`ding-01 :ARG0`: An AMR Corpus for Spontaneous French Dialogue

Jeongwoo Kang[∀]   **Maria Boritchev**[∃]   **Maximin Coavoux**[∀]

[∀] Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
[∃] LTCI, Télécom Paris, 91120 Palaiseau, France

`jeongwoo.jay.kang@gmail.com`
`maria.boritchev@telecom-paris.fr`
`maximin.coavoux@univ-grenoble-alpes.fr`

Abstract

We present our work to build a French semantic corpus by annotating French dialogue in Abstract Meaning Representation (AMR). Specifically, we annotate the DinG corpus, consisting of transcripts of spontaneous French dialogues recorded during the board game *Catan*. As AMR has insufficient coverage of the dynamics of spontaneous speech, we extend the framework to better represent spontaneous speech and sentence structures specific to French. Additionally, to support consistent annotation, we provide an annotation guideline detailing these extensions. We publish our corpus under a free license (CC-SA-BY). We also train and evaluate an AMR parser on our data. This model can be used as an assistance annotation tool to provide initial annotations that can be refined by human annotators. Our work contributes to the development of semantic resources for French dialogue.

1 Introduction

Abstract Meaning Representation (Banarescu et al., 2013, AMR) encodes the meaning of a text as a rooted, directed, and acyclic graph (see Figure 1). Representing meaning in a structured form offers several advantages for information systems. AMR reduces semantic ambiguity by explicitly specifying one plausible interpretation among others. Furthermore, because AMR abstracts away from surface variations — especially syntactic variations — sentences with the same underlying meaning share the same AMR representation (*e.g.*, "The police arrested the thief." and "The thief was arrested by the police."). This canonical representation reduces the search space for models, making AMR a useful tool for various NLP tasks, such as machine translation (Wein and Schneider, 2024), automatic text summarization (Liao et al., 2018; Liu et al., 2015), and human-robot interaction (Bonial et al., 2019, 2023).

Training an AMR parser to automatically generate an AMR graph from a given text requires a dataset consisting of texts associated with their corresponding AMR graphs. However, AMR datasets for French are currently scarce, since most available AMR resources are in English. This imbalance in semantic resources limits the development of French semantic parsers, which hinders the progress of French NLP systems that rely on them. Furthermore, most existing AMR data are based on written texts such as newspaper articles and online forums. In contrast, dialogue data, which exhibits unique linguistic features due to its interactive and spontaneous nature –*e.g.,* French discourse markers such as *alors* (then), *du coup* (so), *donc* (so), and backchannels– remain underrepresented.

To fill this gap in French semantic resources, particularly for dialogue, we manually annotate the DinG corpus (Boritchev and Amblard, 2022) in AMR. DinG consists of transcriptions of dialogues recorded during board game sessions of *Catan*, capturing various linguistic features of spoken interaction in French.

However, the standard AMR framework, as currently defined,[1] has limitations in representing speech-specific features. Therefore, we extend AMR by introducing additional relations to (i) annotate two pragmatic phenomena: discourse markers and backchannel expressions, (ii) represent coreference across multiple turns of speech.

To summarize, our main contributions are as follows:

- We publish `ding-01`,[2] a new AMR corpus of spontaneous French dialogue containing 1,830 turns of speech. We aim to expand the

---

[1]The current version of the annotation guideline is available at `https://github.com/amrisi/amr-guidelines/blob/master/amr.md`
[2]`https://doi.org/10.5281/zenodo.15537425`

corpus to cover 3,000 turns of speech by the end of 2025. We also release a *data statement* with the corpus to describe all relevant metadata and potential biases, following best practices for data production for NLP (Bender and Friedman, 2018; McMillan-Major et al., 2024).

- We adapt AMR to represent spontaneous speech phenomena in French, including discourse markers and backchannels.

- We provide an annotation guideline for two purposes: 1) ensure annotation consistency by clarifying aspects not specified in the original AMR annotation guideline 2) newly define how to annotate linguistic features specific to French dialogue.

- We train and evaluate an AMR parser on our dataset to showcase its practical use case. This model is further expected to serve as an annotation assitance tool.

We expect our corpus to contribute to the future development of semantic parsers for French dialogue, along with future (computational) linguistics research on French dialogical data. As noted by Wein and Opitz (2024), AMR corpora and tools are an underexplored source of data for linguistic investigation. The corpus is already getting some interest from the semantics research community, as it has been integrated in Grew (Amblard et al., 2022) and can now be explored in the tool.[3]

2 Background and Related Work

2.1 Introduction to AMR

AMR represents the meaning of texts using directed, acyclic, and rooted graphs. In an AMR graph, the nodes are 1) predicates predefined in Propbank[4] (Palmer et al., 2005), *e.g.*, `break-01` in Figure 1 or 2) English words, *e.g.*, `man` and `window` in Figure 1 or 3) AMR-specific keywords, *e.g.*, `date-entity`.

