LEAMR (Linguistically Enriched AMR) Alignments
Alignment Guidelines

Alignments in this scheme are designed to be mutually exclusive and comprehensive alignments between connected AMR subgraphs and English token spans. This alignment scheme also includes several layers of alignments with primary alignments for nodes and edges and secondary alignments for duplicate subgraphs and reentrancies.

Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29.

```
(j / join-01
      :ARG0 (p / person
              :name (p2 / name :op1 (x0 / "Pierre") :op2 (x1 / "Vinken"))
              :age (t / temporal-quantity :quant x3 / 61
                      :unit (y / year)))
      :ARG1 (b / board
              :ARG1-of (h / have-org-role-91
                      :ARG0 p
                      :ARG2 (d2 / director
                              :mod (e / executive :polarity (x4 / -)))))
      :time (d / date-entity :month (x5 / 11) :day (x6 / 29)))
```

Primary alignments: [Pierre Vinken]$_{p\ :name\ (p2\ :op1\ x0\ :op2\ x1)}$, [61 years old]$_{:age\ (t\ :quant\ x3\ :unit\ y)}$, will$_\varnothing$ join$_{j\ :ARG0\ :ARG1}$ the$_\varnothing$ board$_b$ as$_\varnothing$ a$_\varnothing$ nonexecutive$_{:mod\ (e\ :polarity\ x4)}$ director$_{:ARG1-of\ (h\ :ARG0\ :ARG2\ d2)}$ [Nov. 29]$_{:time\ (d\ :month\ x5\ :day\ x6)}$ ·

Secondary alignment: p reentrancy (under h :ARG0) is PRAGMATIC, aligned to "director" token

Token indices look like:
0/Pierre 1/Vinken 2/, 3/61 4/years 5/old 6/, 7/will 8/join 9/the 10/board 11/as 12/a 13/nonexecutive 14/director 15/Nov. 16/29 17/.

j	join-01	8	
p	person	0,1	
p2	name	0,1	
x0	"Pierre"	0,1	
x1	"Vinken"	0,1	

t	temporal-quantity	3,4,5	
x3	61	3,4,5	
y	year	3,4,5	
b	board	10	
h	have-org-role-91	14	
d2	director	14	
e	executive	13	
x4	-	13	
d	date-entity	15,16	
x5	11	15,16	
x6	29	15,16	
Edge alignments			
j :ARG0 p	join-01 :ARG0 person	8	
p :name p2	person :name name	0,1	
p2 :op1 x0	name :op1 "Pierre"	0,1	
p2 :op2 x1	name :op2 "Vinken"	0,1	
p :age t	person :age temporal-quantity	3,4,5	
t :quant x3	temporal-quantity :quant 61	3,4,5	
t :unit y	temporal-quantity :unit year	3,4,5	
j :ARG1 b	join-01 :ARG1 board	8	
b :ARG1-of h	board :ARG1-of have-org-role-91	14	
h :ARG0 p	have-org-role-91 :ARG0 person	14	
h :ARG2 d2	have-org-role-91 :ARG2	14	

	director		
d2 :mod e	director :mod executive	13	
e :polarity x4	executive :polarity -	13	
j :time d	join-01 :time date-entity	15,16	
d :month x5	date-entity :month 11	15,16	
d :day x6	date-entity :day 29	15,16	
Reentrancy alignments			
h :ARG0 p	have-org-role-91 :ARG0 person	14	pragmatic

General Principles
- Comprehensive/Mutually Exclusive: Every node & edge must be aligned to exactly one token span
- When it is completely ambiguous which of two spans to align to, you can align to whichever appears first in the sentence. For example, when a name is repeated several times in the sentence, you may align to the first mention.
- Connected: nodes aligned to a span must form a connected subgraph[1] with two exceptions:
  - Duplicate Subgraphs: A span may be aligned to multiple subgraphs if one subgraph is a duplicate of the others. This is often necessary when dealing with ellipsis where there is more semantic content in the AMR than is pronounced in the sentence and thus several identical parts of the AMR must be aligned to the same span. See the section below on how to align duplicate subgraphs.
  - Cases like "never": A span may be aligned to two terminal nodes that have the same parent. For example, "never" will generally align to ":polarity - :time ever", two nodes and two edges which share the same parent.

Spans
- Spans are chunks of one or more tokens that are used to identify named entities or multiword expressions as a single unit. All alignments need to specify a span which must be a contiguous list of token indices (e.g., "5,6,7" but not "4,8"). In this schema, spans must be mutually exclusive and cannot be nested (Though that is allowed by other researchers, for example, in Named Entity Recognition).

