

Lifetime Achievement Award 2023

My Big, Fat 50-Year Journey

Martha Palmer

University of Colorado Boulder, CO

My most heartfelt thanks to ACL for this tremendous honor. I'm completely thrilled. I cannot tell you how surprised I was when I got Iryna's email. It is amazing that my first ACL conference since 2019 in Florence includes this award. What a wonderful way to be back with all of my friends and family here at ACL. I'm going to tell you about my big fat 50-year journey. What have I been doing for the last 50 years? Well, finding meaning, quite literally in words. Or in other words, exploring how computational lexical semantics can support natural language understanding. This is going to be quick. Hold onto your hats, here we go.

1. Texas

I'm from Texas. I went to the University of Texas in Austin where I did my bachelor's degree in philosophy and then did the first interdisciplinary master's degree in computer science and psychology. (Long story, ask me later). I was very fortunate to first get to work with Woody Bledsoe, the father of Automatic Theorem Proving, as a Research Assistant, on pattern recognition of fruit fly chromosomes, in Fortran(!). Later I was equally fortunate to have Robert Simmons¹ (Figure 1) as my MS advisor. He himself had a Psychology PhD since there were no PhDs in Computer Science in the 1950s. He was already widely known as the father of semantic nets, along with Quillian (Simmons 1973). Semantic nets had been used to excellent effect in Terry Winograd's SHRDLU (see below). Simmons had also pointed out the similarities between semantic nets and predicate calculus (Simmons and Bruce 1971). Anyone who's using RDF triples or any kind of knowledge graph today is using a semantic net.

It was a very exciting time to be learning about Natural Language Processing. Winograd's SHRDLU² had just come out—a natural understanding program proof-of-concept (Winograd 1970). He had a virtual robot arm that could move virtual blocks around and follow instructions. It could pick up a red block and put it in a box (see Figure 2). It was all based on procedural semantics. Winograd believed quite firmly that understanding the verb *grasp* meant knowing how to grasp something. For instance, if you need to pick up Block1 and you're already holding Block2, you first must get rid of Block2. Figure 3 has the Lisp code that told the robot arm how to get rid of block two

¹ <https://www.cs.utexas.edu/users/novak/simmons.html>.

² <https://hci.stanford.edu/~winograd/shrdlu/name.html>.

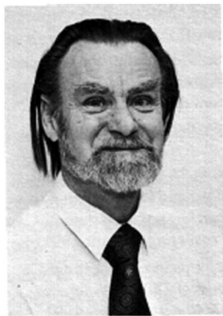


Figure 1
Robert Simmons.

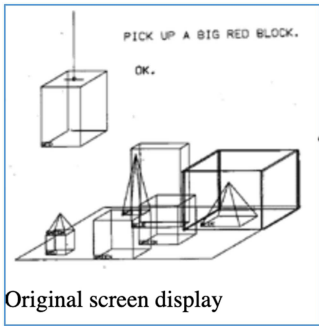


Figure 2
SHRDLU Robot Arm moving a block.

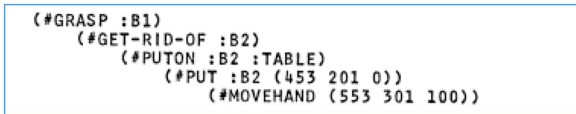


Figure 3
Lisp code for “GRASP.”

by putting it on a table. Simmons had a great group: Gary Hendrix, Jonathan Slocum, Robert Amsler, Mike Smith, and Craig Thompson. Gordon Novak and David Matuszek were also around.

That’s also when Schank’s Conceptual Dependency Nets came out. I still remember reading the Margie paper for the first time and getting incredibly excited about the idea of an abstract transfer of ownership, **ATRANS**, based on physical transfers. If *John gives Mary an aspirin*, then *John* is initiating an event where that *aspirin* will be **transferred from John to Mary**. Schank also had a **PTRANS**, which was the physical transfer. If *I throw you a frisbee* that *frisbee* has physically **moved** from me to you. Then finally an **MTRANS** and that was so cool, a **mental transfer** of information.

Austin, Texas (Figure 4) was a great place to be in those days. I got to attend Willie Nelson’s first 4th of July picnic at Dripping Springs, Texas, where my grandfather is buried, still an annual event. We had Eeyore’s birthday party at Pease Park, another recurring event. I can highly recommend both!



Figure 4
The University of Texas and Eeyore's Birthday Party.

2. Edinburgh

Partly thanks to Woody Bledsoe, I ended up going to Scotland to the University of Edinburgh for my PhD, arriving in 1974. Edinburgh (Figure 5) was very different from Austin. Of course, it was full of brilliant people, Bob Kowalski of Prolog fame, Rod Burstall, Robin Popplestone, and equally brilliant students. You might recognize some of these names: Gordon Plotkin, Chris Mellish, Fernando Pereira, and so forth. But it was also, especially after Texas, a cold gray, dreary, *dreich* place. The pubs all closed at 10 p.m. If one of my male fellow graduate students suggested I join them in a pub that



Figure 5
The Royal Mile, Edinburgh, Scotland.

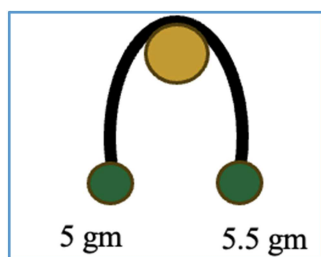


Figure 6
Two particles hung from a pulley.

night, we all had to sit in the Ladies' Lounge because women weren't allowed in the main pub room. That didn't happen very often.

I went to Scotland with the plan of creating a geometry theorem proving tutor, being very keen on computer assisted instruction. However, having a background in NLP from working with Simmons, I ended up implementing the front end of a physics problem solver instead. The idea was to process the sentences that laid out the pulley problems, something like, *A 5 g particle is hung from a pulley and offset by a 5.5 g particle* (Figure 6). My task was to come up with predicate logic representations for the sentences from which equations could be derived. I ended up representing the verb semantics as Horn clauses, with the verb as the consequent. A *hanging* event could be deduced if an **intermediary** was **effecting** a **support** relationship between two objects, like the *pulley* and the *particle*.

$$\text{hang} \leftarrow \text{effect}(\text{intermediary}, \text{support}(\text{pulley}, \text{particle}))$$

For example, a *string* could be the **intermediary** that was *hanging* from the *pulley* to the *particle*. That **effect** relationship could be deduced if there were two **support** relationships between the *pulley* and the **intermediary** and between the **intermediary** and the *particle*.

$$\begin{aligned} &\text{effect}(\text{intermediary}, \text{support}(\text{pulley}, \text{particle})) \leftarrow \\ &\quad \text{support}(\text{pulley}, \text{intermediary}) \\ &\quad \text{support}(\text{intermediary}, \text{particle}) \end{aligned}$$

Then of course, the **offset** relationship must be represented as well. One of the reasons the **intermediary** needs to be made explicit is because the sentence doesn't say anything about the string at all. Since it is implicit, the existence of the string must be deduced.

My dissertation included a 5-step approach to doing natural language understanding (Palmer 1981, 1983, 1990b), as outlined in Table 1.