The edges of the AMR graph are labeled to indicate the relation between nodes. For example, `:ARG0` and `:ARG1` in Figure 1 respectively indicate that `man` is the agent of the predicate `break-01` and that `window` is the object of the

same predicate. This predicate-argument structure is defined in Propbank.[5] An AMR graph can also be represented in textual form (see Figure 2). Although AMR is initially designed for English texts, it is also commonly used to represent non-English texts (Damonte and Cohen, 2018; Xu et al., 2021; Liu et al., 2020). In multilingual settings, two sentences in different languages that convey the same meaning (*i.e.*, sentences that are translations of each other) will share the same AMR graph.
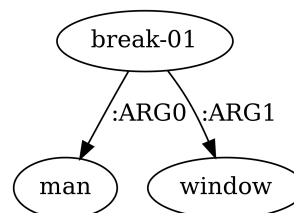

Figure 1: AMR graph for "A man breaks a window" or « Un homme a cassé la fenêtre ».

```
(b / break-01
    :ARG0 (m / man)
    :ARG1 (w / window))
```

Figure 2: AMR graph linearized in text format.

2.2 AMR Datasets

Most large-scale AMR datasets, including AMR 3.0 (Knight et al., 2020) and Massive-AMR (Regan et al., 2024), are available exclusively in English. AMR 3.0 is the most popular dataset for training and evaluating AMR parsers. It contains around 60,000 annotated examples from various sources such as news articles, blogs, and online forums. Massive-AMR, the largest manually annotated AMR dataset, consists of 84,000 utterances addressed to a virtual assistant. Most sentences in Massive-AMR are short questions or requests.

For French, a few datasets are available: *Le Petit Prince* AMR (Kang et al., 2023), Massive-AMR French (Regan et al., 2024) and ReMEDIATE (Druart, 2024). For *Le Petit Prince* AMR, the authors manually aligned the entire English dataset, *The Little Prince* AMR,[6] with the original French text. The French Massive-AMR consists of a part of Massive-AMR English (Regan et al., 2024), manually translated into French. ReMEDIATES is anno-

---

[3] https://semantics.grew.fr/?corpus=ding-01

[4] https://propbank.github.io/v3.4.0/frames/

[5] https://propbank.github.io/v3.4.0/frames/break.html#break.01

[6] https://github.com/flipz357/AMR-World/blob/main/data/reference_amrs/amr-bank-struct-v3.0.txt

tated semi-automatically in French using a trained annotation model. Unlike two previous datasets, ReMEDIATES is not built on pre-existing English data. In terms of corpus type, *The Little Prince* AMR is a literary piece of work. Massive-AMR consists of requests sent to virtual assistants. Finally, ReMEDIATES contains interactions between a virtual assistant and its user to make reservations. Note that ReMEDIATES uses the syntax of AMR graphs but adapts all the concepts and edge labels for Task-Oriented Dialogues (TOD).

Our work stands out from prior work in several key ways. First, we annotate spontaneous conversations between multiple speakers. Our corpus captures real-world interactions, reflecting the dynamics of spontaneous speech in French. Furthermore, *The Little Prince* AMR and Massive-AMR were initially annotated in English and then adapted to other languages through manual translation or crosslingual alignment (assuming that translated sentences should have the same semantic graph as its original sentence). This process can introduce bias, making the data potentially English-centric. We directly annotate French dialogues in AMR without relying on prior English annotations, ensuring that the semantics of French are preserved throughout the annotation process. Finally, while ReMEDIATES is annotated semi-automatically, we annotate the data manually. It is worth emphasizing that large generative language models remain unreliable for semantic annotation tasks, even for English (Ettinger et al., 2023).

2.3 AMR for Dialogues

Although standard AMR provides various semantic roles to present meanings of *texts*, several efforts have been made to extend it to capture various aspects of *dialogue*. DMR (Hu et al., 2022) and Dialogue-AMR (Bonial et al., 2020), as well as the work of Druart (2024) are among these extensions. These three approaches primarily focus on task-oriented dialogues, in which an agent requests an action to a robotic or virtual agent. Therefore, they integrate fine-grained instructions and introduce additional roles to represent, for example, illocutionary force or the speakers' intended contribution (Bonial et al., 2020).

However, these roles are not ideally suited to our corpus, which consists of spontaneous conversations among multiple speakers. While we aim to follow standard AMR conventions as closely

as possible by adhering to the established annotation guidelines, the nature of our data—French dialogue—introduces linguistic phenomena specific to natural oral interaction, such as backchannels and discourse markers.