---

[1] Nodes *V* and edges *E* such that every edge has at least one adjacent node in *V* and such that (ignoring edge direction) there is a path between any two nodes in *V*. There may be "dangling edges" connected to only one node in *V* (incoming or outgoing).

- Named Entities: Names such as *John Smith* or *Democratic Republic of the Congo* will be aligned as a span. Any nodes or edges, etc. aligned to these spans must specify the token index of every token in the span.
- Multiword Expressions: When several tokens form a multiword expression where the words together form a specific idiomatic meaning, possibly unrelated to the tokens' individual meanings, we annotate these cases as a span. For example, *in order to* and *look after* are multiword expressions and will be treated as spans. In cases where a multiword expression is discontiguous (e.g. "**think** the plan **over**"), align only to the first contiguous part of the expression (i.e. **think**).
- Dates: When aligning date and time expressions, the span should include relevant date and time entities (e.g., "**September 9th, 2021**") and cardinal or ordinal numbers (e.g., "**3 years**"), but should not include modifiers, prepositions, or postpositions at the beginning or end of the expression (e.g. "earlier this **year**", "**one week** ago"). If an expression is an interval, both date expressions should be included (e.g. "**5 to 7**", "**September to November**"). If the date expression includes a multiword expression, make sure this is included (e.g. "**10 years old**").
- Quantities & Money: For currency or other quantities any units or cardinal numbers should be included (e.g. "**$ 30 billion**") but no modifiers, prepositions, or postpositions at the beginning or end of the expression (e.g., "more than **30 kilos**").

Nodes
It is usually intuitive which node and which span should be aligned. Here are a few notes for resolving ambiguous cases.
- Named entities: named entities will generally be aligned to several nodes and edges including a node specifying the type of entity (e.g. *New York => city, name, "New", "York"*). A rare exception is when the entity type is pronounced by a separate token in the sentence such as "the company IBM" where the alignment must be *company => company* and *IBM => name, "International", "Business", "Machines".* The *:name* edge is then aligned with *IBM.*
- Punctuation: Alignments to punctuation may be rare but are necessary in a few notable cases:
  - *and* nodes can be aligned to comma (,) or semicolon (;) if there is no other span to align to.
  - *multi-sentence* nodes will usually be aligned to a period (.), semi-colon (;), or other punctuation introducing a new sentence.
  - *:mode interrogative* will usually be aligned to a question mark (?).
  - Rarely, *mean-01* must be aligned to a colon (:) or other punctuation.
- AMR specific concepts: some nodes in AMR are notational and don't correspond to a specific linguistic frame or entity (e.g., *multi-sentence, include-91, have-org-role-91*). Often these may be aligned as a part of a larger subgraph, or in some cases to an unaligned token or punctuation. Use your best judgement to determine which span is more indicative of the concept. I.e., which span, if you saw it, would most accurately be predictive of the concepts presence in the gold AMR.
- Comparative constructions: comparative expressions like "*bigger", "more than", "less", "most", "too", "as...as*" may be aligned to the concept *have-degree-91*. This is a fairly complex frame

which often takes multiple arguments and triggers a reentrancy. Given the large number of constructions it can occur with, it is important to align it correctly. If the sentence has a discontiguous multiword expression such as "*as...as*" or "*more...than*", align to the first span to appear in the sentence.

Edges
- :ARGX edges should always be aligned to the same span as the **parent** (:ARGX-of edges should always be aligned to the child)
- :opX edges should always be aligned with the **parent**.
- :sntX edges should always be aligned with the **parent**.
- :domain edges should always be aligned with the **parent**. (Don't align these edges to copula tokens as it tends to make parsing difficult.)
- :poss and :part edges may be aligned to a token (i.e. *'s* or *of*) or with the **parent.**
- :name edges should always be aligned with the **child**. (Usually, the child and parent will be aligned to the same span.)
- Other edges may be aligned to the same span as the **child** or to an unaligned span such as a preposition or subordinating conjunction. (e.g., *:location => in* or *:time => at*)

Duplicate Subgraphs
Here is an example.