The first step is to establish the referents of noun phrases, then map from these syntactic constituents to semantic roles. (Sounds just like semantic role labeling!) Then recover the fillers of any implicit roles (like the string, above) and draw inferences, but not too many. It's important to keep that constrained. Then situate the result with respect to the discourse context. This was all done through the procedural interpretation of logic based on the Horn clauses that captured the verb semantics. Those Horn clause representations primarily drove the first three steps but continued to feature all the way

Table 1
Steps in Natural Language Understanding.

Step #	Task
1	Establish noun phrase referents
2	Map from syntactic constituents to semantic roles (SRL)
3	Recover implicit roles
4	Draw inferences (but not too many - constrain carefully)
5	Situate with respect to the discourse context

through. This was all in Prolog, of course. It worked well even if it wasn't at all what I had planned to do when I went to Scotland.

Life is what happens to you when you are busy making other plans—John Lennon

I also did not marry the English fiancé who had been one of the reasons I had gone to Edinburgh in the first place. I married a Scotsman instead (Figure 7).



Figure 7
The wedding.

3. Back to the States

Partly because of my husband and his interest in trying to work in corporate America, we left Scotland in 1978 and spent a year and a half in North Carolina. We next moved to Philadelphia where he landed a job with Scott Paper Company. I visited Duke while we were in North Carolina, and Penn as soon as we got to Philadelphia. Being in those locations and meeting the people there was a gift. Alan Biermann is one of the nicest guys in the world. Aravind Joshi and Bonnie Webber were terrific friends and



Figure 8
Alan Bierman, Bonnie Webber, Aravind Joshi.



Figure 9
Barbara Grosz.



Figure 10
Kathy McKeown, Kathy McCoy, Julia Hirschberg, Martha Pollack.

colleagues. I had spent four years in Scotland immersed in implementing this Prolog natural language understanding system, but had never really talked to anyone about what I was doing. It turns out it's *really important* to be able to *explain* what you are doing to the natural language community. Alan, Aravind, and Bonnie (Figure 8) gave me invaluable help, support, and guidance. They showed me how to articulate what I had been working on and to situate it with respect to the literature. I don't think I would have ever finished my dissertation without them.

I also got to meet Barbara Grosz (Figure 9), who was visiting Penn for a while—another fantastic mentor and helper. In addition, there were all these, oh, my goodness, *FEMALE* graduate students at the University of Pennsylvania: Kathy McKeown, Kathy McCoy, Julia Hirschberg, and Martha Pollack (Figure 10)! Yes, that's right. We had Kathy's McCoy and McKeown and Martha's Palmer and Pollock all at Penn at the same time. It caused more than a little confusion. It was terrific.

4. The PUNDIT System

Next came my first real job at Unisys, also in a suburb of Philadelphia, in Research and Development. We took my Prolog code from my dissertation and put it in a whole new system, first called PUNDIT, and later Kernel. We were doing natural language processing of telexes from Navy ships for DARPA using the same Horn clause approach. Figure 11 is a representation of *replace* where the Agent uses an Instrument to **cause** an **exchange** of one Patient for another Patient. In this dataset, it was quite often *starting air compressors* that were being *replaced*. We used the same mapping rules (see Figure 12) that showed how to map from the syntactic constituents to the semantic roles. A Patient was quite likely to be a syntactic object, but it could also be a syntactic subject if it was in the context of an exchange relationship, as in “*The new starting air compressor replaced the old starting air compressor.*”

The telexes from the Navy ships talked about equipment failures like *during routine start of main propulsion gas turbine, air pressure decreased. resulted in aborted engine start.* and then later, messages about military engagements. We were invited to the first Message Understanding Conference, then called MUC (Palmer, Finin, and Walters 1990). Now they’re called TAC for Text Analysis Conferences. Maybe the DOD was worried we were doing too much *mucking about*, and wanted us to be more serious? For the military engagements, temporal relations were critical and sometimes quite subtle. For a sentence like *an F-14 downed an inbound Mig*, it’s very important to recognize that the *Mig* is approaching the aircraft carrier before it gets shot down by the *F-14*. We don’t need to start a war for the wrong reason. An even closer integration of temporal reasoning with semantics and pragmatics was needed for the military engagements than for equipment failures. As it turned out, we were the only group performer at the second MUC conference that managed to get the temporal ordering right, thanks to Rebecca Passonneau, our temporal expert. (Palmer et al. 1986, 1993; Dahl, Palmer, and Passonneau 1987; Passonneau et al. 1991; Palmer 1990a). We were justifiably very proud of ourselves (see Figure 13)!

It therefore wasn’t too surprising to receive an invitation to give a talk at Bell Labs. Bell Labs at the time was the Google Research or AI2 or “your favorite big tech lab” of today. They were famous for having given us UNIX—thanks very much! They had also

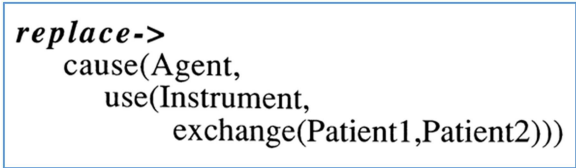


Figure 11
Horn clause for “replace.”

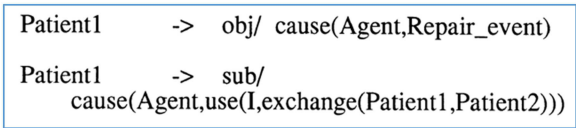


Figure 12
Mapping rules.



Figure 13
The PUNDIT team at Unisys: left to right, Francois Lang, Susan Ball, John Dowding, Shirley Steele, Deborah Dahl, Rebecca Passonneau, Martha Palmer, Marcia Linebarger, Carl Weir, Lynette Hirschman.

put a lot of open source, state-of-the-art speech recognition tools in the public domain. They were very widely respected, making this talk invitation quite a coup. Which made their reaction to my talk that much more disconcerting. “Yeah, OK, this all works fine for this tiny little domain, but you’re never going to be able to scale it up, so why bother?”

5. An Unexpected Change

Around the same time, Unisys hit rocky financial times. Not an uncommon phenomenon—it’s been happening recently to tech companies, especially startups, with all the trouble with certain investment banks and with venture capital. Suddenly our whole lab was in jeopardy. Which is when Scott Paper Company decided to transfer my husband to Singapore (Figure 14).

Life is what happens to you when you are busy making other plans—John Lennon

As it turned out, the three years I spent in Singapore from 1990 to 1993 were incredibly rich. I was introduced to Chinese verb semantics by Wu Zhibiao and other



Figure 14
Life in Singapore.



Figure 15
Beth Levin.

