The Illinois Intentional Tort Qualitative Dataset

Joseph BLASS [a,1] and Kenneth FORBUS [a]
[a] *Qualitative Reasoning Group, Northwestern University*

Abstract. We introduce the Illinois Intentional Tort Qualitative Dataset, a set of Illinois Common Law cases in Assault, Battery, Trespass, and Self-Defense, machine-translated into qualitative predicate representations. We discuss the cases involved, the natural language understanding system used to translate the cases into predicate logic, and validation measures that serve as performance baselines for future AI research using the dataset.

Keywords. Legal Corpus, Predicate Logic, Tort Law

1. Introduction and Background

In Common Law legal systems, cases are resolved by reference to prior cases involving similar claims. Datasets of such cases, and their developers' formal commitments and assumptions, have been important to AI & Law research. We present a dataset of case texts translated into predicate logic using a publicly available natural language understanding system.

Several large datasets collect cases in their original language [1, 2], and researchers using large-scale machine-learning techniques have recently developed legal retrieval and reasoning systems that operate over raw text [3]. But many AI legal reasoning systems require formal machine-interpretable representations, with which researchers have had to annotate their cases. The most widely-used formalism is *factors* [4], legally-salient principles relevant to finding case outcomes and that favor a party. An expert identifies the total list of factors in a domain, then cases are tagged with them (often by hand, but automatically as well [5]). Originally developed for use in HYPO-style reasoners [4], factors and similar formalisms continue to be used in a variety of AI & Law techniques. For example, Horty's *reason model* learns and applies defeasible rules from factored cases [6]. *Abstract dialectical frameworks* encode relationships between arguments and outcomes to generate arguments in factored cases [7]. And Verheij's *case models* are logically consistent sets of cases in propositional logic – not factors, since the statements do not favor a side – that together encode rules [8].

Symbolic case representations have indeed long been useful to AI & Law research. They are a natural fit for legal reasoning, because rationality, explainability, consistency, and transparency of reasoning are hallmarks of good *human* legal reasoning. But in an era of big data and large language models, it is increasingly uncomfortable to rely on human annotation. Semantic interpreters that translate text into symbolic representations can split the difference and scale beyond hand-encoding while providing rich representations for legal reasoning.

---

[1] Corresponding Author: Joseph Blass, joeblass@law.northwestern.edu

2. The Companions Natural Language Understanding System (CNLU)

CNLU [9] is the semantic interpreter for the Companions cognitive architecture [10]. Companions use an ontology derived from Cyc [11]. The knowledge base, NextKB, supplements OpenCyc and contains over 18,000 concepts and over 1000 relations constrained by nearly a million facts. Microtheories partition knowledge and can be linked to form logical environments. NextKB meets the ontological requirements described in [12].

CNLU produces hierarchical parse trees using Allen's bottom-up chart parser [13]. At the leaf nodes (lexical tokens), subcategorization frames are used to generate choice sets of nodes' possible semantic interpretations. Consistent sets are selected to create sentence interpretations, manually or automatically [14]. Coreference resolution merges references to the same entity. CNLU operates over simplified English syntax to focus on semantic over syntactic breadth, that is, to prioritize the system's understanding of ideas over sentence forms. CNLU uses Discourse Representation Theory [15] to handle negation, implication, quantification, and counterfactuals using nested discourse representation structures. Those structures are converted to standard CycL representations in the last language processing step.

The lexical information was semi-automatically extracted from a public domain edition of Webster's dictionary, then augmented. Each word, for each part of speech, has one or more semantic translations derived from FrameNet [16] that map the lexicon to the NextKB ontology to express possible meaning. Syntactic analysis maps complements to role relations. For composability, CNLU uses neo-Davidsonian event representation (Figure 1). Thus "Dave eats ice cream" maps Eat-TheWord to an EatingEvent, and Dave to its performer. Finally, *Narrative Functions* (NFs) [17] support interpretation with abductive explanation. NFs operate across choice sets to build parse explanations, making abductive assumptions as needed.

3. The Illinois Intentional Tort Qualitative Dataset

Illinois was chosen as the dataset's jurisdiction because our institution is located there. CNLU cannot currently handle statutes' linguistic complexity and legalistic formalism, so we sought pure common-law doctrines, where judicial opinions express rules in plain English. We included the Tort doctrines of Trespass, Assault, Battery, and Self-Defense. We found cases using WestLaw and LexisNexis and traced their cited and citing cases. Cases were excluded if unreasoned, overturned, decided before 1870, or unrelated to prior or subsequent cases.

Collected cases were organized by doctrine and annotated with decision year, court, and case reporter metadata. Case facts and conclusions were manually identified and stored as a string argument to a metadata fact. In eleven cases, appellate courts laid out alternate set of facts left ambiguous by a lower court and identified the corresponding legal conclusions.