Backchannels and discourse markers convey pragmatic information in dialogue. However, standard AMR does not take this type of information into account, as specified in its annotation guidelines. Despite this, we chose to annotate the pragmatic information conveyed by backchannels and discourse markers for two main reasons. First, unlike AMR 3.0, which relies primarily on textual data, our corpus consists of dialogues rich in pragmatic content. We believe that annotating this information provides a valuable resource for the study of French dialogue. Furthermore, the additional roles we propose can be easily removed, ensuring compatibility with AMR 3.0.

Second, although the AMR annotation guideline states that pragmatic information is not included, in practice, AMR incorporates some pragmatic elements. For example, the choice of the root node in AMR often depends on the primary focus of the sentence, reflecting pragmatic information. In addition, some predicates (*e.g.,* `know-05` and `see-03`) are used for their discourse functions (*e.g.,* as in "you know" and "you see."), which are also closely related to pragmatics. Thus, adding pragmatic elements to our annotations is not entirely incompatible with standard AMR practices. To account for this pragmatic information, we introduce new roles, which are detailed in Section 5.

3 The DinG Corpus

We annotate the DinG corpus[7] (Boritchev and Amblard, 2022), a collection of manually transcribed multi-party dialogues among French-speaking players of the board game Catan.[8] Catan is a strategic board game centered on resource management and exchange. Thus, players often negotiate resource exchanges with each other, and their actual interactions are recorded in the corpus. We select this corpus for two main reasons.

First, DinG is available under a free license.[9] As

Number of utterances (non-empty)	1,667
Number of tokens covered	17,887
Number of speakers	9

Table 1: Basic statistics on our data.

our goal is to make our data public, selecting open data is a crucial requirement. Second, DinG consists of natural dialogues among speakers. Since the environment is not controlled by the data collectors and the players are free to interact during the game, this dataset captures a natural conversational flow and includes a wide variety of dialogic phenomena. As such, its semantic annotations will serve as an ideal testbed for evaluating pre-trained language models on spontaneous speech transcriptions.

4 `ding-01`

In this section, we present some statistics on the corpus, the annotation process, and the data quality assessed by inter-annotator agreement.

The annotation was carried out over a six-months period, during which approximately 1,830 (see Table 1 for other statistics) turns of speech were annotated using AMR.[10] Among these 1,830 turn takings, some examples only consist of non-annotable words, *e.g., [toux]* (cough), *[rire]* (laugh). The number of utterances (non-empty) in Table 1 excludes these non-annotable examples.

Among these examples, there are 459 discourse markers and 36 instances of *backchannel*. The corpus was primarily annotated by the first author of this article using the *metAMoRphosED* annotation tool (Heinecke, 2023, see Figure 3). Approximately 15% of the examples in the entire corpus were validated by two other annotators, who are co-authors of this article. Specifically, the lead annotator and the two annotators met regularly throughout the annotation process (once a week or every two weeks) to check the validity of the examples one by one and record any difficulties encountered. In case of disagreement among the three annotators, the example was corrected or modified during the discussion.

We encountered several challenges during the annotation process. One example concerned the word '*donc*' (so), which appears frequently in DinG. In most cases, it functions more as a discourse marker (used to start a speech turn or as a filler word) than as a causal connector. However, its usage was often ambiguous, and both interpretations could be valid depending on the context. To reduce ambiguity and improve consistency between annotations, we established the following rule: systematically annotate '*donc*' as a discourse marker, provided that its removal does not change the meaning of the sentence. Our method for addressing other similar challenges by defining clear directions is detailed in our annotation guidelines. Furthermore, when faced with complex cases, or cases where multiple annotation choices were correct, we referred to existing AMR 3.0 data in English to choose the most plausible annotation. These examples contain comments with references to the AMR 3.0 sentences that justify these choices.

To assess the quality of the annotations, 160 examples from our corpus were annotated by two annotators. The agreement score was measured using the SMATCH (Cai and Knight, 2013) score. SMATCH is an evaluation metric for AMR calculated by counting the number of triplets (node, labeled edge, node) in common. We obtained a score of 71.6. For comparison, Banarescu et al. (2013) reports inter-annotator agreement scores ranging from 71 to 83, depending on the data source and the annotators' level of expertise.

After this evaluation, we performed an annotation conflict resolution step to produce our final *gold* corpus. All three authors jointly reviewed these 160 annotation examples. In cases of disagreement, the group resolved conflicts by choosing one of the existing annotations or agreeing on a new alternative.