NUS CS PhD students. I had fantastic colleagues at NUS in both CS and Linguistics. Alain Polguère was in the English Department Linguistics Program and taught me both syntax and dependency parsing. Dependency parsing was very new at that time, but Alain had studied under Igor Mel'čuk (Mel'čuk 1988). A highlight was when Beth Levin (Figure 15) came to visit and we went to Bali with the proofs of her book, *A Preliminary Classification of English verbs* (Levin 1993) paid for by Scott Paper Company (my husband is a brilliant negotiator!). Then, four weeks after I'd finally negotiated exactly the position I wanted at NUS, teaching natural language processing and supervising PhD students, Scott Paper announced that they were bringing us home a year early, in three months—I had my first panic attack. *What was I going to do with my students?*

6. Back to Penn – IRCS and VerbNet

I was a Visiting Professor at Delaware for one very enjoyable year. I was also an Adjunct Professor and then eventually an Associate Professor at Penn. I managed to bring one of my students with me, Wu Zhibiao, as well as my newfound interest in Chinese–English machine translation. At Penn we all learned about Aravind Joshi's very elegant Synchronous Tree Adjoining grammars for machine translation (MT). Zhibiao and I wrote an ACL paper on Selectional Restrictions on verb arguments for transfer lexicons for MT (Wu and Palmer 1994). This is now my most highly cited paper (> 5,000 citations). Thank you, Zhibiao! He stayed in Philadelphia for a job at LDC and eventually got a green card. He ended up with a very good job at Oracle, and later PayPal, and his son graduated from Penn just a few years ago.

Meanwhile, I hadn't forgotten about that challenge from Bell Labs about scaling up. We had Beth's book on verb classes now. Some Penn students (Joseph Rosenzweig, Hoa Dang, and Karin Kipper, see Figure 17 for Karin and Hoa) and I started working on coming up with predicate argument structure representations for the Levin classes and that became VerbNet (Dang et al. 1998, 2000; Kipper et al. 2000; Kipper 2005). Figure 16 has an example class for *break*. The members are verbs like *chip*, *crack*, *crash*, *crush*, *fracture*, *rip*. They're all very semantically similar to *break*. This is one of the more semantically homogenous classes, which is nice. But they are also in this class, not just because they're semantically similar, but also because they're syntactically similar. They can all appear in these same syntactic frames: *John broke the vase*, *the vase broke*, *vases break easily*. (The causative/inchoative alternation, or transitive/intransitive, and the middle construction). I found that especially appealing because I thought, even back then, that if we could just get access to enough text, to get enough examples of these verbs appearing in their different syntactic alternations, then we could cluster them

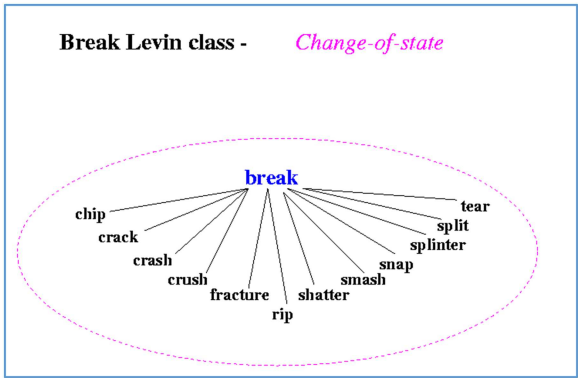


Figure 16
Verb class for Break.



Figure 17
Graduation - Karin Kipper, Martha Palmer, and Hoa Dang.

together automatically. Semantics is almost impossible to work on. It's very subtle, it's very big and complicated, and it's very subjective. You never really know if you're right or not. Syntax is much more concrete and observable. If syntax could help give us insights into semantics, that would be a fantastic boon.

We continued to work on VerbNet for another 25 years (Brown et al. 2022; Kazeminejad et al. 2022; Stowe et al. 2021). The most recent incarnation incorporates James Pustejovsky's (Figure 19) dynamic event structure giving us the same subevent structure as the Generative Lexicon (Pustejovsky 1991). Figure 18 shows some of the other fantastic colleagues that have recently contributed to this project.

Figure 21 provides an example of *give*. In the example sentences, *They lent me a bicycle* or *John gave Mary an aspirin*, there's an Agent, a Recipient, and a Theme. In the first subevent, e1, the Agent has **possession** of the Theme and the Recipient does not ($\neg$ is for negation). In the second subevent, e2, the Agent **transfers** the Theme to the Recipient. This causes the third subevent, e3, where the Recipient now has **possession** of the Theme, and the Agent does not. **ATRANS**. We also have a *Transfer-mesg-37.1* class for information transfer, or **MTRANS**. The big difference here is that the Agent has the information, or the Topic, at the beginning and the Recipient has it at the end, but

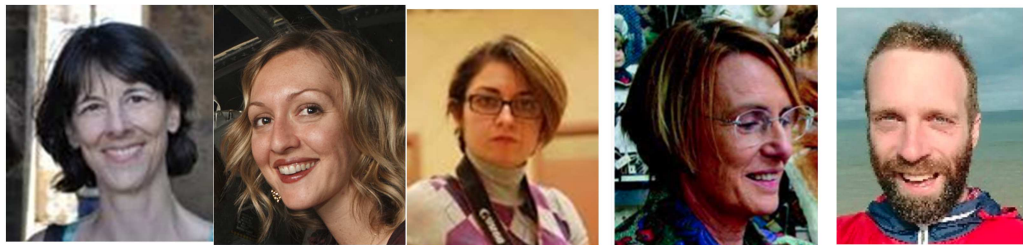


Figure 18
Susan Brown, Julia Bonn, Ghazaleh Kazeminejad, Annie Zaenen, Kevin Stowe.



Figure 19
James Pustejovsky.



Figure 20
James Gung.

the Agent also still has it. When you tell somebody something, you don't lose that piece of information, or at least not until you're about my age. Of course, we also have several classes that are the equivalent of a **PTRANS** class. For example, *Run-51.3.2* is a change-of-location class where the Theme starts in an Initial Location, then it's in **motion** and that **causes** it to end up no longer in the Initial Location but instead at a Destination. We have an automatic parser³ that provides these representations automatically, very quickly, thanks to James Gung (Gung and Palmer 2021). This requires first assigning the correct VerbNet class to the instance, but James always did like surmounting large obstacles (Figure 20).

This now goes part of the way towards scaling up my original plan for natural language understanding from Table 1. We're establishing noun phrase referents and mapping from syntactic constituents to semantic roles. We're using the verb semantics

³ <https://verbnetparser.com/>.

