
Annotation Guidelines for Word Sense Structure

1 OVERVIEW

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

Senses of <i>evening</i>	
<i>evening</i> ₁	[<i>eve, even, eventide</i>] 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”
<i>evening</i> ₂	a later concluding time period, e.g. “it was the evening of the Roman Empire”
<i>evening</i> 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 relations between a word’s senses, including which senses are metaphorical. Annotators are shown all the senses of a word. They should first read through the senses, and flag any that they did not previously know [§2]. The procedure is then as follows:

1. Determine which senses are the “core” senses of the word, and label them. [§3.1]
2. Determine which senses (if any) are connected to another sense via “association”, and label them. [§3.2]
For each associated sense:
 - Determine which core sense it connects to, and label this.
 - Determine whether the association is regular or irregular, and label this.
3. Determine which senses (if any) are connected to another sense via “metaphor”, and label them. [§3.3]
For each metaphorical sense:
 - Determine which core or associated sense it connects to, and label this.
 - Identify (and record) features which are preserved and changed in the metaphorical transformation.

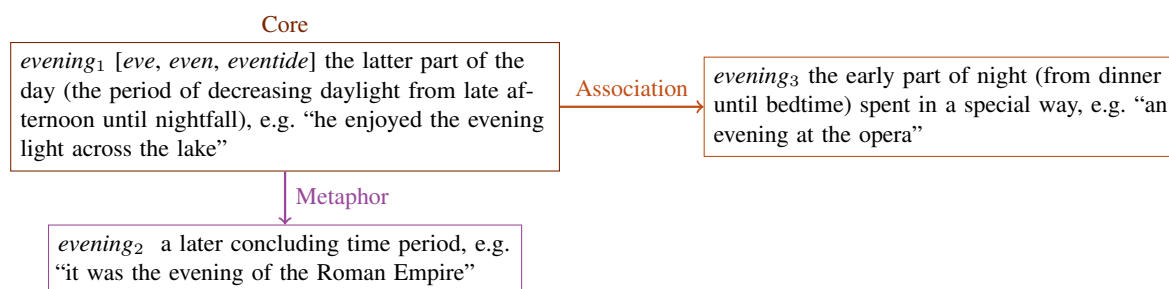
Once this procedure is complete, all of a word’s senses should be labelled. Occasionally, a word is unsuitable for this annotation task. In these cases, annotators have three additional options which they can use to properly treat it:

- In rare cases, a sense might combine a metaphorical sense with a non-metaphorical (core/associated) sense. In these cases, annotators will need to “split” the sense in two. [§4.1]
- In rare cases, the sense to which a metaphorical sense should connect is missing. In these cases, annotators will need to add a “virtual sense”. [§4.2]
- In rare cases, an associated sense might connect to a metaphorical sense or another associated sense (rather than a core sense), or a metaphorical sense might connect to another metaphorical sense (rather than a core/associated sense). In these cases, annotators have the option to mark the sense it connects to as a “secondary core”. [§4.3]

An example of complete annotation is shown overleaf.

Senses of <i>evening</i>	Label	Features
<i>evening</i> ₁ [<i>eve, even, eventide</i>] 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
<i>evening</i> ₂ a later concluding time period, e.g. “it was the evening of the Roman Empire”	Metaphor (from <i>evening</i> ₁)	is the end of any time period
<i>evening</i> ₃ the early part of night (from dinner until bedtime) spent in a special way, e.g. “an evening at the opera”	Association (from <i>evening</i> ₁)	

We can also visualise this annotation in graphic form (below). Note however 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. Annotators should indicate if there is a sense which they do not know. They should also indicate 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

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

*attack*₅ [*attempt*] the act of attacking

This could refer to either a physical attack, or a verbal attack. To disambiguate the definition, it is possible to look 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 mousing over words in the definition.

Broad Senses The wording of a sense’s definitions can sometimes be very broad and all-encompassing. For example, consider these two senses of *granny*:

*granny*₁ [*grandma, grandmother, grannie, gran, nan, nanna*] the mother of your father or mother

*granny*₂ an old woman

The second sense, *granny*₂, is very broad. Based on its wording, it could include any elderly woman, including those covered by *granny*₁. However, annotators should assume that all senses are mutually exclusive, and do not overlap: a sense with a broad definition would not cover the meaning of more specific senses. Because of this, *granny*₂ should be taken to mean “an old woman **who is not** the mother of your father or mother”.