Common conflicts involved edge labels such as `:ARG0`, `:ARG1`, and `:ARG2`, typically resulting from annotation mistakes that were straightforward to correct once identified. Another recurring issue concerned the selection of synonymous PropBank concepts. For instance, `own-01` and `possess-01` convey the same meaning and share the same two semantic roles (`:ARG0` for the owner and `:ARG1` for the owned item). In the English AMR data, the choice between these concepts is guided by the specific lexical item used in the sentence. We used these cases of conflict to refine our annotation guidelines, ensuring a consistent selection between such synonymous concepts.
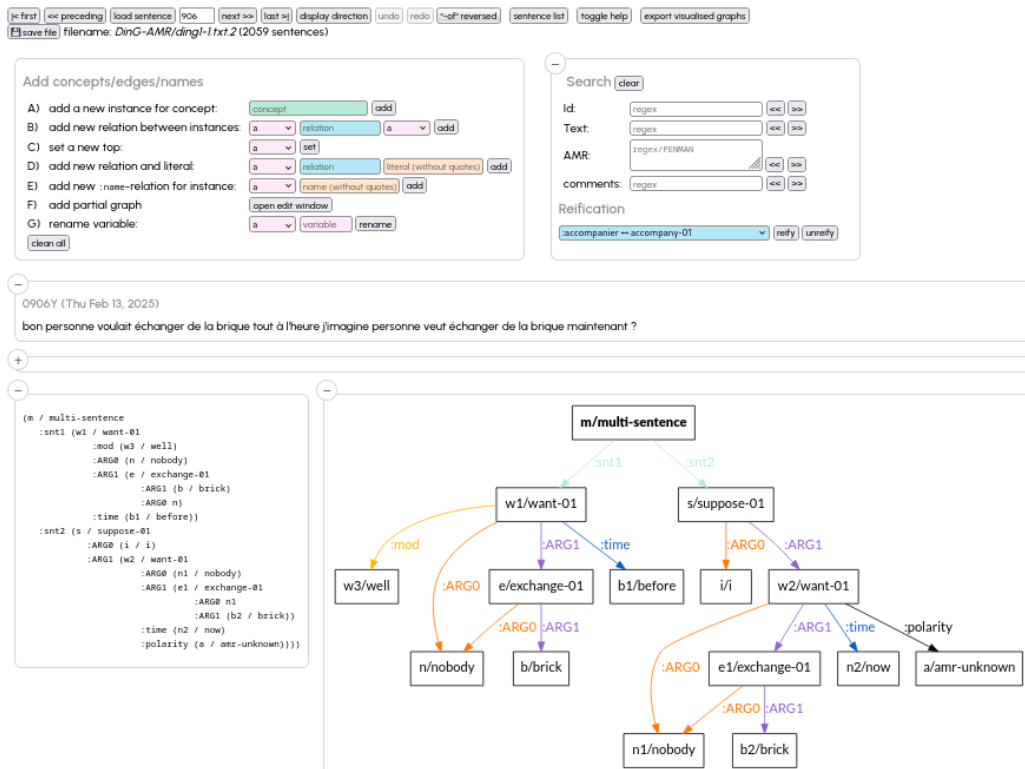
---

[10]We followed the original turn-taking divisions as defined in the DinG corpus.

Figure 3: Screenshot illustrating the annotation process with metAMoRphosED.

5 AMR Adapted for DinG

While adhering as closely as possible to standard AMR, we introduce some extensions to better capture the specific features of spontaneous French speech. Some of these key features are outlined below. In addition, we annotate inter-instance coreference, which is an addition that sets our corpus apart from AMR 3.0. We also adapt the standard AMR concept of *focus* to represent focalization strategies in spoken French. Further details on these extensions are provided in our annotation guideline.

For `ding-01` use cases requiring compatibility with the English AMR 3.0 corpus, these extensions are designed to be easily removable.

5.1 Discourse Markers

Discourse markers are short words or phrases used by speakers to structure their discourse, for example, *donc* (so), *et* (and). They are used to begin an utterance, or can serve as fillers in the middle of an utterance or during a hesitation. We introduce a new role, `:discourse-marker`, to annotate them (see Figure 4). This role can also be reified with the concept `be-discourse-marker-91`.

```
#::id 0780B
(p / put-01
   :ARG0 (y / you)
   :ARG1 (r / road)
   :mode imperative
   :ARG2 (h / here)
   :polarity -
   :discourse-marker "donc")
```

Figure 4: « **Donc** mets pas ta route ici » (So don't put your road here).[11]

5.2 Backchannels

Backchannels refer to short interjections made by a listener while another person is speaking (e.g., *hum*, *mmh-mmh*) to signal attention to the conversation. We annotate them using a new relation `:back-channel`, which can be reified with the concept `be-back-channel-91`. Figure 6 is an annotation of backchannel to a previous utterance (Figure 5).