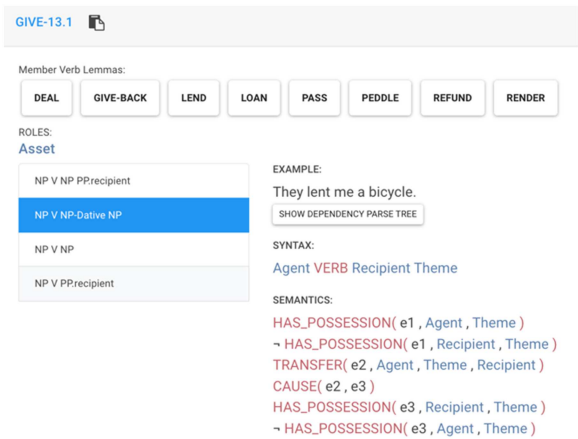


Figure 21
The VerbNet Give 12.1 class.

to drive those first two steps. Matt Gerber and Joyce Chai achieved the third step when they showed that VerbNet predicate argument structures could be used to suggest implicit roles, guiding the recovery of the implicit information (Gerber and Chai 2010). This is achieved by finding referents for the roles, similarly to pronoun reference, mirroring how it was done in PUNDIT (Palmer et al. 1986). IBM’s recent Neuro-Symbolic AI Workshop⁴ provides some hints on how these representations could also be used to draw inferences (step 4). Are we finished? Not at all. How well have we really scaled up? My dissertation had less than 20 verbs, and the PUNDIT/Kernel system less than 100. Now, with VerbNet we have 4,500 lemmas with almost 7,000 senses. That doesn’t seem too shabby until you realize that WordNet has over 9,000 English verbs, twice that number. Oh, dear, so much still to do, and it is SO HARD! I’ve collaborated with colleagues who have worked on constructing VerbNets for Arabic, Basque, Catalan, French, Spanish, and Urdu. They’ve done a wonderful job yet still have less coverage than English does, and it is a sloooow process. It requires a deep understanding of syntax AND semantics and is very, very time-consuming. There’s also been a lot of effort to try to automatically build classes that are similar to VerbNet classes using the idea of clustering verbs based on their syntactic alternations. It turns out that quite a bit of semantics is also needed. A bunch of us have been working on this since the 1990s (Merlo and Stevenson 2001; Schulte im Walde 2000, 2006; Lopez de Lacalle et al. 2014; Di Fabio, Conia, and Navigli 2019). Suzanne Stevenson and her student, Chris Persian, had some of the best results with the CHILDES database,⁵ but only by adding semantic features (Parisien and Stevenson 2010). Recent efforts include Kawahara, Peterson, and Palmer (2014), Peterson et al. (2016), Peterson and Palmer (2018), Peterson, Brown, and Palmer (2020), and Majewska et al. (2021). It feels as if we’ve been beating our heads against a brick wall for the last 30 years.

The large language models (LLMs) could be a complete game changer here. They have all the crucial semantic information about the words that is needed to complement the syntax. In fact, the syntax certainly already plays an important role in LLMs,

4 <https://research.ibm.com/events/ibm-neuro-symbolic-ai-workshop-jan2023>.
5 <https://chilides.talkbank.org/>.

whether it's latent or explicit, for them to be able to capture the similarities they capture. Let this be a challenge to a new generation of researchers, probing the LLMs to determine what kind of verb class generalizations they can make automatically. It certainly isn't necessary to match VerbNet classes or FrameNet frames (Baker et al. 1998; Baker 2014), but they could provide guidance. It's not like VerbNet is the perfect classification by any means. Every time we look at a class we go, "*Oh, did we leave a verb out? Should that verb really be there?*" It's hard to do and it should be done in a very probabilistic continuous way anyway. It should be possible to do something exceptional now.

7. Proposition Banks

In the meanwhile, if more coverage is needed than can be supplied by VerbNet, there is PropBank, which came a little later but is probably better known (Kingsbury and Palmer 2003; Palmer, Gildea, and Kingsbury 2005; Palmer, Gildea, and Xue 2010; Pradhan et al. 2022). It's the same idea of frames for semantic role labeling. The goal is a canonical predicate argument structure for the different syntactic realizations of a specific verb as well as for other verbs that are quite similar. A simple predicate logic representation of:

- (1) *When Powell met Zhu Rongji on Thursday, they discussed the return of the spy plane.*

would be:

- (2) `meet(Powell, Zhu), discuss([Powell, Zhu], return(X, plane))`

The sentence could just have easily been,

- (3) *Powell and Zhu Rongji met on Thursday and discussed the return of the spy plane.*

with the same logical representation, which would also work well for similar contexts with verbs like *consult*, *debate*, *join*, *wrestle*, *battle*. Semantic role labeling (SRL) is the process of consistently annotating the arguments that fill the semantic roles of the predicate argument structure. A key component of the PropBank annotation process is the ability to consult PropBank Frame Files, individual lexical entries for the different senses of verbs, adjectives, nominalizations, or other kinds of predicating elements.⁶ The Frame File for *discuss*, in Figure 22, says there's an Arg0 Prototypical Agent-like thing that is the *discussant*, and an Arg1 that is the *topic* (Dowty 1991). There might also be an explicit mention of a conversational partner if they haven't already been mentioned. Figure 23 shows how example (3) would be annotated, where the span of the entire conjunction is labeled as the Arg0. In both sentences the *return* event is the Arg1 and

⁶ Initially PropBank had verb entries and NomBank (Meyers et al. 2004) had eventive nouns. The entries have since been merged (O'Gorman et al. 2018a).

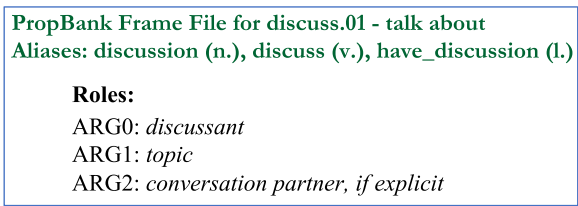


Figure 22
Frame file for “discuss.”

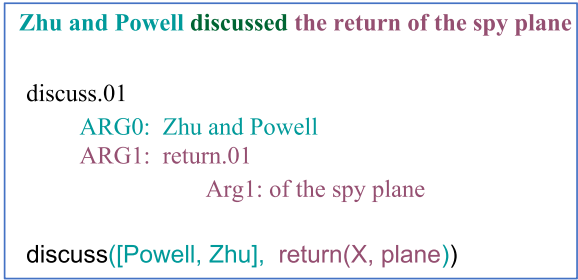


Figure 23
Semantic role labels for Ex (3).

that event in turn has its own Arg1 for *the spy plane*. The ability to consult the Frame Files greatly increases the speed and consistency of the annotation, resulting in ITA figures of 84% and above (Palmer et al. 2005).

There is an equivalent tree representation, with exactly the same information, in Figure 24. The Frame File argument structure can also be used to posit an implicit argument. We know that *return* events often have agents. This *return* event has a quite likely agent. If you know enough about the discourse context, you know that a US spy plane landed in China and Powell really wants Zhu to return it. So hopefully Zhu will be the agent of the return, as depicted in Figure 25.

There’s lots of annotated data for training and fine-tuning purposes, and at this point the Frame Files have 11,436 frameset ID entries. English PropBank has over 2M annotated tokens, Chinese over 1M, Arabic 0.5M, and Hindi/Urdu combined 0.6M.⁷ There is a small Korean PropBank, and multiple efforts in other languages, including Spanish, French, German, Basque, Catalan, and so on. For English, the SemLink resource (Stowe et al. 2021), provides mappings from PropBank frameset IDs to VerbNet, FrameNet (Baker 2014), WordNet (Miller 1995; Fellbaum 1998), and OntoNotes (Weischedel et al. 2011). IBM relied on the English PropBank annotation to project semantic role labels onto 23 other languages, as part of their Universal Proposition Bank project⁸ (Jindal et al. 2022). The Computational Linguistics group in Prague have a similar tectogrammatical approach which they applied first to Czech (Sgall et al. 1986) and then mapped to English (Mikulová et al. 2006; Hajič et al. 2020). They’re adding German and Spanish. They also have links to these other resources and are moving toward multilingual event

⁷ <https://proppbank.github.io/>.
⁸ <https://universalpropositions.github.io/>.

