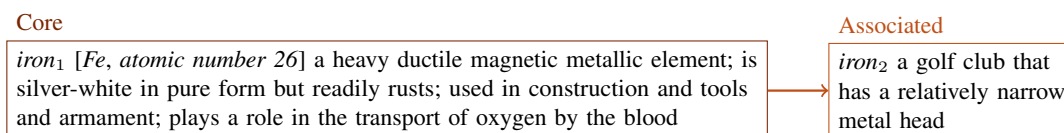
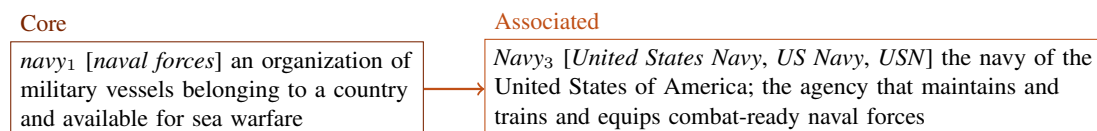| Core	Associated
wheel$_1$ a simple machine consisting of a circular frame with spokes (or a solid disc) that can rotate on a shaft or axle (as in vehicles or other machines)	*wheel$_7$* [*bicycle*, *bike*, *cycle*] a wheeled vehicle that has two wheels and is moved by foot pedals
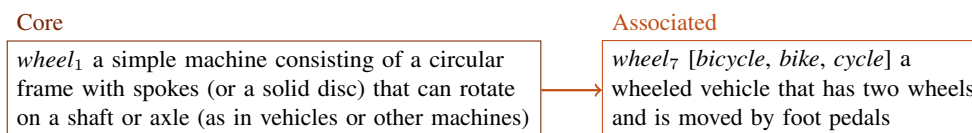
Property–Object Metonymy Another common type of metonymy is when a word describing a property of an object (e.g. its material) is used to refer to the object itself. For example, the core sense of the word *iron* is a type of metal (*iron$_1$*). The word *iron* also has numerous metonymic associations, one of which refers to golf club made of that material (*iron$_2$*):

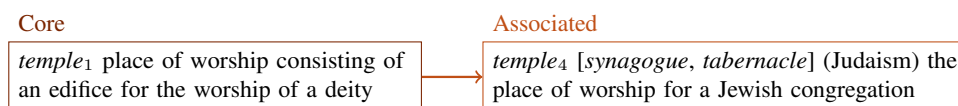| Core	Associated
iron$_1$ [*Fe, atomic number 26*] a heavy ductile magnetic metallic element; is silver-white in pure form but readily rusts; used in construction and tools and armament; plays a role in the transport of oxygen by the blood	*iron$_2$* a golf club that has a relatively narrow metal head

This association is not necessarily reversible: if the core sense of the word was the object (rather than the property), then the property sense may be treated as metaphor (see "abstractions" in §3.4).

3.2.3 Restrictive Associations

Sometimes, a word has a sense which is more specific than the core sense. This commonly occurs with naming. For example, the core sense of the word *navy* is the sea-faring wing of a military force (*navy$_1$*). However, the word is also used specifically to refer to the US Navy (*Navy$_3$*), which is a restrictive association:

Core	Associated
navy$_1$ [*naval forces*] an organization of military vessels belonging to a country and available for sea warfare	*Navy$_3$* [*United States Navy*, *US Navy*, *USN*] the navy of the United States of America; the agency that maintains and trains and equips combat-ready naval forces

This also occurs with examples that are not naming. For example, the core sense of *temple* refers to a place of worship (*temple$_1$*), but it is also used to refer to specifically Jewish places of worship (*temple$_4$*):

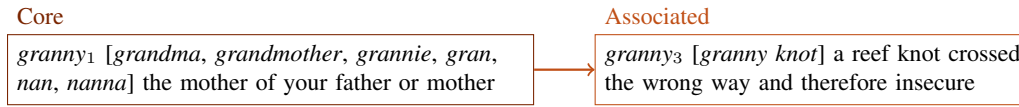| Core	Associated
temple$_1$ place of worship consisting of an edifice for the worship of a deity	*temple$_4$* [*synagogue*, *tabernacle*] (Judaism) the place of worship for a Jewish congregation

This association is not necessarily reversible. In both of the previous examples, the core sense of the word was the more general sense. However, if the specific sense was the core sense, then the more general sense will **not** be an association: it will be a metaphor (see "generaliations" in §3.4).