3 THREE TYPES OF SENSE

There are three labels that can be given to senses. The labels are “core”, “association”, and “metaphor”. 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. Core senses often relate to the physical world, but not always.

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 denotes multiple completely unrelated senses. All other senses of a word are related to a core sense, either by general association (§3.2) or by metaphorical similarity (§3.3).

Selecting The Best Core Sense Sometimes, there are multiple senses which are closely related, and could each be a candidate core sense. In these cases, only one should be labelled as a core sense, and it is up to the annotator to determine which they think is the best. Consider these two senses of *shower*:

Senses of <i>shower</i>		Label
<i>shower</i> ₁	a plumbing fixture that sprays water over you, e.g. “they installed a shower in the bathroom”	
<i>shower</i> ₂	[<i>shower bath</i>] washing yourself by standing upright under water sprayed from a nozzle, e.g. “he took a shower after the game”	Core

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). Senses which are related to a core sense will be treated later (§3.2).

Disregard History Sometimes it is quite easy to predict which of a word’s senses came first historically. However, annotators should not factor into account their intuitions (or knowledge) about the history of a word’s senses when annotating. Instead, they should consider what the core sense is of a word is today. Consider these two senses of the word *train*:

Senses of <i>train</i>		Label
<i>train</i> ₁	[<i>railroad train</i>] 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
<i>train</i> ₃	[<i>caravan, wagon train</i>] a procession (of wagons or mules or camels) traveling together in single file, e.g. “they joined the wagon train for safety”	

Although *train*₁ (locomotive) is unlikely to be the earliest sense of *train*, is it the core sense which most people today would think of.

Multiple Core Senses Words can have multiple core senses. This happens when words denote several senses (or groups of senses) that are completely unrelated. Consider these three senses of the word *bridge*:

Senses of <i>bridge</i>		Label
<i>bridge</i> ₁	[<i>span</i>] a structure that allows people or vehicles to cross an obstacle such as a river or canal or railway etc.	Core
<i>bridge</i> ₅	any of various card games based on whist for four players	Core
<i>bridge</i> ₉	[<i>bridge deck</i>] an upper deck where a ship is steered and the captain stands	Core

These three senses are completely unrelated to each other, and therefore are all core senses. When determining whether two senses are related or not, once again the history of these senses should be disregarded. Historically speaking, *bridge*₉ (ship deck) derives from *bridge*₁ (crossing): on paddle steamers the place the ship was commanded from was a physical bridge connecting two paddle houses, and although the bridge there is gone, the term has survived. However, annotators should **not** consider historical information like this when determining whether two senses are related or not. If they perceive an obvious relation between two senses, then only one should be labelled as a core sense, but if they cannot perceive any obvious relation, then both should be labelled as cores, regardless of the sense's historical derivation. Therefore, *bridge*₉ is a core sense.

3.2 Associated Senses

Words often have several senses which are closely related to a core sense. These associations should be labelled as such, and connected to the core sense which they are associated with. Importantly, this label should **not** be used for senses which bare metaphorical similarities to other senses. Consider the word *neck*:

Senses of <i>neck</i>	Label
<i>neck</i> 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
<i>neck</i> ₂ a narrow elongated projecting strip of land	
<i>neck</i> ₃ a cut of meat from the neck of an animal	Association (from <i>neck</i> ₁)
<i>neck</i> 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”	
<i>neck</i> ₅ [neck opening] an opening in a garment for the neck of the wearer; a part of the garment near the wearer's neck	Association (from <i>neck</i> ₁)

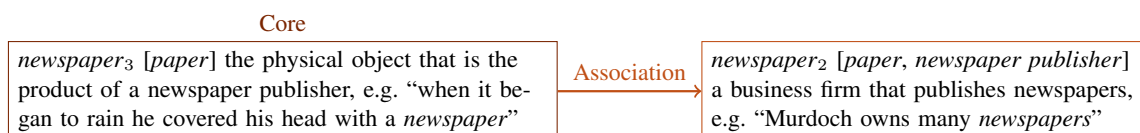
The core sense of *neck* refers to the part of the body connecting the head to the torso (*neck*₁). The other four senses are all related to this sense. It has two senses which are linked by association: meat from that part of the body (*neck*₃), and a hole in clothing for that part of the body (*neck*₅). These senses are clearly related to *neck*₁, but are not metaphorically similar to it. Senses *neck*₂ and *neck*₄ are metaphorically similar to *neck*₁, so they should not be labelled as associated (see §3.3).