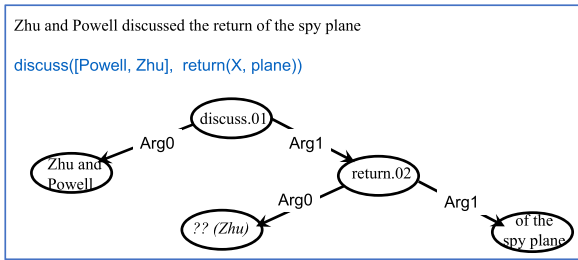


Figure 24
Our representation situated w/re context.

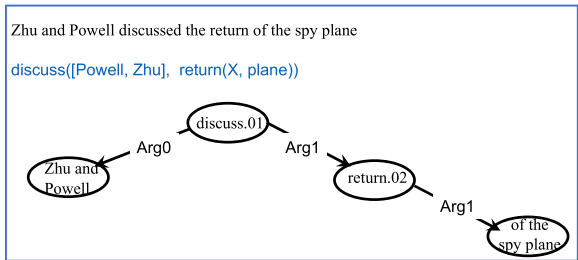


Figure 25
A proposition as a tree.



Figure 26
Contributors to PropBanks.

ontology.⁹ I owe so much to so many people who have contributed to all of these resources (see Figure 26). Working on verb semantics in multiple languages has been an incredible privilege. The job is already a blessing since the primary obligation is to work on interesting problems. However, since I don’t speak any of these languages myself, I have been very reliant on others. This has spawned interactions with intelligent, motivated, thoughtful people, both students and faculty, from all over the world. It has provided an opportunity to learn about their languages and their cultures—a constant delight—and by far the best part of the job. Even better is getting to come to ACL conferences and seeing all of these great collaborators in person again.

⁹ <https://ufal.mff.cuni.cz/synsemclass>.

8. Abstract Meaning Representations

Prop Bank is also the basis for Abstract Meaning Representations (AMRs), a joint NSF project that started as a collaboration between Kevin Knight at ISI, Dan Gildea at Rochester, Nianwen Xue at Brandeis, Kathy McKeown at Columbia, and Jim Martin and me at Colorado. When it shifted to DARPA funding, LDC and CMU joined the fray and the bulk of the annotation was done by an outstanding group in Romania, thanks to Daniel Marcu. The most recent LDC release has over 60,000 sentences with AMR annotations (Banarescu et al. 2013; Bonial et al. 2018; O’Gorman et al. 2018b). A subset of it has been automatically translated into Italian, Spanish, German, and Mandarin Chinese. The field is now trying to decide if English AMRs can be projected onto other languages similarly to the way IBM projected English PropBank.

For the differences between PropBank SRL and an AMR, see the tree in Figure 27, which illustrates how AMR drops determiners and function words, and adds Named Entity tags and the Wikipedia links.

AMR also provides more structure for the *return of the spy plane*. We now know that the plane is an Arg0 of a *spying* event. We can also recover our implicit Arg0 argument from the PropBank frameset, and, using intra-sentential coreference, link it to the first mention of *Zhu* in the sentence. The reentrancy for the reference to *Zhu* is what makes the AMR a graph, as seen in Figure 28. (For more information about the annotation of implicit arguments and multi-sentence coreference for AMR, see O’Gorman et al. [2018b] and O’Gorman [2019]). In summary, AMRs can be thought of as more abstract labeled semantic dependency trees without function words. Many of the nouns and adjectives have predicate argument structures as well as the verbs. They have Named Entity tags with Wikilinks. There are abstract discourse relations like the Penn Discourse

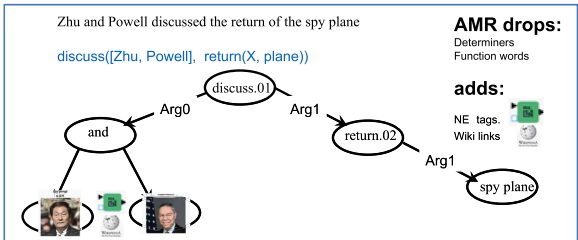


Figure 27
An Abstract Meaning Representation as a tree.

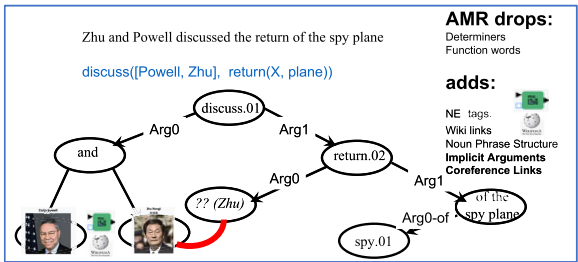


Figure 28
An Abstract Meaning Representation as a graph.

Treebank relations, a partial interpretation of modality and negation, and a few implicit arguments and relations. The PropBank frameset IDs provide the previously mentioned links to VerbNet, FrameNet, and OntoNotes. The equivalence relations for coreference are what make it a graph, a directed acyclic graph. Each tree is an individual graph, but the coreference links along with causal and temporal relations between events create connections between the individual sentences. In this way a document becomes a forest of trees or graphs that then becomes a rich connected knowledge graph—also known as a semantic net!

9. Progress with Natural Language Understanding Applications

Going back to the original plan of our 5 steps for natural language understanding in Table 1, how much progress have we made? We are establishing a lot of the noun phrase references (step 1). We’re still doing the semantic role, labeling (step 2), and recovering implicit roles (step 3). We’re drawing some inferences and we’re partially situating with respect to the discourse context (steps 4 and 5), although those two steps are far from complete. With the goal of advancing progress on steps 4 and 5, we have a current DARPA project to map the PropBank Frameset IDs to Wikidata items, which will in turn make the Wikidata inheritance relations accessible from AMRs. We’ve already mapped all of LDC’s Named Entity and relation types to Wikidata (Spaulding et al. 2023) as well as a set of temporal relations. This is all in the public domain, in a JSON file that is downloadable from a GitHub.¹⁰ Heng Ji, Zoey Li, and other outstanding colleagues at UIUC have used this resource to bootstrap a general-purpose event detection system for 3,500 distinct event types and their associated Wikidata items (Zhan et al. 2023). AMRs are also being used to excellent effect in a medical informatics joint project led by Guergana Savova, thanks to the strenuous efforts of Kristin Wright-Bettner, Skatje Myers, Jon Cai, and Jim Martin. We’ve annotated about 6,000 sentences from electronic patient records on colon cancer with AMRs. The parser trained on this data gets a Smatch score of over 80% (2 papers under review).