---

[1]This is more likely to be realised in plural form, as in e.g. "check out my new wheels", which refers to a vehicle as a whole, as opposed to just its wheels.

3.2.4 Ad-hoc Associations

Ad-hoc associations are unpredictable, and each one is unique. Were these senses to be forgotten, it is unlikely that they would re-emerge. For example, the word *granny* refers to both an elderly woman who has grandchildren ($granny_1$), but is also used in sailing to refer to a type of knot that is the sort that would be tied by a granny ($granny_3$). This second sense is associated with the first, but not in any usual systematic way.

Core

$granny_1$ [*grandma, grandmother, grannie, gran, nan, nanna*] the mother of your father or mother

→

Associated

$granny_3$ [*granny knot*] a reef knot crossed the wrong way and therefore insecure

3.3 Metaphorical Senses

A metaphorical sense is a sense which resembles a core or associated (non-metaphorical) sense, often in an abstract way. In other words, a metaphorical sense is "like" one of the non-metaphorical senses. Senses like this should be labelled as metaphors, and connected to the non-metaphorical sense they relate to to. For example, the core sense of the *leaf* refers to a part of a plant ($leaf_1$). The word is also used to refer to a page of a book ($leaf_2$). This second sense resembles the first sense, so it is metaphorical. This example is shown below, in both tabular and graphic forms:

Senses of *leaf*		Label
$leaf_1$	[*leafage, foliage*] the main organ of photosynthesis and transpiration in higher plants	Core
$leaf_2$	[*folio*] a sheet of any written or printed material (especially in a manuscript or book)	Metaphorical (from $leaf_1$)

Core

$leaf_1$ [*leafage, foliage*] the main organ of photosynthesis and transpiration in higher plants

Metaphorical

$leaf_2$ [*folio*] a sheet of any written or printed material (especially in a manuscript or book)

Each metaphorical sense is similar to the non-metaphorical sense it connects to in some respects, but different in others. The second sense of *leaf* has some features in common (e.g. they are both flat), but it is missing an important feature (it is not part of a plant). Annotators additionally need to record this feature transformation (§3.3.1).

3.3.1 Feature Transformations

For each metaphorical sense, annotators need to record the feature transformation underlying the metaphor. This is done using the following procedure:

1. Add features to the core or associated sense which the metaphorical sense connects to. Features are plain text, and fill the gap "This thing ___". For example, the core sense of *bird* might have the features "has wings", "is an animal", and "lays eggs" (all are valid ends to the previously mentioned sentence). Good features are briefly worded, and not too broad nor too specific:

Senses of *leaf*		Label	Features
$leaf_1$	[*leafage, foliage*] the main organ of photosynthesis and transpiration in higher plants	Core	is flat enables photosynthesis is part of a plant
$leaf_2$	[*folio*] a sheet of any written or printed material (especially in a manuscript or book)	Metaphorical (from $leaf_1$)	

7

2. For each feature, decide whether the metaphorical sense:

   (a) also has the feature ("kept")

   (b) does not necessarily have the feature ("lost")

   (c) transforms the feature ("modified")

   For features labelled as modified, edit them so they fit the metaphorical sense. Features should be designed to require few edits to capture differences (e.g. single word substitutions).

Thus, for the *leaf* example, a complete annotation (with one kept feature, one lost feature, and one modified feature) could be:

Senses of *leaf*		Label	Features
leaf$_1$	[*leafage*, *foliage*] the main organ of photosynthesis and transpiration in higher plants	Core	is flat enables photosynthesis is part of a plant
leaf$_2$	[*folio*] a sheet of any written or printed material (especially in a manuscript or book)	Metaphorical (from *leaf*$_1$)	is flat ~~enables photosynthesis~~ is part of a book

Every metaphorical sense must have **either**

- at least one kept feature and one lost feature, or

- at least one modified feature.

This is a minimum requirement; senses can have more features that this. However, annotators should not attempt to create features which capture all of the meaning of a sense. Instead, they should try to create the minimal set of informative feature transformations, that (1) capture both the similarity and difference of a metaphor, and (2) tease apart the differences between multiple metaphorical senses (see §3.4). Often, a modified feature can be expressed as one kept feature and one lost feature, or vice versa; annotators should use whatever feature transformations they think best meet these criteria.

