Legal Text Interpretation: Identifying Hohfeldian Relations from Text

Wim Peters¹ and Adam Wyner²

¹Department of Computer Science, University of Sheffield, Sheffield, United Kingdom

w.peters@sheffield.ac.uk

²Department of Computing Science, University of Aberdeen, Aberdeen, United Kingdom

azwyner@abdn.ac.uk

Abstract

The paper investigates the extent of the support semi-automatic analysis can provide for the specific task of assigning Hohfeldian relations of Duty, using the General Architecture for Text Engineering tool for the automated extraction of Duty instances and the bearers of associated roles. The outcome of the analysis supports scholars in identifying Hohfeldian structures in legal text when performing close reading of the texts. A cyclic workflow involving automated annotation and expert feedback will incrementally increase the quality and coverage of the automatic extraction process, and increasingly reduce the amount of manual work required of the scholar.

Keywords: legal relations, Hohfeld, text analysis, information extraction

1. Introduction

The automatic acquisition of knowledge from text is a complex, incomplete and incremental process. Knowledge is quite often available but left implicit in text, or depends on previous analysis steps.

Legal experts need tools that help to extract and interpret large amounts of legal texts in a uniform way. Of particular interest, experts wish to access the norms, which express the duties, rights, etc. of the parties discussed in the law. At a more fine-grained level, it is important to access who bears what role with respect to the norm, that is, who is the responsible agent or the receiving party within the action. Yet, it is widely acknowledged that given the complexity of legal language, this is a difficult task. One step towards facilitating this task is to establish a semantic model of the norms, giving the structure of the parties, the roles they play, and the interconnections among the different forms of norms. Such a model provides a target to guide identification and extraction of key textual components, which would move us closer to the goal of making the contents of legal texts accessible in greater detail, variety, and volume. The task we have set ourselves in this paper is to make selected, specific semantic content, e.g. the Hohfeldian duties, from a sample of European directives explicit and accessible for manual expert interpretation, hypothesis testing and further research. We do this by facilitating the inclusion of language and information technology into legal interpretation workflows customised to scholars' research questions. Also, our wider aim is to provide tools for focused close reading, data interpretation and exploration, collaborative annotation and discussion.

The work described in this paper investigates the extent of the support semi-automatic analysis can provide for the specific task of assigning Hohfeldian relations, and works towards a principled methodology for computer-assisted interpretation of legal texts, on the basis of an evaluated text analytic tool for supporting automated or semi-automated identification and extraction of Hohfeldian knowledge from the source text.

The paper is organised as follows. In Section 2., we discuss

the theoretical background to this work in legal theory. In Section 3., we provide a more detailed description of our methodology. The workflow is described in Section 3.1.. The results of the analysis are presented in Section 4.. We end the paper with a discussion of related work in Section 5., and a discussion of future plans in Section 6.

2. Legal Theory - Hohfeldian Relations

We take as our starting point the semantic model for annotating and reasoning over normative provisions (Francesconi, 2015). This semantic model, a *Provision Model*, is presented as an ontology of Hohfeldian relations (Hohfeld, 1913). As noted in (Francesconi, 2015), there are models of legal relations in legal philosophy, e.g. (Raz, 1980), or computer science, e.g. (Hoekstra et al., 2009; Sartor, 2006). However, these are not suitable to the Semantic Web, which requires models that are computationally tractable; (Francesconi, 2015) provides just such a rich and tractable model.

For our purposes, we focus on the so-called Hohfeldian relations, which are regulative rules, *provision types*, that specify which actions can or cannot be performed, the roles of the parties to the actions, and the objects of the actions, which are *attributes* of provision types. The running example is a clause from European Directive 2002/65/EC:

The supplier shall communicate to the consumer all the contractual terms and conditions and the information referred to in Article 3(1) and Article 4 on paper or on another durable medium available and accessible to the consumer in good time before the consumer is bound by any distance contract or offer.

This is understood as a legal provision that specifies a Duty. The conception of Duty is that a party is required to perform an action such that should she not, it is a violation of the law. The various attributes are: the 'hasDutyBearer' attribute is the party that is responsible for meeting the duty; the 'hasDutyObject' attribute is the entity that the required action is performed on; the 'hasDutyAction' attribute is just the action performed; and the 'hasDutyCounterpart' attribute is the other party to the duty with respect to whom the bearer must perform the action. In our example above, the following Hohfeldian elemments can be identified:

hasDutyBearer	=	'Supplier'
hasDutyObject	=	'Contractual terms
		and conditions
hasDutyAction	=	'Communication'
hasDutyCounterpart	=	'Consumer'

For the human analyst, identifying the values of the attributes in the text is a "relatively" straightforward matter, though in practice and across other examples it is more subtle than it would appear.