We’ve also been working on dialogues, very much still a wild frontier for natural language processing. Julia Hockenmaier and Anjali Narayan-Chen used Minecraft to create a blocks world environment where an architect and a builder could chat about building structures with blocks (Narayan-Chen et al. 2019). This enabled the collection of dozens of dialogues that have now been annotated with AMRs. The AMR annotations required explicit Frame Files for spatial relations, resulting in the need to either create or revise almost 200 Frame Files (Bonn et al. 2020). We now have about 25,000 annotated sentences and others are working with us on adding coreference and discourse relations.

We also have an educational application where we’re planning to use AMRs. We’re trying to put a social collaborative AI partner into student breakout groups of 10- to 13-year-olds in classrooms.¹¹ We will need to explain to the students *why* our interactive partner says and does everything that it says and does, as well as to their parents and teachers (Cao et al. 2023; Cai et al. 2023). Explainability and transparency are crucial to the success of this project.

¹⁰ <https://github.com/e-spaulding/xpo>.

¹¹ <https://www.colorado.edu/research/ai-institute/>.

10. Uniform Meaning Representations

These are exciting projects, but once more, the focus is all on English, English AMRs, and English applications. There are still all of these other languages. Which is why here at Colorado we are also working on another NSF project with Brandeis and the University of New Mexico on Uniform Meaning Representations (UMRs),¹² cross-lingual AMRs (Van Gysel et al. 2021). We're looking very carefully at the English guidelines and formats for AMRs and how we can adapt them to make them more cross-lingually effective (Xue et al. 2014). We're making sure they are suitable for low resource languages such as Arapaho, a polysynthetic language. Other languages we are considering include Kukama, English, Chinese, Hindi, Arabic, Spanish, Sanapaná, Hua, and Czech. This is why it has been so important to work with a linguistic typologist like Bill Croft, who is helping us ensure that our revised guidelines will be generally applicable. We're also adding number agreement, aspect and modality, and logical form so that the UMRs will provide a solid basis for generation and for reasoning. Uniform Meaning Representations are intended to provide a lightweight, flexible, cross-linguistically general format that can capture figurative language, implicit arguments, temporal and causal relations, rich spatial relations, logical form, aspect, and modality both within and across sentences. The thinking is that by creating UMRs for all these different languages with the same format, we will provide scaffolding for the task of mapping between different languages, improving our ability to bootstrap LLM applications for low resource languages.

11. Remaining Challenges

We are at least making progress on scaling up and expanding to multiple languages. However, there is still a lot of work to do. This has been something of a history tour of the field of NLP, but these symbolic representations are not merely of historical interest. In spite of the exciting performance of GPT4, there is still a need for explainability, transparency, and replicability. Neither are LLMs infallible. Let me first refer you to Yejin Choi's excellent TED talk, where she points out some of the commonsense failings of LLMs.¹³ Here's another example, thanks to Susan Brown, Annie Zaenen, and Felix Zhang, of an odd language interpretation made by GPT4: Given, *John mowed the lawn for 30 minutes*, and then the question *Is the lawn completely mown?*, GPT4 will answer *Yes*. That is quite likely, if the lawn isn't too large, given that John has stopped mowing. However, when told that the lawn is an acre, or even a square mile, GPT4 will still assume the lawn is completely mown after 30 minutes and answer *Yes*. This is no longer at all possible but knowing that requires common sense and an understanding of the significance of Dowty's Incremental Themes, that determine whether a task is completed (Dowty 1991). English is, after all, a human language, right? It's our language. So, we're allowed to say how we think it should be interpreted.

I don't have another 50 years. I am passing the baton to all of you. Natural language understanding is not solved, and it is such a hard problem that we need every tool in the toolbox. We can't afford to throw anything away. If we can use rich symbolic lexical resources with deep semantic representations to improve the ability of large language models to see the implications, to draw the right conclusions, to exercise common sense,

¹² <https://github.com/umr4nlp/references>.

¹³ <https://www.youtube.com/watch?v=SvBR00GT5VI>.



Figure 29
Boulder, Colorado. Susan Brown, Mans Hulden, Maria Pacheco, Alexis Palmer, Jim Martin, Wayne Ward, Katharina Kann, the University of Colorado, a CLASIC/CLEAR Open House, Derek Palmer, Martha Palmer, Neil Palmer, my birthday, the Rockies.

why wouldn't we do that? Is there some rule that says no one is allowed to use these kinds of resources with large language models? Will neuro-symbolic approaches get us where we want to go, or do we need something brand new that no one has even thought of yet? There are a lot of wonderful problems out there. You guys are going to have a great time.

A few last thoughts. I loved the order and predictability of logic and math to begin with. That's why I wanted to do a geometry theorem prover. But I learned that I loved the subtleties, the idiosyncrasies, and the mysteries of language even more. With semantics, you're never sure you're right.

This means we have to stand in an inconspicuous mysterious place, a place where we are not sure that we're sure—where we are comfortable knowing that we do not know very much at all.—Richard Rohr¹⁴

Or as Aravind Joshi put it more succinctly, in his talk upon receiving the Benjamin Franklin Medal 2005,¹⁵

The mystery of language is big enough to keep you awake a long time.

With respect to LLMs,

Knowledge is proud that it knows so much, wisdom is humble that it knows no more.
—William Cowper

14 <https://cac.org/daily-meditations/praying-for-wisdom-2021-12-28/>.
15 <https://www.fi.edu/en/awards/laureates/aravind-k-joshi>.

My big, fat 50-year journey,
Recent and current collaborations marked with★



Figure 30
The journey.

Wisdom might be the difference between people and large language models.

Finally, when I was trying to decide, in Texas, if I should even go to graduate school, this is what one of my philosophy professors told me,

Follow your heart and in the end, you will find you have come to the place you want to be.

My last move was from Philadelphia to Colorado in 2005 (Figures 29, 30). And you know what, he was right. It's true!

Acknowledgments

In addition to the many outstanding mentors, colleagues, postdocs, and students that I have already thanked above, I want to express my very deep appreciation and gratitude to the steady stream of thoughtful, helpful, and dedicated program managers from NSF, DARPA, DTRA, and NIH who have funded this work, as well as to Lockheed Martin, Google, iPSoft, and Anthem who also provided funding. Without this funding, very little of this would have been accomplished. I would especially like to thank Tanya Korelsky, Joe Olive, and Boyan Onyshkevych for their guidance, their unwavering faith in me, and their friendship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of NSF,

DARPA, DTRA, NIH, or the U.S. government or any of the companies.