3.4 Additional Details and Special Cases

The Threshold for Metaphoricity   As mentioned above, a metaphorical sense is similar to a core or associated sense in some respects, but different in others. It might sometimes be difficult to determine whether two senses are sufficiently different to merit labelling one as a metaphor. In general, in cases like these, annotators should be generous about labelling metaphors, and should design features which are specific about the distinctions. For example, the word *flesh* can refer to either the flesh of an animal (*flesh*$_1$), or the flesh of fruit (*flesh*$_2$). Annotators should label subtle distinctions like this as metaphorical, and design features to highlight the similarities and differences:

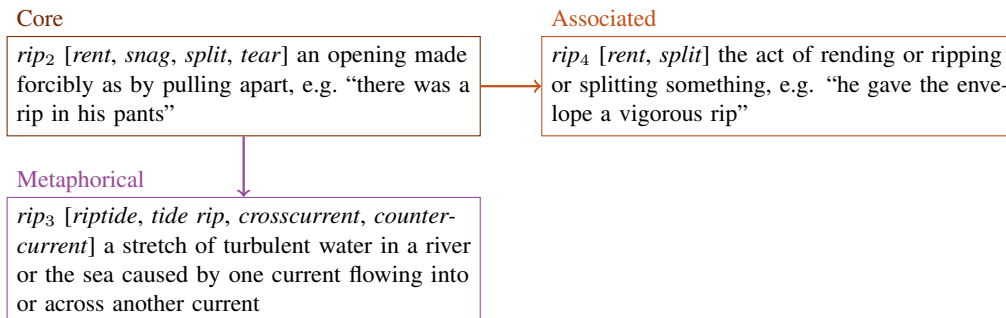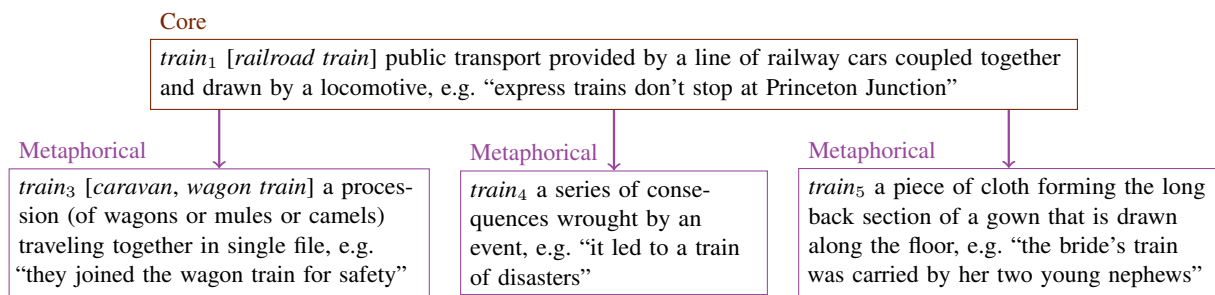| Senses of *flesh*		Label	Features	
flesh$_1$	the soft tissue of the body of a vertebrate: mainly muscle tissue and fat	Core	is soft tissue is edible needs to be cooked is moist is part of an animal	Core *flesh*$_1$ the soft tissue of the body of a vertebrate: mainly muscle tissue and fat
flesh$_2$	[*pulp*] a soft moist part of a fruit	Metaphorical (from *flesh*$_1$)	is soft tissue is edible ~~needs to be cooked~~ is moist is part of a fruit	Metaphorical *flesh*$_2$ [*pulp*] a soft moist part of a fruit

Choosing the Sense a Metaphor is Similar To   Often there are several related senses which a metaphorical sense could connect to. Connecting it to the correct sense requires annotators to consider the semantic type of the senses. An event should connect to an event, an object should connect to an object, an action should connect to an action, and so on. Consider the word *rip*, which can refer to an opening made by ripping (rip_2), the act of ripping (rip_4), or a riptide (rip_3):

Core
rip_2 [*rent*, *snag*, *split*, *tear*] an opening made forcibly as by pulling apart, e.g. "there was a rip in his pants"

Associated
rip_4 [*rent*, *split*] the act of rending or ripping or splitting something, e.g. "he gave the envelope a vigorous rip"

Metaphorical
rip_3 [*riptide*, *tide rip*, *crosscurrent*, *countercurrent*] a stretch of turbulent water in a river or the sea caused by one current flowing into or across another current

Here, there are two non-metaphorical senses (rip_2 and rip_4). The third sense (rip_3) is metaphorical, and could connect to either of the other senses. However, it connects to rip_2, because both refer to the rip itself (rather than the act of creating a rip, as in rip_1).