There is a range of Hohfeldian legal relations which, it is claimed, are in logical relationships, related to the *squares of opposition* of classical propositional logic. Below we illustrate relations between the provisions of Duty, Right, NoRight, and Privilege (others are Power, Liability, Disability, and Immunity):

Right	correlative	Duty
opposite		opposite
NoRight	correlative	Privilege

The 'opposite' relation has a natural interpretation as negation. The 'correlative' is particularly interesting since it indicates a sort of lexical semantic symmetry between the provisional concepts. Such symmetries allow us to infer, given one statement, another statement, for example: *Bill sold a piano to Jill* implies *Jill bought a piano from Bill.*, and vice versa. Though the statements here are logically equivalent, the still seem to convey different information; that is, in the first, *Bill* is the agent of the selling, while *Jill* is the recipient, while in the second *Jill* is the agent of buying, while *Bill* is the source. We see somewhat similar correspondences of roles between correlative provisions (as well as other Hohfeldian relations), representing the provisions as predicates with arguments of attributes:

Duty(hasBearer='Supplier', hasCounterpart='Consumer') correlates with Right(hasBearer='Consumer', hasCounterpart='Supplier')

The significance of the logical relations among the Hohfeldian provisions (including logical relation of opposition) is that from one representation of a Hohfeldian provision, we might infer the others, which are implicit in the legal meaning of the text. By developing a consistent, rich formal, populated ontology of the legal provisions, the model from (Francesconi, 2015) enables to draw inferences and query the knowledge base. The difficulty, however, is in populating the ontology automatically from the source text.

Given the model, the next significant problem is populating it from source text, preferably by automated means. As we discuss in Section 5., there have been limited efforts to apply NLP tools to extract information to frames such as the Hohfeldian attribute-value constructions.

The approach we take is scoped in several ways. We do not yet consider all the Hohfeldian legal provisions and their relations, but work with a subset in order to define and refine the methodology and tools to extract the relevant information. Furthermore, we do not discuss the relationship between Hohfeldian concepts and deontic logic, i.e. the logic of *obligation*, *prohibition*, and *permission*, though there are several obvious overlaps. Instead, our analysis concentrates on the identification of linguistic deontic constructs and their mapping onto Hohfeldian notions (see section 3.1.4. below).

3. Methodology

In order to achieve maximum result from automated analysis, a flexible methodology is required that allows dynamic integration and accumulation of knowledge into the automatic analysis process.

The analysis we pursue for this purpose is positioned on the qualitative side of the analysis spectrum seen as a scale between quantitative and qualitative analysis (Burdick et al., 2012). Whereas quantitative analysis applies techniques to derive generalisations from large amounts of data, qualitative analysis is characterised by work to identify specific information on data of smaller scale. Starting from baseline automatic analysis, our bottom-up approach of incrementally adding/changing/deleting text annotations captures an increasing but non-exhaustive body of acquired knowledge. This knowledge can then be used for progressive information filtering in order to obtain a workable search space for further information extraction.

These methodological elements highlight the importance of scholarly close reading in this process, leaving the scholar to add, change, or delete any automatically acquired knowledge. The tool and its workflow are solely to assist.

From a general perspective, our methodology assists legal researchers and practitioners by means of an incremental automation of legal interpretation. This is done by creating/changing/adding text annotations in any number of iterations, increasingly capturing relevant information for the task of identifying Hohfeldian actors and relations. Where each next iteration requires inclusion of changed or complex knowledge beyond the scope of the present automatic analysis, we introduce a feedback mechanism within the workflow involving both a legal and technical expert and the human interpreter of the legal text.

Our approach has the advantage that the complexity of the analysis and any unwanted annotations involved in the process can be withheld to any extent from the scholar if this hinders the interpretation process.

Scholarly insight should find its way into the next cyclic application by means of an adaptation of the automatic analysis based on the scholarly feedback. When the resulting annotations are deemed acceptable by the expert, these annotations are then serialized into RDF triples according to a data model.