References

- Baker, C. F., C. J. Fillmore, and J. B. Lowe. 1998. The Berkeley FrameNet project. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, 1:86–90. <https://doi.org/10.3115/980845.980860>
- Baker, Collin F. 2014. FrameNet: A knowledge base for natural language processing. In *Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929–2014)*, pages 1–5. <https://doi.org/10.3115/v1/W14-3001>
- Banarescu, Laura, 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*, held at ACL-2013.
- Bonial, Claire, Bianca Badarau, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Tim O’Gorman, Martha Palmer, and Nathan Schneider. 2018. Abstract Meaning representation of constructions: The more we include, the better the representation. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- Bonn, Julia, Martha Palmer, Jon Cai, and Kristin Wright-Bettner. 2020. Spatial AMR: Expanded spatial annotation in the context of a grounded Minecraft corpus. In *Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC 2020)*.
- Brown, S., J. Bonn, G. Kazeminejad, A. Zaenen, and J. Pustejovsky, and M. Palmer. 2022. Semantic representations for NLP using VerbNet and the generative lexicon. *Frontiers of Artificial Intelligence*. <https://doi.org/10.3389/frai.2022.821697>, PubMed: 35493615
- Cai, Jon Z., Brendan King, Margaret Perkoff, Shiran Dudy, Jie Cao, Marie Grace, Natalia Wojarnik, Ananya Ganesh, James H. Martin, Martha Palmer, Marilyn Walker, and Jeffrey Flanigan. 2023. Dependency dialogue acts—Annotation scheme and case study. In *13th International Workshop on Spoken Dialogue Systems Technology*.
- Cao, Jie, Rachel Dickler, Marie Grace, Jeffrey B. Bush, Alessandro Roncone, Leanne M. Hirshfield, Marilyn A. Walker, and Martha S. Palmer. 2023. Designing an AI partner for jigsaw classrooms. In *Language-Based AI Agent Interaction with Children@IWSDS’23*.
- Dahl, Deborah, Martha Palmer, and Rebecca Passonneau. 1987. Nominalizations in PUNDIT. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, 6 pages. <https://doi.org/10.3115/981175.981194>
- Dang, Hoa Trang, Karin Kipper, Martha Palmer, and Joseph Rosenzweig. 1998. Investigating regular sense extensions based on intersective Levin classes. In *Proceedings of Coling/ACL-98*, 36th Association of Computational Linguistics Conference, pages 293–300. <https://doi.org/10.3115/980845.980893>
- Dang, Hoa Trang, Karin Kipper, and Martha Palmer. 2000. Integrating compositional semantics into a verb lexicon. In *Eighteenth International Conference on Computational Linguistics, COLING-2000*, 8 pages. <https://doi.org/10.3115/992730.992795>
- Di Fabio, Andrea, Simone Conia, and Roberto Navigli. 2019. VerbAtlas: A novel large-scale verbal semantic resource and its application to semantic role labeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 627–637. <https://doi.org/10.18653/v1/D19-1058>
- Dowty, David. 1991. Thematic proto-roles and argument selection. *Language*, 67(3):547–619. <https://doi.org/10.1353/lan.1991.0021>
- Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. Bradford Books. <https://mitpress.mit.edu/9780262561167/>, <https://doi.org/10.7551/mitpress/7287.001.0001>
- Gerber, Matthew and Joyce Chai. 2010. Beyond NomBank: A study of implicit arguments for nominal predicates. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1583–1592.
- Gung, James and Martha Palmer. 2021. Predicate representations and polysemy in VerbNet semantic parsing. In *14th International Conference on Computational Semantics (IWCS)*.
- Hajič, J. et al. (2020). Prague Dependency Treebank - Consolidated 1.0 (PDT-C 1.0). LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. <http://hdl.handle.net/11234/1-3185>
- Jindal, Ishan, Alexandre Rademaker, Michał Ulewicz, Ha Linh, Huyen Nguyen, Khoi-Nguyen Tran, Huaiyu Zhu, and Yunyao Li. 2022. Universal Proposition Bank 2.0. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1700–1711.
- Kawahara, Daisuke, Daniel W. Peterson, and Martha Palmer. 2014. A step-wise usage-based method for inducing polysemy-aware verb classes. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL2014)*.

- <https://doi.org/10.3115/v1/P14-1097>
- Kazeminejad, G., M. Palmer, S. Brown, and J. Pustejovsky. 2022. Componential analysis of English verbs. *Frontiers of Artificial Intelligence*. <https://doi.org/10.3389/frai.2022.780385>, PubMed: 35707764
- Kipper Schuler, Karin. 2005. *VerbNet: A Broad-coverage, Comprehensive Verb Lexicon*. Ph.D. dissertation, University of Pennsylvania.
- Kipper, Karen, Hoa Trang Dang, and Martha Stone Palmer. 2000. Class-Based construction of a verb lexicon. *AAAI/AAI*, 2000:691–696.
- Kingsbury, Paul and Martha Palmer. 2003. PropBank: The next level of TreeBank. In *The Second Workshop on Treebanks and Linguistic Theories*, (TLT 2003).
- Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press.
- Lopez de Lacalle, Maddalen, Egoitz Laparra, and German Rigau. 2014. Predicate matrix: Extending SemLink through WordNet mappings. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 903–909.
- Majewska, Olga, Charlotte Collins, Simon Baker, Jari Björne, Susan Windisch Brown, Anna Korhonen, and Martha Palmer. 2021. BioVerbNet: A large semantic-syntactic classification of verbs in biomedicine. *Journal of Biomedical Semantics*, 12(1):12. <https://doi.org/10.1186/s13326-021-00247-z>, PubMed: 34266499
- Mel'čuk, Igor. 1988. *Dependency Syntax: Theory and Practice*. The SUNY Press.
- Merlo, Paola and Suzanne Stevenson. 2001. Automatic verb classification based on statistical distributions of argument structure. *Computational Linguistics*, 27(3):373–408. <https://doi.org/10.1162/089120101317066122>
- Meyers, Adam, Ruth Reeves, Catherine Macleod, Rachel Szekely, Veronika Zielinska, Brian Young, and Ralph Grishman. 2004. Annotating Noun Argument Structure for NomBank. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*. <https://api.semanticscholar.org/CorpusID:12699810>
- Mikulová, M., A. Bémová, J. Hajič, E. Hajičová, J. Havelka, V. Kolářová, L. Kučová, M. Lopatková, P. Pajas, J. Panevová, M. Sevcikova, P. Sgall, J. Štěpánek, Z. Uresová, K. Veselá, and Z. Žabokrtský. 2006. Annotation on the tectogrammatical level in the Prague Dependency Treebank. Annotation manual. Technical Report 30. Institute of Formal and Applied Linguistics, Charles University, Prague, Czech Republic. https://ufal.mff.cuni.cz/pdt-c/publications/tr_en_def.pdf; https://ufal.mff.cuni.cz/pdt-c/publications/tr_en_def.pdf; <https://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/t-layer/html/index.html>
- Miller, G. A. 1995. Wordnet: A lexical database for English. *Communications of the ACM*, 38:39–41. <https://doi.org/10.1145/219717.219748>
- Narayan-Chen, Anjali, Prashant Jayannavar, and Julia Hockenmaier. 2019. Collaborative dialogue in Minecraft. In *Proceedings of ACL*. <https://doi.org/10.18653/v1/P19-1537>
- O'Gorman, Tim, Sameer Pradhan, Julia Bonn, Katie Conger, James Gung, and Martha Palmer. 2018a. The new PropBank: Aligning PropBank with AMR through POS Unification. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- O'Gorman, Tim, Michael Regan, Kira Griffitt, Martha Palmer, Ulf Hermjakob, and Kevin Knight. 2018b. AMR beyond the sentence: The Multi-sentence AMR corpus. In *Proceedings of the International Conference on Computational Linguistics (COLING 2018)*, Santa Fe, NM.
- O'Gorman, Timothy J. 2019. *Bringing Together Computational and Linguistic Models of Implicit Role Interpretation*. Ph.D. dissertation. University of Colorado at Boulder.
- Palmer, Martha. 1981. A case for rule-driven semantic analysis. In *Proceedings of the 10th Annual Meeting of the Association for Computational Linguistics*, 6 pages. <https://doi.org/10.3115/981923.981958>
- Palmer, Martha. 1983. Inference-driven semantic analysis. In *Proceedings of the American Association for Artificial Conference*, 6 pages.
- Palmer, Martha. 1990a. Customizing verb definitions for specific semantic domains. *Machine Translation*, 5:45–62. <https://doi.org/10.1007/BF00310040>
- Palmer, Martha. 1990b. *Semantic Processing for Finite Domains*. ACL Book Series.