Multiple Metaphorical Extensions   Sometimes, multiple metaphorical senses all connect to the same non-metaphorical sense. Consider the word *train*, this time with additional senses shown:

Core
$train_1$ [*railroad train*] public transport provided by a line of railway cars coupled together and drawn by a locomotive, e.g. "express trains don't stop at Princeton Junction"

Metaphorical
$train_3$ [*caravan*, *wagon train*] a procession (of wagons or mules or camels) traveling together in single file, e.g. "they joined the wagon train for safety"

Metaphorical
$train_4$ a series of consequences wrought by an event, e.g. "it led to a train of disasters"

Metaphorical
$train_5$ a piece of cloth forming the long back section of a gown that is drawn along the floor, e.g. "the bride's train was carried by her two young nephews"

Each of these metaphors transforms the meaning of $train_1$ in different ways. Each will modify the same set of features from $train_1$; annotators should therefore design features for $train_1$ which will capture these differences. Adding more features can be helpful:

Senses of *train*		Label	Features
$train_1$	[*railroad train*] public transport provided by a line of railway cars coupled together and drawn by a locomotive, e.g. "express trains don't stop at Princeton Junction"	Core	is pulled from the front is a sequence of carriages runs on a railway is a means of transit
$train_3$	[*caravan*, *wagon train*] a procession (of wagons or mules or camels) traveling together in single file, e.g. "they joined the wagon train for safety"	Metaphorical (from $train_1$)	~~is pulled from the front~~ is a sequence of wagons ~~runs on a railway~~ is a means of transit
$train_4$	a series of consequences wrought by an event, e.g. "it led to a train of disasters"	Metaphorical (from $train_1$)	~~is pulled from the front~~ is a sequence of consequences ~~runs on a railway~~ ~~is a means of transit~~
$train_5$	a piece of cloth forming the long back section of a gown that is drawn along the floor, e.g. "the bride's train was carried by her two young nephews"	Metaphorical (from $train_1$)	is pulled from the front ~~is a sequence of carriages~~ ~~runs on a railway~~ ~~is a means of transit~~

Each of the metaphorical senses transforms the features in different ways:

- The sense $train_4$ (consequences) is a sequence of things, similar to how $train_1$ is comprised of a sequence of carriages. The second feature captures this.

- Like $train_4$, the sense $train_3$ (wagons) is also a sequence of things, but it has more in common with $train_1$, because it is also a means of transit. The fourth feature thus distinguishes this metaphor from $train_4$.

- The sense $train_5$ (gown) has a very different metaphorical similarity: it is pulled from the front by the bride, like how $train_1$ is drawn along from the front by an engine carriage. The first feature captures this.

Generalisations   Sometimes, one of a word's senses is a generalisation of another. If a generalisation covers cases which are clearly metaphorical, it should be treated as a metaphor. For example, the word *granny* refers primarily to women with grandchildren ($granny_1$), but it can also be used to refer to any old women, including those that do not have grandchildren ($granny_2$). This case is clearly metaphorical:

Senses of *granny*		Label	Features
$granny_1$	[*grandma*, *grandmother*, *grannie*, *gran*, *nan*, *nanna*] the mother of your father or mother	Core	is an elderly woman has grandchildren
$granny_2$	an old woman	Metaphorical (from $granny_1$)	is an elderly woman ~~has grandchildren~~

Generalisations can also apply to proper nouns. For example, the core sense of *Shylock* is a Shakespearean character, known for his merciless money-lending ($Shylock_2$). This word is now used metaphorically, to refer to anyone who does this ($shylock_1$):

Senses of *shylock*		Label	Features
$shylock_1$	[*usurer*, *loan shark*, *moneylender*] someone who lends money at excessive rates of interest	Metaphorical (from $Shylock_2$)	lends money at high interest ~~is a Shakespearean character~~
$Shylock_2$	a merciless usurer in a play by Shakespeare	Core	lends money at high interest is a Shakespearean character

If the general sense was the core sense, then the more specific sense would be a restrictive association, not a metaphor (see §3.2.3).