---

[11] `#::id` specifies the identifier of the example in our corpus. The identifier is composed of a number (i.e., `0780`) and the letter (i.e., `B`) that denotes a speaker.

```
#::id 0851B
(p / possible-01
   :ARG1 (e / exchange-01
      :ARG1 (t / thingy))
   :ARG1-of (r / request-confirmation-91)
   :discourse-marker "du coup"

   :time (n / now))
```

Figure 5: « du coup là on peut échanger des trucs c'est ça ? » (So now we can exchange thingies, right?).

```
#::id 0852Y
(b / be-back-channel-91
   :ARG2 "hum")
```

Figure 6: « hum » (hmm).

5.3 Inter-Instance Coreference

Since the DinG corpus captures interactions between players throughout the game, coreference can span multiple utterances or instances. To ensure a complete representation of meaning, we annotate multi-instance coreferences by marking antecedents that appear in different utterances. For example, the node s0080b_s_stone in Figure 8 indicates that its antecedent comes from the example identified by the ID 0080b in Figure 7 and the concept s / stone associated with that example.

```
# ::id 0080B
(w / want-01
   :ARG0 (y / you)
   :ARG1 (s / stone)
   :polarity (a / amr-unknown))
```

Figure 7: « Tu veux de la **pierre** ? » (You want stone?)

```
# ::id 0082B
(e / exchange-01
   :ARG0 (I / I)
   :ARG2 (y / you)
   :ARG1 (s / sheep
      :quant 3)
   :ARG3 (s1 / s0080B_s_stone))
```

Figure 8: « Je te l'échange contre 3 moutons » (I trade you 3 sheep for it).

5.4 Inter-Instance Verb Ellipsis

Speakers often omit verbs when the meaning remains clear without them (verb ellipsis). When this occurs across different instances (inter-instance level), the omitted verb is mentioned in a previous utterance, and may be spoken by another speaker. We annotate such ellipses similarly to inter-instance coreference, by referencing the utterance ID of the original verb (see Figure 9 and 10).

```
# ::id 0061R
(a / and
   :op2 (p / possible-01
      :ARG1 (p1 / put-01
         :ARG0 (w / we)
         :ARG1 (c / settlement)
         :ARG2 (i / intersect-01)
         :mod (o / only))))
```

Figure 9: « On peut poser les colonies que sur les intersections. » (We can put the settlements only on intersections).

```
# ::id 0062Y
(s / s0061R_p1_put-01
   :ARG0 (w / s0061R_w_we)
   :ARG1 (r / road)
   :ARG2 (e / edge
      :mod (o / only))
```

Figure 10: « et les routes que sur les arêtes » (and roads only on edges).

5.5 Focus Representations

In AMR, the *focus* of a sentence is indicated by a root node. We apply this principle to the annotation of cleft structure, a sentence structure commonly used in French for emphasis. The cleft structure follows the pattern « C'est [subject] qui ... » ("it's [subject] who/that..." in English) used to emphasize the [subject]. To reflect this emphasis on the subject, we select it as the root of the AMR graph. Figure 11 presents an example of a sentence with a cleft structure, accompanied by its annotation in AMR. We adopt the same strategy for cases of left dislocations with pronominal resumption, as in the example: «moi, je veux 2 blés» ("me, I want 2 grains," in English). This type of structure, very common in spoken French, is also a way of expressing focus. In this case, the concept i will be the root of the AMR graph.

```
#::id 0095Y
(y / you
   :ARG0-of (c / choose-01
      :ARG1 (p1 / place
         :ARG2-of (p / put-01
            :ARG1 (t / they))))
   :polarity (a / amr-unknown))
```

Figure 11: « **C'est toi qui** choisis où est-ce que tu les mets ? » (It's **you** who choose where you put them?).

5.6 Disfluencies

Disfluencies are common in spontaneous dialogues. Disfluency markers (*e.g.*, euh, eh), repetitions (*e.g.*, «franchement t'es t'es franchement» "frankly

35

you're you're frankly" in English) and false starts (*e.g.*, «j'ai be- j'ai pas de bois» "I nee- I don't have lumber" in English) are often observed in the DinG corpus. In standard AMR, disfluency markers are not annotated. In line with this convention, we do not annotate disfluency markers, repetitions or short false starts. However, if a false start has interpretable semantic content, we annotate it using :reparandum (see Figure 12) following de Marneffe et al. (2021), who employed this label to mark overridden disfluencies in syntactic annotations.

```
# ::id 0314R
(t / thing
   :value 7
   :ord (o / ordinal-entity
      :value 1)
   :ARG1-of (f / fall-01)
   :ARG1-of (h / have-degree-91
      :ARG5 (r / roll-01
         :ARG1 (d / dice))
      :ARG2 (c / common
         :reparandum (p / possible-01))
      :ARG3 (m / most))
   :discourse-marker "et"
   :discourse-marker "donc"
   :discourse-marker "hein"

   :discourse-marker "et")
```

Figure 12: « et au premier 7 qui va tomber qui est donc euh le lancé de dés **le plus possible hein le plus courant** » (and the first 7 to fall, which is the most posssible the most common dice roll).