(isa treat123 IceCream) (eats Dave treat123)	(isa treat123 IceCream) (isa eat456 EatingEvent) (doneBy eat456 Dave) (objectConsumed eat456

Figure 1. "Dave eats Ice Cream," traditional (L) vs. neo-Davidsonian event representation (R).

Original Text, Bishop v. Ellsworth: "On July 21, 1965, defendants, Mark and Jeff Ellsworth and David Gibson, three small boys, entered [the plaintiff Dwayne Bishop's] salvage yard premises at 427 Mulberry Street in Canton, without his permission, and while there happened upon a bottle partially embedded in the loose earth on top of a landfill, wherein they discovered the sum of $ 12,590 in United States currency. [The] boys delivered the money to the municipal chief of police who deposited it with [the] Canton State Bank. The defendants caused preliminary notices to be given as required by Ill Rev Stats, chapter 50, subsections 27 and 28, (1965)."

Simplified Text: "The plaintiff owns a salvage yard. The defendants are young boys. The defendants entered the salvage yard. The plaintiff did not permit them to enter the salvage yard. The defendants found a bottle containing $12590 of money in the plaintiff's salvage yard. The defendants brought the money to the chief of police. The defendants placed notices about the money in the newspaper."

Figure 2. Original vs. Simplified Case Facts. Bishop v. Ellsworth, 91 Ill. App. 2d 386 (1986)

Each of these was converted into two dataset cases, for each set of facts and conclusion. The dataset comprises 88 cases illustrating 112 distinct tort claims. These include 17 assault cases (12 positive cases where the court found an assault had occurred, 5 negative cases where the legal standard was not met), 40 battery cases (30 positive, 10 negative), 43 trespass cases (29 positive, 14 negative), and 12 self-defense cases (5 positive, 7 negative). Positive cases outnumber negatives because they are more likely to be published and later relied upon.

Judges' descriptions of case facts often include run-on sentences, long lists and descriptions, and asides, so case texts were simplified for CNLU (Figure 2). Parties' names were reduced to party designations. Names (people, places, and things), prices, and dates were removed. For cases with unrelated causes of action, facts identified as only relevant to an unrelated claim were removed. Words not in CNLU's vocabulary were replaced with synonyms (or added to the vocabulary). Longer sentences were broken down into simpler clauses, complex grammatical structures were rephrased, and compound nouns were sometimes rephrased as declarative sentences (e.g., a long sentence referring to "the plaintiff's salvage yard" became several sentences including "The plaintiff owns a salvage yard."). Texts were then processed by CNLU, with choice sets selected manually to ensure maximum fidelity.

Legal reasoning operates over complex real-world situations, so a rich, accurate understanding of legal texts is critical. NFs can infer sentence meaning beyond strict semantics. To illustrate: given the phrase "the plaintiff climbed to the balcony," CNLU might yield the facts in Figure 3: a climbing event, done by the plaintiff, ending at the balcony. Missing is the fact that *the plaintiff is now on the balcony*. Trespass involves being on private property without permission, so that missing fact is needed to understand the plaintiff may be trespassing. Similarly, understanding Assault or Battery means understanding when actions constitute threats or physical contact, but such information is so obvious to humans that it is rarely explicitly stated. NFs can make such inferences within CNLU's language processing and understanding (not post-processing), to understand what words mean, not just what they say.

```
(isa balcony4180 BalconyLevelInAConstruction)
(isa climb4579 Climbing) (isa plaintiff4574 Plaintiff)
(performedBy climb4579 plaintiff4574)
(to-Generic climb4579 balcony4180)
(isa plaintiff4574 Plaintiff)
(toLocation climb4579 balcony4180)
```

Figure 3. "The plaintiff climbed to the balcony."

We wrote NFs to make commonsense inferences for frequently recurring situations (200 NF detection rules for 93 NFs). Most were to infer (1) object locations (like the climbing example), (2) when events cause damage, (3) transitive ownership (e.g., a building's owner owns its balcony), (4) when events involve physical contact, (5) part/whole relationships, and (6) when actions create new entities (e.g., Alan suing Bob creates a lawsuit entity). To ensure we only wrote language understanding rules, not legal rules, we only wrote rules that would be true outside legal proceedings, and did not write rules to infer legal conclusions (i.e., facts the opinions indicated were true as a matter of law). Still, determining what is a commonsense versus legal inference is tricky; our use of NFs is discussed below in the limitations section.