Abstractions   Abstractions are extensions from a sense that refers to an object, to a sense that describes a property of that object. A common example is colour names that are derived from real-world referents. For example, the word *amber* refers to both fossil resin, and the colour of the fossil resin:

Senses of *amber*		Label	Features
$amber_1$	[*gold*] a deep yellow color, e.g. "an amber light illuminated the room"	Metaphorical (from $amber_2$)	is a deep yellow colour ~~is fossil resin~~
$amber_2$	a hard yellowish to brownish translucent fossil resin; used for jewelry	Core	is a deep yellow colour is fossil resin

In cases like these, the sense referring to the abstract property ($amber_1$) can be used to describe things that are not the object it derived from ($amber_2$). For example, $amber_1$ can be used to describe eyes (e.g. "his eyes were amber"); this is clearly metaphorical. In cases like these, it can be helpful to construct a simile (e.g. "his eyes were *like* amber") to see the metaphorical similarity—they are both the same yellowish colour, but they are not both resin.

If the core sense was the property sense, then the object sense will be a metonymic association, not a metaphor (see §3.2.2).

Distinguishing Between Associations and Metaphors   The examples of generalisation and abstraction may feel to some annotators like they are better treated as instances of association, not metaphor. Likewise, some of the examples of association (§3.2) may feel better treated as metaphor. A good way to distinguish between the two is using the features. If it is easy to express a feature transformation between one sense and another, it should be treated as a metaphorical sense. Otherwise, it should be treated as an associated sense.

4   ADDITIONAL OPERATIONS

In rare cases, the senses of a word will be unsuitable for this annotation task. In these cases, annotators have additional operations they can use to make a word properly treatable. There are three operations they can use: splitting a sense in two, adding a new "virtual" sense, or adding an additional "conduit" tag to a sense. Each of these operations should be used sparingly, only when necessary.

4.1   Splitting Senses

Sometimes, a sense conflates a metaphorical sense with a non-metaphorical sense. This is an error of the dictionary we use. Consider this sense of the word *birth*:

Senses of *birth*
$birth_1$ the time when something begins (especially life), e.g. "they divorced after the birth of the child"; "his election signaled the birth of a new age"

From its definition, it is clear that $birth_1$ covers both a metaphorical sense of *birth* ("the *birth* of a new age"), and a non-metaphorical sense ("the *birth* of the child"). In this case, you should "split" the sense into two separate senses, one corresponding to its non-metaphorical component, and the other to its metaphorical component. You should also edit the definition in each case to make the distinction clear. Each half should then be treated in isolation as separate senses, following the standard procedure:

Senses of *birth*		Label
$birth_{1A}$	the time when life begins, e.g. "they divorced after the birth of the child"	Core
$birth_{1B}$	the time when something begins other than life, e.g. "his election signaled the birth of a new age"	Metaphorical (from $birth_{1A}$)

The metaphorical half, $birth_{1B}$, is necessarily labelled as a metaphor. The non-metaphorical half, $birth_{1A}$, can either be a core sense or an associated sense (this will depend on the other senses of the word, which are not shown here). Splitting should be done sparingly, only in cases where a sense unambiguously conflates metaphorical and non-metaphorical senses.

4.2   Virtual Senses

Sometimes, the sense which a metaphorical or associated sense connects to might be missing. In these cases, annotators have the option of defining an additional "virtual sense". Virtual senses are new senses, with a definition provided by the annotator, which are either core or associated. Virtual senses must be the same part of speech as the other senses (i.e. they must be nouns).

Virtual Associated Senses   Sometimes, it will be clear that a metaphorical extension is present, but it cannot be expressed because a sense needed is missing. Consider the word *bee*. A *bee* is an insect, known for their communal activities (bee_1). This is the core sense of *bee*. A *bee* is also a social gathering (of people) to do an activity together (bee_2):

Senses of *bee*		Label
bee_1	any of numerous hairy-bodied insects including social and solitary species	Core
bee_2	a social gathering to carry out some communal task or to hold competitions	