6 Models

We train an AMR parser on the previously described data to showcase its practical use. The trained model can assist in the annotation process in our future work. Specifically, the model automatically annotates the data, which can then be manually refined by a human annotator. This semi-automatic approach is useful for scaling up data annotation.

6.1 Sequence-to-Sequence AMR Parser

Recently, sequence-to-sequence AMR parsers (Konstas et al., 2017; Bevilacqua et al., 2021; Yu and Gildea, 2022) have gained popularity due to their strong performance and methodological simplicity. These models take an input sentence and generate an AMR graph in a textual format. Training such models requires a graph linearization step, which converts the AMR graph into a single-line textual format. It also requires a post-processing step because the model may produce ill-formed outputs, for example, graphs with mismatched parentheses or disconnected components. To address

this, a post-processing step is applied to correct formatting errors and reconstruct a well-formed AMR graph from its linearized representation. These steps are described in more detail in the following sections.

6.2 Experimental Setup

To train a sequence-to-sequence AMR parser, we employ a multilingual language model mBart (Liu et al., 2020). To linearize AMR graph, we traverse the graph with depth first search (DFS) in line with Bevilacqua et al. (2021). As a pre-processing step, we rename variables in AMR graphs so that variable numbering follows an order (*e.g.,* a, a2, a3···) instead of random numbering (*e.g.,* a3, a, a2···). In addition, we added empty space between parentheses (see Figure 13 and 14 for differences between before and after pre-preprocessing).

```
(m2 / multi-sentence
   :snt1 (e / exact)
   :snt2 (m / make-05
      :ARG2 (c1 / settlement
         :ARG1-of (b / build-01
            :ARG0 (y / you)))
      :ARG1 (p / point :quant 1))
   :snt3 (m1 / make-05
      :ARG2 (c2 / city)

      :ARG1 (p1 / point :quant 2)))
```

Figure 13: AMR graph before pre-processing.

```
( m / multi-sentence
   :snt1 ( e / exact )
   :snt2 ( m2 / make-05
      :ARG2 ( s / settlement
         :ARG1-of ( b / build-01
            :ARG0 ( y / you ) ) )
      :ARG1 ( p / point :quant 1 ) )
   :snt3 ( m3 / make-05
      :ARG2 ( c / city )

      :ARG1 ( p2 / point :quant 2 ) ) )
```

Figure 14: AMR graph after pre-processing.

We train two distinct models: one trained solely on our data (hereafter referred to as Domain-specific), and another that is first trained on a larger AMR corpus (Knight et al., 2020) and then fine-tuned on our data (hereafter referred to as Pre-trained+Domain-specific). The aim of the second model is to explore whether leveraging large-scale AMR data can facilitate learning our data, which differs in several key aspects: data types (text vs. dialogue transcripts), domain (general vs. board game-related), and semantic roles (standard AMR vs. AMR adapted for French dialogue). Note that the current large-scale AMR data is only available in English and not in our target language, French.

To obtain such data in French, we translated English AMR 3.0 into French using machine translation[12] following Damonte and Cohen (2018).

We split our data set into train, dev and test sets to respectively train the model, to select the best checkpoint, and to evaluate the model's performance on unseen data. The training and dev set respectively consists of 1,375 and 146 examples.[13] For testing, we used the subset of data that underwent a conflict resolution (see Section 4), which consists of 146 examples after filtering out examples solely consisting of non-annotable words.

The model was trained for 4,000 steps, with evaluations conducted every 50 steps on a dev set to select the best-performing checkpoint. Early stopping was applied, terminating training if the validation score did not improve over 750 consecutive steps. The learning rate was set to $3e-5$. Pre-trained+Domain-specific was initially pre-trained on AMR 3.0 data for up to 40,000 steps, with early stopping triggered after 7,500 steps without improvement. Following pre-training, the model was fine-tuned on our data for 4,000 steps using the same settings described above for the Domain-specific training.

6.3 Results

Figure 2 shows the results of our experiments. The findings indicate that pre-training the model on large-scale data is beneficial to learn our corpus in several ways. First, it helps to learn the correct structure of AMR graphs. For example, while the Domain-specific model produced 3 ill-formed graphs out of 146 that could not be recovered during post-processing, the Pre-trained+Domain-specific model successfully avoided such errors.

Moreover, large-scale pre-training helps the model better identify the appropriate predicates for French text. The Domain-specific model occasionally produced predicates that closely resembled the surface form of the French verb, rather than the correct PropBank predicate. For instance, it generated `poser-01` instead of `put-01` for the phrase «tu peux poser...» (you can put...), and `peux-01` instead of `capable-01` for «tu peux » (you can).