4. Experimental and Baseline Results

Our approaches build probabilistic relational schemas of cases using a modified version of the SAGE analogical generalization engine [18, 19], and use them to reason about other cases. Schemas and ungeneralized cases are stored in a generalization pool (gpool) and reasoned with by analogy using the Structure Mapping Engine (SME, [20]). SME creates a mapping between two cases and uses it to make candidate inferences (CIs) between them. Given a case at bar, our systems make generalizations from the other cases, retrieve from the gpool using the held out case, map the retrieved case onto the new one, and check the CIs for the held out solution. Solutions are predicates saying who did what to whom, so "Carl trespassed on Dan's lawn by driving on it" might be expressed as `(trespassOnPropertyByAction Carl123 DansLawn456 drive789)`. That the system must generate, not select, an answer measures its understanding. Our techniques are examined more closely in [21].

Given a case at bar, the first technique, *Purely Analogical Precedential Reasoning* (PAPR), creates generalizations from all other cases in the same doctrine, generalizing positive cases in one gpool and negative ones in another. If it finds a legal conclusion amongst the CIs (after retrieval and mapping), it proposes it as the case solution; if not, it retrieves again. To measure the system's ability to solve the case given its case base, not its ability to retrieve a good case on the first try, we have it check its work: if its proposed solution is wrong, it retrieves and checks up to six additional mappings, measuring Precision@6.

Our second technique, *Analogical Reasoning with Positive Generalizations* (ARPG), reflects the fact that it is positive cases that encode legal doctrines: negative cases may have in common only the absence of positive case facts. ARPG depends on the assumption that legal cases have sufficiently different legally-irrelevant facts that schemas of positive examples will encode only legally-relevant facts. ARPG retrieves from only positive generalizations (not ungeneralized examples) and inspects the CIs. If they contain a case conclusion,

Table 1. Performance of 4 baselines on Illinois Intentional Tort Qualitative Dataset, in % Accuracy

Technique	All Cases	Assault	Battery	Trespass	Self-Defense	Positive	Negative
ARPG	35%	35%	25%	44%	-	10%	97%
PAPR	72%	94%	68%	67%	-	75%	66%
legalBERT	33%	53%	36%	23%	42%	33%	35%
GPT-J	52%	35%	50%	61%	25%	62%	28%

Table 2. Different techniques' performances compared using proportion tests. Significance reported at $p < 0.05$.

Method	GPT-J	legalBERT	PAPR (part'l credit)
ARPG	GPT-J performs better overall, on Battery and Trespass, and on Pos cases. ARPG performs better on Neg cases.	No sig. diff overall. ARPG outperforms legalBERT on Trespass and Neg cases; legalBERT beats ARPG on Pos cases.	PAPR outperforms ARPG overall, by doctrine, and on Pos cases; ARPG outperforms PAPR on Neg cases.
PAPR (part'l credit)	PAPR outperforms GPT-J, overall and on Assault cases and Neg cases.	PAPR outperforms legalBERT, overall and when broken down by Positive v Negative cases.	-
legalBERT	GPT-J outperforms legalBERT overall and on battery, assault, and positive cases.	-	-

ARPG checks for *other* CIs. Extra CIs mean some schema fact besides the conclusion is missing in the case, so the case is negative and the absent fact is a missing claim element. If the conclusion is the only CI, the case is positive. ARPG is also evaluated using Precision@6.

Both techniques can test whether the system is partway towards an answer by accepting a conclusion CI in which all but one entities are correct. In a partially-true conclusion CI, the system identifies the tortfeasor and *either* the tortious action or the victim. Here we report performance for ARPG with a strict truth test and PAPR with a partial truth test, our floor and ceiling performances in [21]. Finally, we tested our systems only on Trespass, Assault, and Battery; we are still studying modeling affirmative defenses like Self-Defense.

We report two baselines using off-the-shelf ML techniques. LegalBERT is a BERT model specialized on legal texts [22]. It was tested as a multiple-choice system. We created multiple-choice cases by negating conclusions and reversing party roles. GPT-J is an open-source model based on OpenAI's GPT-3 [23]. GPT-J was fine-tuned and tested on our dataset using holdout and 5-ply cross-fold validation. To test GPT-J, we prompted it with the simplified case facts and had it generate six text completions, which we examined for a simple expression of the conclusion. Performances are reported in Table 1 and compared in Table 2.

5. Discussion and Limitations

We discuss implications the analysis reveals about the dataset; methods are discussed in [21].

A greater number of positive than negative cases, as well as cases where the defendant is the one accused of tortious conduct, may lead statistical methods to accuse the defendant and generally perform well. Indeed, statistical methods outperformed ARPG in positive but not negative cases. To assess a reasoning technique's performance, special attention should be paid to negative cases and those where someone other than the defendant is the accused.