The second sense relates metaphorically to bee_1, but not directly: it is not metaphorically similar to the insect, but to the communal work which the insect is known for.[2] To express this, it is necessary to add a virtual sense, bee_3:

| Senses of *bee*		Label
bee_1	any of numerous hairy-bodied insects including social and solitary species	Core
bee_2	a social gathering to carry out some communal task or to hold competitions	Metaphorical (from bee_3)
bee_3 (virtual)	a gathering of bees working on some communal task	Associated (from bee_1)

Core
bee_1 any of numerous hairy-bodied insects including social and solitary species

Associated
bee_3 a gathering of bees working on some communal task

Metaphorical
bee_2 a social gathering to carry out some communal task or to hold competitions

An example of a valid feature transformation for this metaphorical extension might be:

$$\begin{array}{cc} \textbf{Features of } bee_3 & \textbf{Features of } bee_2 \\ \text{is a cooperative group of bees} \quad \Rightarrow & \text{is a cooperative group of people} \end{array}$$

Virtual Core Senses Virtual senses can also add core senses, if the most basic sense is missing. For example, consider the word *road*. The core sense of *road* is a surface, usually made of tarmac, which cars drive on to move between places. This sense is missing:

Senses of *road*		Label
$road_1$	[*route*] an open way (generally public) for travel or transportation	
$road_2$	a way or means to achieve something, e.g. "the road to fame"	Metaphorical (from $road_1$)

The second sense of *road* is a metaphorical extension of the first, but the first is not the core sense. To express this, it is necessary to add a virtual sense:

Core
$road_3$ a surface, usually made of tarmac, which cars drive on to move between places, e.g. "the road was hot without shoes", "get off the road!"

Associated
$road_1$ [*route*] an open way (generally public) for travel or transportation

Metaphorical
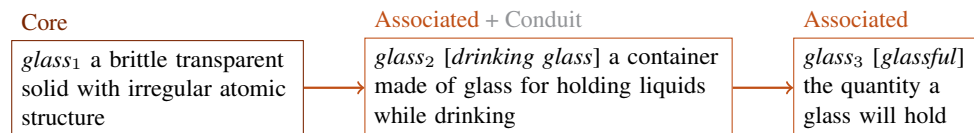$road_2$ a way or means to achieve something, e.g. "the road to fame"
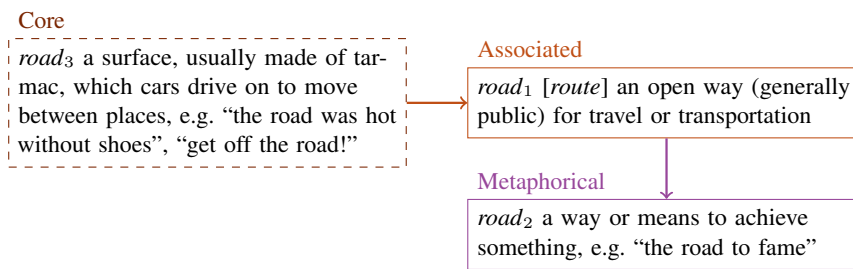
4.3 Conduit Senses

Our annotation procedure requires that each associated sense is connected to a core sense, and each metaphorical sense is connected to either a core sense or an associated sense. However, there are cases where this requirement needs to be violated. To treat these cases, it is possible to label a metaphorical or associated sense as a "conduit". A sense which is labelled as a conduit functions like a core sense, which means that other metaphors or associations can connect to it. This is necessary in three cases:

---

[2]Historically speaking, these two senses exhibit homonymy (they are unrelated). Nevertheless, the sense bee_2 is commonly reanalysed to be metaphorically related to bee_1.
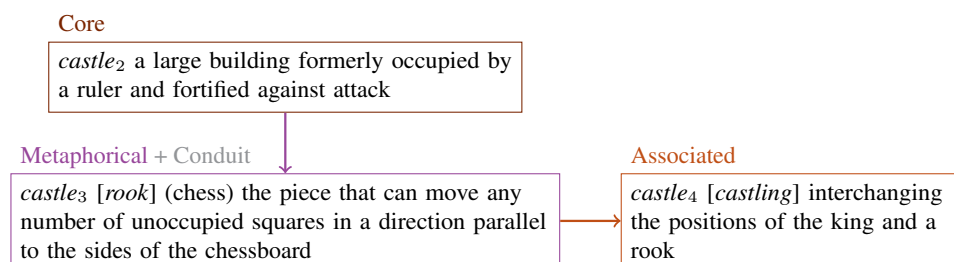
- Sometimes, an associated sense extends another associated sense, rather than a core sense. Consider the word *glass*. The core sense of *glass* is the transparent material ($glass_1$). The word *glass* is also used to refer to a cup made of glass ($glass_2$); this sense is a metonymic association of $glass_1$. The word *glass* is also used to refer to the contents of such a cup ($glass_3$). Nothing directly links $glass_1$ to $glass_3$; instead, this sense is connected to $glass_2$ by a systematic container–containee association. The middle connecting sense ($glass_2$) should be labelled as a conduit, to allow $glass_3$ to connect to it:

Core	Associated + Conduit	Associated
$glass_1$ a brittle transparent solid with irregular atomic structure	$glass_2$ [*drinking glass*] a container made of glass for holding liquids while drinking	$glass_3$ [*glassful*] the quantity a glass will hold

- Sometimes, an associated sense extends a metaphorical sense, rather than a core sense. Consider the word *castle*. The core sense of *castle* is a fort occupied by a ruler ($castle_2$). The word *castle* is also used to refer to a piece in chess ($castle_3$), which visually resembles a fort; this is metaphorically similar to $castle_2$. An example feature transformation might be:

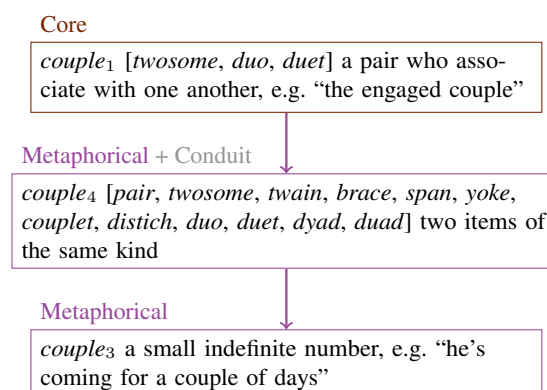| **Features of *castle*$_2$**		**Features of *castle*$_3$**
has crenellations	$\Rightarrow$	has crenellations
is a building	$\Rightarrow$	~~is a building~~

This sense is then associated with a final sense of *castle*, which describes a move in chess involving this piece ($castle_4$). Nothing directly links this sense to $castle_2$. The middle connecting sense ($castle_3$) should be labelled as a conduit, to allow $castle_4$ to connect to it:

Core

$castle_2$ a large building formerly occupied by a ruler and fortified against attack

Metaphorical + Conduit

$castle_3$ [*rook*] (chess) the piece that can move any number of unoccupied squares in a direction parallel to the sides of the chessboard

Associated

$castle_4$ [*castling*] interchanging the positions of the king and a rook

- Sometimes, a metaphorical sense extends another metaphorical sense, rather than a core or associated sense. For example, consider the word *couple*. The core sense of *couple* refers to two people ($couple_1$). A loosening of constraints extends *couple* metaphorically to refer to two of anything ($couple_4$). This sense, in turn, extends further (in US English) to refer a small number of anything ($couple_3$). This is a two-stage feature transformation:

Features of *couple*$_1$		**Features of *couple*$_4$**		**Features of *couple*$_3$**
is two people	$\Rightarrow$	is two of anything	$\Rightarrow$	is a small number of anything

Nothing directly links $couple_1$ to $couple_3$. The middle connecting sense ($couple_4$) should be labelled as a conduit, to allow $couple_3$ to connect to it:

Core

$couple_1$ [*twosome, duo, duet*] a pair who associate with one another, e.g. "the engaged couple"

Metaphorical + Conduit

$couple_4$ [*pair, twosome, twain, brace, span, yoke, couplet, distich, duo, duet, dyad, duad*] two items of the same kind

Metaphorical

$couple_3$ a small indefinite number, e.g. "he's coming for a couple of days"

With conduits, it is theoretically possible to connect any sense to any other sense, in an unconstrained fashion. However, this ability should be used incredibly sparingly, **only when no direct connection to the core is possible**. This is particularly important in the case of metaphorical extensions on other metaphorical extensions, which can usually be explained by linking back to the core sense. A conduit should only be added if there is no way to formulate a direct link. Additionally, it is not permitted to use conduits to create loops (e.g. you cannot label two senses as conduits and connect them to each other).