	SMATCH
Domain-specific	68.1
Pre-trained+Domain-specific	73.5

Table 2: SMATCH scores of the two models.

Despite these improvements, both models exhibited certain weaknesses. Some sentences in the dataset included non-annotable elements such as coughing or laughter, marked with square brackets (*e.g.,* [toux] for coughing, [rire] for laughing). These elements should not be represented in AMR graphs, but our model failed to capture the pattern and incorrectly annotated some of them (see Figures 15 and 16 for an example). Additionally, although the Pre-trained+Domain-specific model generally performed better at predicting PropBank predicates for French verbs, both models struggled with rare verbs. In such cases, they generated incorrect predicates resembling the verb's surface form—for example, `confine-01` instead of `entrust-01` for «on te confie...» (we entrust you with...).

```
(y / yes
    :mod (a / ah))
```

Figure 15: Reference graph for « ah **[pron fin de mot fricative palatele sourde]**+ oui (0.5s) +[pron]» (ah [pronounce voiceless palatal fricative]+ yes (0.5s)).

```
(m / multi-sentence
   :snt1 (a / ah)
   :snt2 (e / end-01
       :ARG1 (w / word
           :mod (f / fricative))
       :ARG2 (y / yes))
   :snt3 (a2 / and
       :op1 (y2 / yes)))
```

Figure 16: Pre-trained+Domain-specific's prediction for Figure 15.

Lastly, concerning new semantic roles added in our adaptation (`:discourse-marker` and `:back-channel`), both models showed good performance at capturing them. Among 43 discourse markers to predict, both models found around 30 discourse-markers (recall around 0.7). However, some of these discourse-markers were attached to wrong parent nodes. As for `:back-channel`, there was only one example in the test set and both models correctly predicted the `:back-channel`.

37

7 Conclusion and Future Work

We presented our ongoing work to annotate the DinG corpus in AMR to contribute to linguistic resources for French. To better represent the dynamics of spontaneous speech in the DinG corpus, we adapted standard AMR by introducing new semantic roles. We provide an annotation guideline detailing these adaptations, as well as a data statement containing metadata of `ding-01`.[14] To demonstrate a practical application of the dataset, we trained and evaluated an AMR parser on our data. The resulting model can also serve as an annotation assistance tool, helping to accelerate the annotation process and scale up the semantic annotation process. In our future work, we aim to expand the annotated dataset to approximately 3,000 utterances.

UMR Uniform Meaning Representation (UMR) have been introduced in Van Gysel et al. (2021) as an extension of AMR to languages other than English, with the ambition of being used to "annotate the semantic content of a text in any language". UMR is developed as AMR with additional features, notably aspect, tense, modality, along with expanded ones, such as quantification & scope, and discourse relations.

While UMR appears as a very promising representation tool, we have not yet used it for our purposes. There is no French-UMR dataset available for now, which makes evaluation difficult, especially for corpora with complex language phenomena such as DinG. We plan to participate in the development of AMR to UMR translation tools, which should result in several silver French-UMR corpora, paving the way for further meaning representation work. The additions we made to AMR in order to annotate DinG are a lighter version of some of the additional annotations needed for UMR annotation; thus our annotation guidelines could also be of use for a middle step between AMR and UMR.

Acknowledgments

---

[14]The annotation guideline, the data statement, and the corpus are available at `https://doi.org/10.5281/zenodo.15537425`.

References

Maxime Amblard, Bruno Guillaume, Siyana Pavlova, and Guy Perrier. 2022. Graph querying for semantic annotations. In *Proceedings of the 18th Joint ACL-ISO Workshop on Interoperable Semantic Annotation within LREC2022*, pages 95–101.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Emily M. Bender and Batya Friedman. 2018. Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, 6:587–604.

Michele Bevilacqua, Rexhina Blloshmi, and Roberto Navigli. 2021. One spring to rule them both: Symmetric amr semantic parsing and generation without a complex pipeline. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14):12564–12573.

Claire Bonial, Lucia Donatelli, Mitchell Abrams, Stephanie M. Lukin, Stephen Tratz, Matthew Marge, Ron Artstein, David Traum, and Clare Voss. 2020. Dialogue-AMR: Abstract Meaning Representation for dialogue. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 684–695, Marseille, France. European Language Resources Association.

Claire Bonial, Julie Foresta, Nicholas C. Fung, Cory J. Hayes, Philip Osteen, Jacob Arkin, Benned Hedegaard, and Thomas Howard. 2023. Abstract Meaning Representation for grounded human-robot communication. In *Proceedings of the Fourth International Workshop on Designing Meaning Representations*, pages 34–44, Nancy, France. Association for Computational Linguistics.