- Cambridge University Press. <https://doi.org/10.1017/CB09780511554414>
- Palmer, Martha, Carl Weir, Rebecca Passonneau, and Tim Finin. 1993. The kernel text understanding system. *Artificial Intelligence*, 63:17–68. Special Issue on Text Understanding. [https://doi.org/10.1016/0004-3702\(93\)90014-3](https://doi.org/10.1016/0004-3702(93)90014-3)
- Palmer, Martha, Dan Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–105. <https://doi.org/10.1162/0891201053630264>
- Palmer, Martha, Daniel Gildea, and Nianwen Xue. 2010. *Semantic Role Labeling*. Synthesis Lectures on Human Language Technology Series, edited by Graeme Hirst. Morgan & Claypool. <https://doi.org/10.1007/978-3-031-02135-0>
- Palmer, Martha, Deborah Dahl, Rebecca [Schiffman] Passonneau, Lynette Hirschman, Marcia Linebarger, and John Dowding. 1986. Recovering implicit information. In *Proceedings of the Sixth Annual Meeting of the Association for Computational Linguistics*. <https://doi.org/10.3115/981131.981135>
- Palmer, Martha, Tim Finin, and Sharon Walters. 1990. Workshop on the evaluation of natural language processing systems. *Computational Linguistics*, 16(3):175–181.
- Parisien, C., and S. Stevenson. 2010. Learning verb alternations in a usage-based Bayesian model. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (CogSci 2010).
- Passonneau, Rebecca, Carl Weir, Tim Finin, and Martha Palmer. 1991. Integrating natural language processing and knowledge based processing: Demand driven reasoning for document analysis. In *Proceedings of the 1991 National Conference on Artificial Intelligence* (AAAI-91), 6 pages.
- Peterson, Daniel, Jordan Boyd-Graber, Martha Palmer, and Daisuke Kawahara. 2016. Leveraging VerbNet to build corpus-specific verb clusters. In **SEM 2016, w/ ACL 2016*. <https://doi.org/10.18653/v1/S16-2012>
- Peterson, Daniel, Susan Brown, and Martha Palmer. 2020. Verb class induction with partial supervision. In *Proceedings of the Thirty-fourth AAAI Conference on Artificial Intelligence*. <https://doi.org/10.1609/aaai.v34i05.6385>
- Peterson, Daniel W. and Martha Palmer. 2018. Bayesian verb sense clustering. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence* (AAAI-18). <https://doi.org/10.1609/aaai.v32i1.12023>
- Pradhan, Sameer, Julia Bonn, Skatje Myers, Kathryn Conger, Tim O’Gorman, James Gung, Kristin Wright-Bettner, and Martha Palmer. 2022. PropBank comes of age—Larger, smarter, and more diverse. In *Proceedings of the 11th Joint Conference on Lexical and Computational Semantics*, pages 278–288. <https://doi.org/10.18653/v1/2022.starsem-1.24>
- Pustejovsky, James. 1991. The generative lexicon. *Computational Linguistics*, 17(4):409–441.
- Schulte im Walde, Sabine. 2000. Clustering verbs semantically according to their alternation behaviour. In *Proceedings of COLI: 18th International Conference on Computational Linguistics*. <https://doi.org/10.3115/992730.992754>
- Schulte im Walde, Sabine. 2006. Experiments on the automatic induction of German semantic verb classes. *Computational Linguistics*, 32(2):159–194. <https://doi.org/10.1162/coli.2006.32.2.159>
- Sgall, P., E. Hajičová, and J. Panevová. 1986. The meaning of the sentence in its semantic and pragmatic aspects 1 ed. Dordrecht: Elsevier, Prague: Academia, ix, pages 353.
- Simmons, Robert F. and Bertram C. Bruce. 1971. Some relations between predicate calculus and semantic net representations of discourse. In *Proceedings of IJCAI’71: Proceedings of the 2nd International Joint Conference on Artificial Intelligence*, pages 524–530.
- Simmons, R. F. 1973. Semantic networks: Their computation and use for understanding English sentences. In Roger C. Schank and Kenneth Mark Colby, editors, *Computer Models of Thought and Language*. W.H. Freeman and Company, pages 63–113.
- Spaulding, et al. 2023. The DARPA Wikipedia Overlay Wikidata as an ontology for natural language processing. In *Proceedings of the 19th Joint ACL-ISO Workshop on Interoperable Semantic Annotation*, pages 1–10, Held in conjunction with IWCS 2023.
- Stowe, Kevin, Jenette Preciado, Kathryn Conger, Susan Windisch Brown, Ghazaleh Kazeminejad, James Gung, and Martha Palmer. 2021. SemLink 2.0: Chasing lexical resources. In *14th International Conference on Computational Semantics* (IWCS).

- Van Gysel, Jens, Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Tim O’Gorman, Andrew Cowell, William Croft, Chu-Ren Huang, Jan Hajić, James H. Martin, Stephan Oepen, Martha Palmer, James Pustejovsky, Rosa Vallejos, and Nianwen Xue. 2021. Designing a uniform meaning representation for natural language processing. *Künstliche Intelligenz*. <https://doi.org/10.1007/s13218-021-00722-w>
- Weischedel, Ralph, Eduard Hovy, Mitchell Marcus, Martha Palmer, Robert Belvin, Sameer Pradan, Lance Ramshaw, and Nianwen Xue. 2011. OntoNotes: A large training corpus for enhanced processing. In Joseph Olive, Caitlin Christianson, and John McCary, editors, *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation*. Springer, pages 54–63.
- Winograd, Terry. 1970. AI Technical Report 235, MIT.
- Wu, Zhibiao and Martha Palmer. 1994. Verb semantics and lexical selection. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 133–139. <https://doi.org/10.3115/981732.981751>
- Xue, Nianwen, Ondrej Bojar, Jan Hajic, Martha Palmer, Zdenka Uresova, and Xiuhong Zhang. 2014. Not an interlingua, but close: Comparison of English AMRs to Chinese and Czech. In *9th Edition of the Language Resources and Evaluation Conference*.
- Zhan, Qiusi, Sha Li, Kathryn Conger, Martha Palmer, Heng Ji, and Jiawei Han. 2023. GLEN: General-purpose event detection for thousands of types. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 2823–2838.