There are two kinds of association, which annotators are asked to distinguish between. These are regular associations (§3.2.1) and irregular associations (§3.2.2), and they are detailed in the following sections.

3.2.1 Regular Associations

Regular associations are systematic and predictable, and appear throughout a language. You would expect new words in a language to develop these kinds of associations. Below is a (non-exhaustive) list of examples of regular associations, with examples shown in graphic form:

- Product–producer alternations, e.g. *newspaper*:



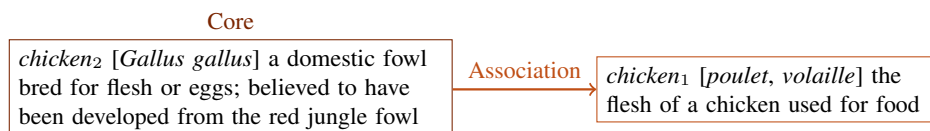
- Result–process alternations, e.g. *merger*, *rip*:



- Fruit–plant alternations, e.g. *apple*, *fig*:



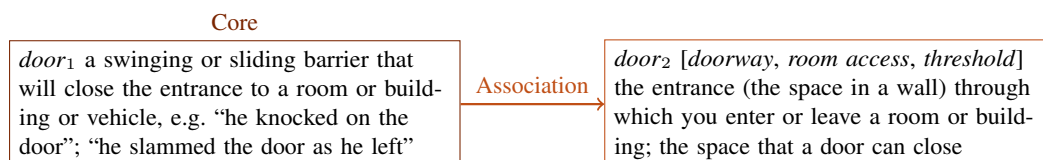
- Animal–meat alternations, e.g. *lamb*, *chicken*:



- Container–contents alternations, e.g. *bottle*, *bin*:



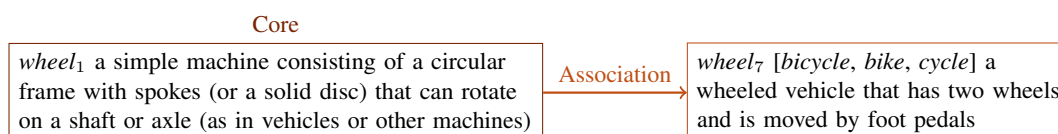
- Figure–ground alternations, e.g. *neck*, *door*:



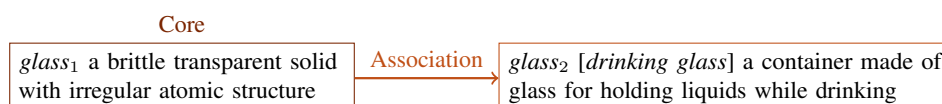
- Area–discipline alternations, e.g. *art*, *biology*:



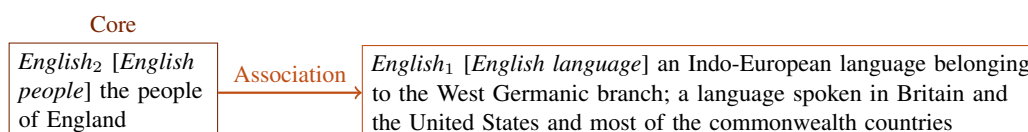
- Part–whole alternations, e.g. *wheel*:



- Material–object alternations, e.g. *iron*, *glass*:



- Nationality–language alternations, e.g. *French*, *English*:



- Class–instance alternations, e.g. *temple*, *helmet*:



All* of these associations could go in either direction (e.g. just as you could have a producer sense associated with a product sense, you could also have a product sense associated with a producer sense). The direction of the relation will depend on which sense the annotators decides is the core sense.

3.2.2 Irregular Associations

Irregular associations are ad-hoc and 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, but is also used in sailing to refer to a type of knot that is the sort that would be tied by a granny. This second sense is associated with the first, but not in a usual systematic way.