Claire N. Bonial, Lucia Donatelli, Jessica Ervin, and Clare R. Voss. 2019. Abstract Meaning Representation for human-robot dialogue. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2019*, pages 236–246.

Maria Boritchev and Maxime Amblard. 2022. A multiparty dialogue ressource in French. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 814–823, Marseille, France. European Language Resources Association.

Shu Cai and Kevin Knight. 2013. Smatch: an evaluation metric for semantic feature structures. In *Proceedings of the 51st Annual Meeting of the Association*

for Computational Linguistics (Volume 2: Short Papers), pages 748–752, Sofia, Bulgaria. Association for Computational Linguistics.

Marco Damonte and Shay B. Cohen. 2018. Cross-lingual Abstract Meaning Representation parsing. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1146–1155, New Orleans, Louisiana. Association for Computational Linguistics.

Lucas Druart. 2024. *Vers une Compréhension Contextuelle et Structurée de la Parole Dialogique Orientée Tâche*. Theses, Université d'Avignon.

Allyson Ettinger, Jena Hwang, Valentina Pyatkin, Chandra Bhagavatula, and Yejin Choi. 2023. "you are an expert linguistic annotator": Limits of LLMs as analyzers of Abstract Meaning Representation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 8250–8263, Singapore. Association for Computational Linguistics.

Johannes Heinecke. 2023. metAMoRphosED, a graphical editor for Abstract Meaning Representation. In *Proceedings of the 19th Joint ACL-ISO Workshop on Interoperable Semantics (ISA-19)*, pages 27–32, Nancy, France. Association for Computational Linguistics.

Xiangkun Hu, Junqi Dai, Hang Yan, Yi Zhang, Qipeng Guo, Xipeng Qiu, and Zheng Zhang. 2022. Dialogue meaning representation for task-oriented dialogue systems. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 223–237, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Jeongwoo Kang, Maximin Coavoux, Didier Schwab, and Cédric Lopez. 2023. Analyse sémantique AMR pour le français par transfert translingue. In *Actes de CORIA-TALN 2023. Actes de la 30e Conférence sur le Traitement Automatique des Langues Naturelles (TALN), volume 2 : travaux de recherche originaux – articles courts*, pages 55–62, Paris, France. ATALA.

Kevin Knight, Bianca Badarau, Laura Baranescu, Claire Bonial, Madalina Bardocz, Kira Griffitt, Ulf Hermjakob, Daniel Marcu, Martha Palmer, Tim O'Gorman, and Nathan Schneider. 2020. Abstract meaning representation (AMR) annotation release 3.0 ldc2020t02. Philadelphia: Linguistic Data Consortium.

Ioannis Konstas, Srinivasan Iyer, Mark Yatskar, Yejin Choi, and Luke Zettlemoyer. 2017. Neural AMR: Sequence-to-sequence models for parsing and generation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 146–157, Vancouver, Canada. Association for Computational Linguistics.

Kexin Liao, Logan Lebanoff, and Fei Liu. 2018. Abstract Meaning Representation for multi-document summarization. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1178–1190, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Fei Liu, Jeffrey Flanigan, Sam Thomson, Norman Sadeh, and Noah A. Smith. 2015. Toward abstractive summarization using semantic representations. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1077–1086, Denver, Colorado. Association for Computational Linguistics.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Angelina McMillan-Major, Emily M. Bender, and Batya Friedman. 2024. Data statements: From technical concept to community practice. *ACM J. Responsib. Comput.*, 1(1).

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Michael Regan, Shira Wein, George Baker, and Emilio Monti. 2024. MASSIVE multilingual Abstract Meaning Representation: A dataset and baselines for hallucination detection. In *Proceedings of the 13th Joint Conference on Lexical and Computational Semantics (*SEM 2024)*, pages 1–17, Mexico City, Mexico. Association for Computational Linguistics.

Jens EL Van Gysel, Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Tim O'Gorman, Andrew Cowell, William Croft, Chu-Ren Huang, et al. 2021. Designing a uniform meaning representation for natural language processing. *KI-Künstliche Intelligenz*, 35(3):343–360.

Shira Wein and Juri Opitz. 2024. A survey of AMR applications. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6856–6875.

Shira Wein and Nathan Schneider. 2024. Lost in translationese? reducing translation effect using Abstract Meaning Representation. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 753–765, St. Julian's, Malta. Association for Computational Linguistics.

Dongqin Xu, Junhui Li, Muhua Zhu, Min Zhang, and Guodong Zhou. 2021. XLPT-AMR: Cross-lingual pre-training via multi-task learning for zero-shot AMR parsing and text generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 896–907, Online. Association for Computational Linguistics.

Chen Yu and Daniel Gildea. 2022. Sequence-to-sequence AMR parsing with ancestor information. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 571–577, Dublin, Ireland. Association for Computational Linguistics.