The limitations of CNLU and of the process by which commonsense NF rules were generated must be acknowledged. The dataset does not yet reach the goal of being generated by feeding raw legal text into a language understanding system, because no system can reliably both handle the complexity of legal texts and generate accurate symbolic representations from them. CNLU features three limitations, each of which is an area of active research. First, CNLU still relies on a human simplifying the original text: the complexity of surface forms CNLU can handle has progressed, but it cannot yet handle arbitrarily complex English input. A stopgap solution may be to train a large language model to simplify texts to CNLU's level. Second, for this dataset CNLU's choice sets were manually selected to ensure semantic fidelity to the text. CNLU can automatically select choice sets quite well [14, 24]; we hand-selected because our goal was to create an accurate dataset, not to evaluate CNLU. Third, something like NFs are necessary to express what a text means, not just what it literally says. Common-sense reasoning is a persistent problem in AI research. For now, the options are to create generally-applicable rules, or to accept that facts obvious to humans remain unknown. Because such facts are critical to understanding legal cases and their outcomes, leaving them unknown guarantees that a computer system will either fail to learn legal concepts or will learn the wrong ones. We invite disagreement and discussion on this point.

References

[1] Petrova A, Armour J, Lukasiewicz T. Extracting Outcomes from Appellate Decisions in US State Courts. Proc's of JURIX Conference 2020; p. 133-142.
[2] Grabmair M, Ashley K, Chen R, Sureshkumar P, Wang C, Nyberg E, Walker, V. Introducing LUIMA: an exp't in legal conceptual retrieval of vaccine injury decisions using a UIMA type sys. & tools. ICAIL 2015.
[3] Branting K, Weiss B, Brown B, Pfeifer C, Chakraborty A, Ferro L, Yeh, A. Semi-supervised methods for explainable legal prediction. In: Proc's of 17th ICAIL; 2019; p. 22-31.
[4] Ashley, KD. Toward a computational theory of arguing with precedents. In: Proc's of ICAIL; 1989; p. 93-102.
[5] Bruninghaus S, Ashley KD. Predicting Outcomes of Case Based Legal Arguments. ICAIL; 2003; p. 233-242.
[6] Horty JF. Reasons and Precedent. In: Proc's of 13th ICAIL; 2011; p. 41-50.
[7] Al-Abdulkarim L, Atkinson K, Bench-Capon T, Whittle S, Williams R, Wolfenden C. Noise induced hearing loss: Building an application using the ANGELIC methodology. Argument & Comp., 2019; 10(1):5-22.
[8] Verheij B. Formalizing Arguments, Rules and Cases. In: Proc's of ICAIL; 2017; p. 199-208.
[9] Tomai E, Forbus K. EA NLU: Practical Language Understanding for Cog. Modeling. Proc's of FLAIRS; 2009.
[10] Forbus K, Hinrichs T, Crouse M, Blass, J. Analogies vs Rules in Cog. Arch. In: Adv. in Cog. Sys.; 2020.
[11] Lenat D. CYC: A large-scale investment in knowledge infrastructure. Comms of ACM; 1995; 38(11):33-38.
[12] Ashley KD. Ontological req's for analogical, teleological, and hypothetical legal reas'g. ICAIL; 2009; p.1-10.
[13] Allen JF. Natural Language Understanding. (2nd ed). 1994; Addison Wesley; Redwood City, CA
[14] Barbella D, Forbus K. Analogical word sense disambiguation. In: Adv. in Cog Sys 2013; 2(1):297.
[15] Kamp H, Reyle U. From discourse to logic: Intro to model-theor'c semantics of nat. lang. 1993; Kluwer Acad.
[16] Ruppenhofer J, Ellsworth M, Schwarzer-Petruck M, Johnson CR, Scheffczyk J. FrameNet II: Extended theory and practice. International Computer Science Institute; 2016.
[17] McFate C, Forbus K, Hinrichs T. Using narrative function to extract qualitative information from natural language texts. In: Procs of the AAAI Conf. on AI; 2014; p. 373-379.
[18] Kuehne S, Forbus K, Gentner D, Quinn B. SEQL: Category learning as progressive abstraction using structure mapping. In: Procs of 22nd Mtg. of Cog Sci Society; 2000; p. 770.
[19] Blass J, Forbus K. Conclusion-Verified Analogical Schema Induction. In: Adv. in Cog. Systems; 2022.
[20] Forbus KD, Ferguson RW, Lovett A, Gentner D. Extending SME to handle large-scale cognitive modeling. Cognitive Science 2017; 41(5):1152-1201.
[21] Blass J, Forbus K. Analogical Reas'g, General'n, & Rule Learning for Common Law Reasoning. In prep (draft manuscript available upon request).
[22] Chalkidis I, Fergadiotis M, Malakasiotis P, Aletras N, Androutsopoulos I. LEGAL-BERT: The muppets straight out of law school. arXiv preprint arXiv:2010.02559; 2020.
[23] Floridi L, Chiriatti M. GPT-3: Its nature, scope, limits, and conseq's. Minds and Machines 2020; 30(4).
[24] Ribeiro D, Forbus K. Combining analogy with language models for knowledge extraction. AKBC Conf. 2021.