"It would **work** like medicare …"
(**w/work-09**
        :ARG1 (i/it)
        :ARG1-of (r/resemble-01
                :ARG2 (**w2/work-09**
                        :ARG1 (p/program
                                :name (n/name
                                        :op1 "Medicare")
                                :wiki "Medicare_(United_States)"))))

In the above example, there are two *work-09* concepts in the AMR but only one pronounced "work" token in the sentence. We can blame this on ellipsis. Under that hypothesis, the sentence could alternatively be pronounced "It would **work** similarly to how medicare **works**." Since we have to annotate the sentences we are given, we handle these situations in a secondary layer of alignments for duplicate subgraphs. We will say that *w2/work-09* is a duplicate of *w/work-09,* that the primary alignments will include the alignment *work => w/work-09,* but there will also be a secondary alignment in the duplicate subgraphs layer that looks like *work => w2/work-09*. This allows us to align every node between the two layers and it allows us to specify that the second alignment is the result of ellipses or

some other linguistic phenomena. As a matter of convention, the first node or subgraph (in the linearized AMR) gets the primary alignment and any duplicates are given secondary alignments. When annotating, you can annotate duplicate subgraph alignments using an asterisk (*) before the token indices. (In the case of more than one duplicate, you can use two asterisks (**), etc.)

Here is an example annotation:

w	work-09	2	
w2	work-09	*2	duplicate
Edge alignments			
w :ARG1 i	work-09 :ARG1 it	2	
W2 :ARG1 p	work-09 :ARG1 program	*2	duplicate

Reentrancies
In addition to all edges being aligned, any edge which is a reentrancy (an edge which re-uses a previously named concept as an argument) must be aligned to a span which signals the reentrancies. We can identify reentrancies by looking at any edges where the target has multiple parents. There are several linguistic reasons an AMR might have a reentrancy, so reentrancies are also annotated with a reentrancy type.

Here is an example:

I actually had some other classmates there , and was going to call them .

(a/and
        :op1 (h/have-03
                :ARG0 (i/i)
                :ARG1 (c/classmate
                        :location (t/there)
                        :mod (o/other)
                        :quant (s/some))
                :ARG1-of (a2/actual-02))
        :op2 (c2/call-02
                **:ARG0 i
                :ARG1 c**))

Reentrancy Edges			
h :ARG0 i	have-03 :ARG0 i	_	
h :ARG1 c	have-03 :ARG1 classmate	_	
c2 :ARG0 i	call-02 :ARG0 i	8	coordination
c2 :ARG1 c	call-02 :ARG1 classmate	13	coref

Note that above, any edges where the target has multiple parents gets listed row by row. For each target one edge is annotated with an underscore (_) marking that as the primary parent. Any other reentrancies are aligned to a token span which is responsible for triggering the reentrancy and annotated with the reentrancy type from the list:

coref	Coreference triggered by a pronoun (including possessive or reflexive)
repetition	Coreference triggered by a repeated name or non-pronominal phrase
coordination	Reentrancies triggered by coordination of two or more phrases sharing an argument
control	Reentrancies triggered by control verbs, control nouns, or control adjectives.
adjunct control	Reentrancies triggered by control within an adjunct phrase
unmarked adjunct control	Reentrancies triggered by control within an adjunct phrase with only a bare verb and no subordinating conjunction
comparative control	Reentrancies triggered by a comparative construction
pragmatic	Reentrancies that must be resolved using context

Above, one reentrancy *call-02 :ARG0 i* is aligned to "and" which triggers the reentrancy by coordination. The other reentrancy *call-02 :ARG1 classmate* is aligned to "them" which triggers a reentrancy through coreference.

- Coreference:

- ○ Pronominal: Reentrancies triggered by a pronoun, including possessive and reflexive pronouns, are aligned to the pronoun and annotated with the label **coref**.
- ○ Repetition: Reentrancies triggered by a named entity or other coreferential phrase being repeated in the sentence are aligned to the repeated phrase and annotated with the label **repetition**. (This includes cases where the phrase is not exactly the same but is coreferential. E.g., "**Chinese** … **China** ..." or "**United States** … **U.S.** ...")
- Coordination: In a sentence with VP coordination such as "The people **cheered and celebrated**.", the same argument (or arguments) is shared by multiple verbs. This can happen not just for verb phrases but for any phrase which takes one or more semantic arguments. In cases where a reentrancy occurs because of coordination, the reentrancy is aligned to the coordinating conjunction (usually "and") and annotated with the label **coordination**. You also need to decide which edge is the primary parent edge and which ones are annotated as reentrancies. By convention, given coordination of the form "A and B", and edges out of A are annotated as primary edges and the edges out of B are treated as reentrancies.
- Control: In sentences with control verbs, control adjectives, or control nouns, a reentrancy is triggered by a grandparent of the reentrancy's target node. Control is a lexical property of some word or phrase where an argument of the control phrase also becomes an argument of a controlled phrase. For example, in the sentence "John was **afraid** to **speak up**", *John* is a semantic argument of both *afraid* and *speak up*, and so the corresponding AMR will have a reentrancy corresponding to the relationship between *John* and *speak up*. *Afraid* is the control phrase and *speak up* is the controlled phrase which is an argument of *afraid.* Both the control phrase and controlled phrase take *John* as an argument. The AMR looks like (f / fear-01 :ARG0 (p / person :name (n / name :op1 "John")) :ARG1 (s / speak-up-02 :ARG0 p)). Since control is a lexical property of *afraid* and other control phrases, we align control reentrancies with the control phrase. So the edge between *John* and *afraid* (fear-01 :ARG0 person) is annotated as the primary edge and the edge between *John* and *speak up* (speak-up-02 :ARG0 person) is annotated as a reentrancy triggered by *afraid.* Reentrancies triggered by a control phrase are annotated with the label **control**.