3.3 Metaphorical Senses

Some senses have a metaphorical similarity with a core or associated sense. Senses like this should be labelled as metaphors, and connected to the non-metaphorical sense they relate to. Crucially, each metaphorical sense is similar to the non-metaphorical sense it connects to in some respects, but different in others. For example, consider these two senses of the word *leaf*, shown in both tabular and graphic forms:

Senses of <i>leaf</i>		Label
<i>leaf</i> ₁	[<i>leafage, foliage</i>] the main organ of photosynthesis and transpiration in higher plants	Core
<i>leaf</i> ₂	[<i>folio</i>] a sheet of any written or printed material (especially in a manuscript or book)	Metaphor (from <i>leaf</i> ₁)

The first sense, *leaf*₁ (plant), is the core sense. The second sense, *leaf*₂ (page), has some features in common with this sense (e.g. they are both flat), but an important feature missing (it is not part of a plant). Because of this, it is metaphorical. Annotators additionally need to record this feature transformation. For each metaphorical sense, this is done using the following procedure:

1. Add features to the core/associated sense it connects to. Features are plain text, and fill the gap “This thing ____”. For example, the literal 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.
2. For each feature sense, decide whether the metaphorical sense:
 - (a) has the feature (“**kept**”).
 - (b) does not necessarily have the feature (“**lost**”).
 - (c) transforms the feature (“**modified**”).
3. 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).

Annotators should not attempt to create features which capture all of the meaning of a sense. Instead, they should pick features which draw out the metaphorical distinctions.

*The exception is class–instance alternations. If an instance is the core sense, then the class will be a metaphor (see the ‘Generalisations’ section in §3.3).

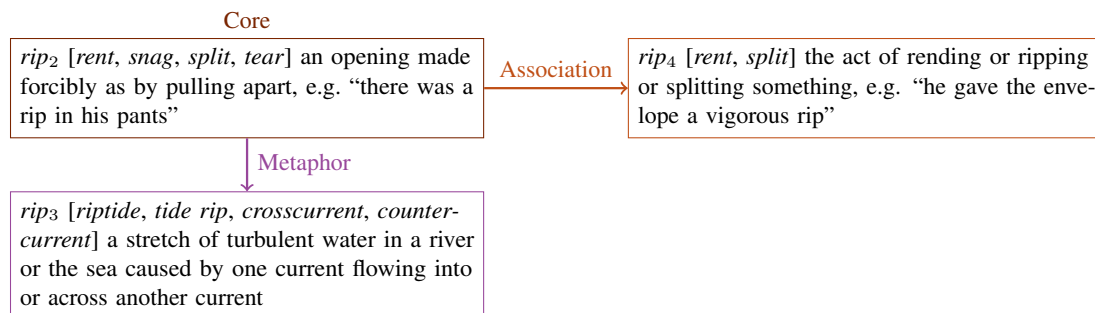
Every metaphorical sense must **either**:

- have at least one kept feature and one lost feature
- have at least one modified feature

Thus, for the *leaf* example, a valid feature transformation (which meets the first requirement) would be:

Features of <i>leaf</i> ₁ :		Features of <i>leaf</i> ₂ :
is flat	⇒	is flat
is part of a plant	⇒	is part of a plant

Choosing the Sense a Metaphor is Similar To Often there are several associated senses which a metaphorical sense could connect to. Connecting it to the correct sense requires annotators to be precise about the type of a sense. For example, if a metaphorical sense refers to an event, then the sense it connects to should also refer to an event; if a metaphorical sense refers to an action, then the sense it connects to should also refer to an action. Consider this annotation for these senses of *rip*:



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 sense *rip*₂, because both refer to the rip itself (rather than the act of creating a rip, as in *rip*₁).

Multiple Metaphorical Extensions When multiple metaphorical senses all connect to the same non-metaphorical sense, they should each transform the same set of features. In these cases, adding more features can be helpful. Annotators should design features which distinguish between different types of metaphorical similarity. For example, consider again the word *train*, this time shown with additional metaphorical senses:

Senses of <i>train</i>	Label	Features
<i>train</i> ₁ [<i>railroad train</i>] 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
<i>train</i> ₃ [<i>caravan, wagon train</i>] a procession (of wagons or mules or camels) traveling together in single file, e.g. “they joined the wagon train for safety”	Metaphor (from <i>train</i> ₁)	is pulled from the front is a sequence of wagons runs on a railway is a means of transit
<i>train</i> ₄ a series of consequences wrought by an event, e.g. it led to a train of disasters	Metaphor (from <i>train</i> ₁)	is pulled from the front is a sequence of consequences runs on a railway is a means of transit
<i>train</i> 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”	Metaphor (from <i>train</i> 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 same set of features from *train*₁ (locomotive). However, they transform the features in different ways. For example, *train*₄ (consequences) is a sequence of things, similar to how *train*₁ is comprised of a sequence of carriages. The sense *train*₃ is also a sequence of things, and has more in common with *train*₁, because it is also a means of transit. The final sense’s metaphorical similarity is very different: *train*₅ (gown) is pulled from the front by the bride, like how a *train*₁ is drawn along from the front by an engine carriage. In this example, features have been selected to capture each of these differences.

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. Consider these two senses of the word *granny*:

Senses of <i>granny</i>	Label	Features
<i>granny</i> ₁ [grandma, grandmother, grannie, gran, nan, nanna] the mother of your father or mother	Core	is an elderly woman has grandchildren
<i>granny</i> ₂ an old woman	Metaphor (from <i>granny</i> ₁)	is an elderly woman has grandchildren

Recall that *granny*₂ does not overlap with *granny*₁ (see §2). With this in mind, it describes specifically old women who do not have grandchildren. It should therefore be labelled as a metaphor.

Abstractions A special case of generalisation is an abstraction from a sense that refers to an object, to a sense that describes a property of that object (which can be used to describe things that are not that object). This should be labelled as metaphorical. This commonly occurs with colour names that are derived from real-world referents, e.g. *amber*:

Senses of <i>amber</i>	Label	Features
<i>amber</i> ₁ [gold] a deep yellow color, e.g. “an amber light illuminated the room”	Metaphor (from <i>amber</i> ₂)	is a deep yellow colour is fossil resin
<i>amber</i> ₂ a hard yellowish to brownish translucent fossil resin; used for jewelry	Core	is a deep yellow colour is fossil resin

The sense *amber*₂ can be used in clear metaphorical cases (e.g. “his eyes were amber”), and so it should be labelled as a metaphor. 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.

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 find a suitable feature transformation, it should be treated as a metaphor. Otherwise, it should be treated as association.

4 ADDITIONAL OPTIONS

In rare cases, the senses of a word will be unsuitable for this annotation task. In these cases, annotators have additional options to make a word properly treatable. There are three options they have: splitting a sense in two, adding a new “virtual” sense, or labelling a sense as a “secondary core”.

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*

<i>birth</i> 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. Each have should then be treated in isolation, as usual:

Senses of <i>birth</i>	Label
<i>birth</i> _{1A} 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” (non-metaphorical part only)	
<i>birth</i> _{1B} 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” (metaphorical part only)	Metaphor (from <i>birth</i> _{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. For ambiguously worded senses, refer to §2.

4.2 Virtual Senses

Sometimes, the sense which a metaphor 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.

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 <i>bee</i>	Label
<i>bee</i> ₁ any of numerous hairy-bodied insects including social and solitary species	Core
<i>bee</i> ₂ a social gathering to carry out some communal task or to hold competitions	

The second sense relates metaphorically to *bee*₁, but not in terms of the specifics: it is not metaphorically similar to the insect, but to the communal work which the insect is known for. To express this, it is necessary to add a virtual sense expressing this, *bee*₃:

Senses of <i>bee</i>	Label
<i>bee</i> ₁ any of numerous hairy-bodied insects including social and solitary species	Core
<i>bee</i> ₂ a social gathering to carry out some communal task or to hold competitions	Metaphor (from <i>bee</i> ₃)
<i>bee</i> ₃ (virtual) a gathering of bees working on some communal task	Association (from <i>bee</i> ₁)

Core

*bee*₁ any of numerous hairy-bodied insects including social and solitary species

Association

*bee*₃ a gathering of bees working on some communal task

Metaphor

*bee*₂ a social gathering to carry out some communal task or to hold competitions

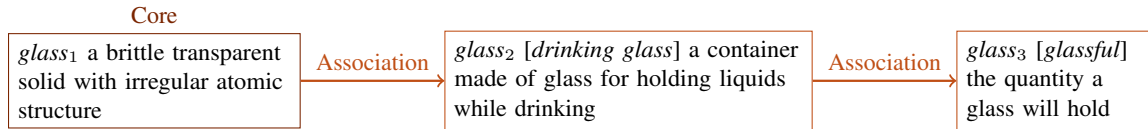
An example of a valid feature transformation for this metaphorical extension might be:

Features of <i>bee</i> ₃ :	Features of <i>bee</i> ₂ :
is a cooperative group of bees	⇒ is a cooperative group of people

4.3 Secondary Cores

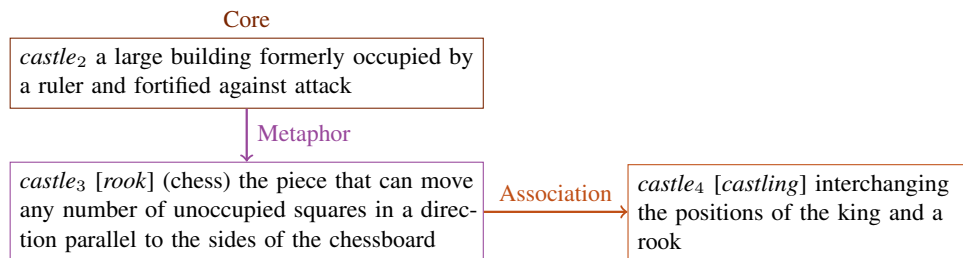
Our annotation procedure requires that each associated sense is connected to a core sense, and each metaphorical sense is connected to either a core sense or an associated sense. However, there are cases where this requirements needs to be violated. This happens in three cases:

- Sometimes, an associated sense extends another associated sense, rather than a core sense, e.g. *glass*₃:



The core sense of *glass* is *glass*₁ (transparent material). This is associated with *glass*₂ (cup) by a material–object association. This sense is, in turn, associated with *glass*₃ (contents of cup) by a container–contents association. Both of these extensions are typical regular associations (see §3.2.1). However, nothing directly links *glass*₁ to *glass*₃.

- Sometimes, an associated sense extends a metaphorical sense, rather than a core sense, e.g. *castle*₄:

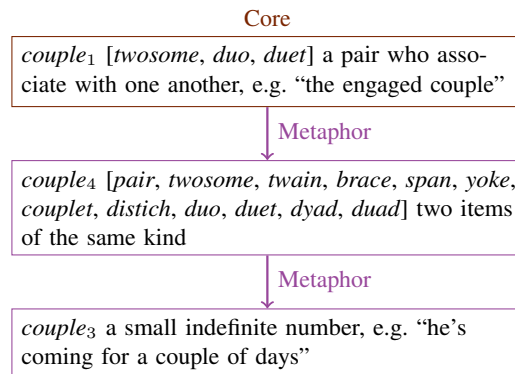


The core sense of *castle* is *castle*₂ (fort). This is metaphorically similar to *castle*₃ (chess piece) by simple resemblance; an example feature transformation might be:

Features of <i>castle</i> ₂ :		Features of <i>castle</i> ₃ :
has crenellations	⇒	has crenellations
is a building	⇒	is-a building

This sense is then associated with *castle*₄ (chess move). Nothing directly links this sense to *castle*₂.

- Sometimes, a metaphorical sense extends another metaphorical sense, rather than a core or associated sense, e.g. *couple*:



The core sense of *couple* is *couple*₁ (two people). A loosening of constraints extends this metaphorically to *couple*₄ (two of anything). This, in turn, extends further to *couple*₃ (a small number of anything). This is a two-stage feature transformation:

Features of <i>couple</i> ₁ :		Features of <i>couple</i> ₄ :		Features of <i>couple</i> ₃ :
is two people	⇒	is two of anything	⇒	is a small number of anything

Nothing directly links *couple*₁ to *couple*₃.

When any of these three cases occur, it is necessary to override our annotation requirements. This is done by marking a non-core sense as a “secondary core”. Senses which are labelled as secondary cores function as core senses, which means that other metaphors or associations can connect to them. Any sense which is connected to another sense via metaphor or association can be labelled as a secondary core. In the above examples, *glass*₂, *castle*₃, and *couple*₄ are all secondary cores.

With this option, 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 true in the case of metaphorical extensions on other metaphorical extensions, which can usually be explained by linking back to the original core sense; in these cases, a secondary core should only be added if there is no way to formulate a direct link.