Reentrancy Edges			
f :ARG0 p	fear-01 :ARG0 person	_	
s :ARG0 p	speak-up-02 :ARG0 person	2	control

- Adjunct control:
  Several types of control do not fit the pattern described above. Most of these fall into the category of adjunct control. Reentrancies triggered by adjunct control are annotated with the label **adjunct control**.
  - ○ Infinitive Purpose Clauses: An infinitive purpose clause is a clause like "I went to the store **to buy some milk**" where *to buy some milk* is an adjunct clause specifying the

purpose of going to the store. Infinitive purpose clauses take the infinitive form of a verb, hence the name. Reentrencies triggered by this type of clause should be aligned to the infinitival *to* token.

- ○ Subordinate Gerund Clauses: Another type of adjunct control can be seen in "Mary did her homework **while listening to music**". In this and similar cases, adjunct control is introduced by subordinating conjunction (such as *after*, *while,* etc.) followed by a verb in gerund form. In these cases, the reentrancy should be aligned to the subordinating conjunction.
    - ■ More rarely, you may see a sentence like "Mary did her homework, listening to music". This is like a subordinate gerund clause, but no subordinating conjunction is present. In cases like this, the reentrancy should be aligned to the gerund verb and you should annotate it with the special label **unmarked adjunct control**.
- ● Comparative control: comparative expressions like "*bigger", "more", "less", "most", "too", "as...as"* may be aligned to the concept *have-degree-91* which often includes a reentrancy to the attribute whose degree is being described. This type of reentrancy gets the label **comparative control**. The first mention of the attribute in the linearized AMR is annotated as the primary edge, while the second attribute is annotated as a reentrancy.
  *"The key is to be **as** objective **as** possible ."*
  (k/key-02
        :ARG1 (o/objective
            :ARG2-of (**h/have-degree-91**
              :ARG1 (y/you)
              :ARG3 (e/equal)
              :ARG4 (p/possible-01
                 :ARG1 o))))
  For example, the reentrancy pointing to the node *objective* in the above AMR should be aligned to *as* (the first token in *as...as* construction above).

- ● Pragmatic: In many cases, there is not a clear way to infer a reentrancy or how it should be resolved using either syntax or coreference. In these cases, the reentrancy is aligned to the same span as the edge itself is aligned and is annotated with the label **pragmatic**. These often come up when a frame has an argument which is obvious to a speaker because of world knowledge but which is not otherwise easy to resolve. Reentrancies of this type are labelled **pragmatic**.

  *"**Precursory** signs of the earthquake"*
  (s/sign
        :ARG1-of (h/have-03
            :ARG0 (e/earthquake))
        :ARG1-of (**p/precede-01**
            :ARG2 e))

For example, the reentrancy pointing to the node *earthquake* in the above AMR should be aligned to "*precursory".* The intuition is that the word "precursory" has an implicit semantic argument, but this semantic argument is not resolved by normal processes like syntactic composition or coreference. Instead, a natural English speaker uses their general knowledge to pick *"earthquake"* as the most likely candidate. The reentrancy should be aligned to the span with an implicit argument, which in this case is *"precursory"*.

Another example:
*"John met up with a **friend**"*
(m/meet-up-04
      :ARG0 (p/person
           :name (n/name
                :op1 "John"))
      :ARG1 (p2/person
           :ARG0-of (h/have-rel-role-91
           :ARG1 p
               :ARG2 (f/friend))))
In this example, there is an implicit argument of *"friend"* answering the question "a friend of whom". This is represented in the AMR above by a reentrancy to p/person. From context, it is obvious that "a friend" refers to a friend of John's, but inference must be resolved by pragmatic knowledge since there is no syntactic or coreferential way to resolve it. Like in the previous example, the reentrancy should be aligned to the span with an implicit argument, which in this case is *"friend".*