The eventual semantic web oriented goal of this exercise is to link up scholarly activity with the semantic web. At the end of the workflow the semantics of EU directives will be represented by a semantically maximally exhaustive set of RDF triples expressing a complex network of legal vocabulary, facts, statements and relations.

3.1. Workflow

The workflow consists of a number of stages. Its foundation is linguistic analysis, after which Hohfeldian concepts are step-wise discovered in a heuristic fashion.

Figure 1 shows the main stages of the workflow, which can generally be characterised as involving a cyclic improvement of the automatic annotation process by means of an intervention of the scholar in the form of new/changed/deleted text annotations. Also, detailed nontextual feedback on what information is missing or wrong, will be exploited and operationalised by the text engineer for the further improvement of the automatic processing in the next iteration, by adapting the output of the GATE system according to the experts judgements.

3.1.1. GATE

The fundamental building blocks of our approach are text annotations. By creating and combining annotations the required patterns emerge.

The tool we are working with is GATE¹ (General Architecture for Text Engineering) (Cunningham et al., 2002), which is an open-source framework for language engineering applications. It provides an interface for viewing, adding, and creating text annotations, which have been produced by a purpose-built automatic text analysis pipeline. GATE ensures repeatability of application pipelines and reusability of the results of previously run applications.

3.1.2. Ingestion of the directive into GATE and linguistic pre-processing

Using GATE, we applied existing pipelines to the directives texts for tokenisation, part of speech tagging, lemmatisation and term extraction. This provides us with a normalised linguistic framework for further processing.

3.1.3. Term Selection

In order to focus our Hohfeldian analysis we selected important terminology from the directives under consideration. Our reasoning behind this is that in this way we will be able to extract the main Hohfeldian framework for each directive, and discard peripheral Hohfeldian constructs with a minimum of risk.

We considered as important only the terms that are explicitly defined in a directive (see Figure 2), and terms that are used in these definitions.

3.1.4. Identification of deontic structures

Our linguistic analysis identifies various deontic modalities, which are annotated using a linguistic typology of deontic structures containing standard linguistic descriptors for deontic modality², in our case the GOLD ontology³ (Farrar and Langendoen, 2003). This results in the following subtypes:

- PermissiveModality ⁴
- ObligativeModality ⁵

3.1.5. Identification of Hohfeldian concepts

In this stage we map our linguistic deontic structures onto the hohfeldian concept *Duty*. Language analysis through the use of patterns, the Stanford parser (Klein and Manning, 2003) (syntactic dependency) and the lexical resource VerbNet (Kipper et al., 2006) (semantic information for verbs and their arguments) provides input for heuristics for the identification of relevant role bearers within Hohfeldian constructs. These heuristics are all based on annotation types that have been automatically added to the text.

For instance, the syntactic subject of an *ObligativeModality* structure, within which the main verb requires the thematic role Agent (according to VerbNet), is annotated as both the *Agent* and the Hohfeldian *DutyBearer*. In this way we try to identify relations within Hohfeldian *Duty* constructs which involve any defined term or term definition elements.

3.1.6. Expert Evaluation and Feedback

Working with and evaluating the system output involved a close reading exercise by a legal expert using the GATE graphical user interface as illustrated in figure 3. Close reading of the directives and deleting, adding and changing annotations of the texts yielded the annotations needed for the computation of the initial system performance.

3.1.7. RDF Serialization

One the expert judges the annotations to be correct and complete, the data need to be made available in a semantically explicit data structure. In order to make the results available on the Semantic Web, and thus embed the analysis results in a potentially much wider semantic context, the correct annotation structure is mapped onto the existing Hohfeld ontology (Francesconi, 2015). The Hohfeldian annotation instances populate the ontology by means of an RDF serialisation, which makes all Hohfeldian constructs available as they are explicitly stated in the text. Implicit Hohfeldian constructs can then be derived from the ontological structure through the opposition and correlation relations, as discussed in section 2..

4. Evaluation

We tested our system and workflow on two directives: 2002L0065 concerning the distance marketing of consumer financial services and 2007L0023 on the marketing of pyrotechnic articles.

4.1. Duty

The exercise concentrated on the Hohfeldian notion of *Duty*. It was found that many expressions can be deemed duties in the directives. In order to provide a workable scope to our evaluation, we needed a precise definition of *Duty*. For our evaluation we defined as valid *Duty* instances those in which the manufacturer (within 2007L0023), the supplier or the consumer (both within 2002L0065) is required to perform an action, and, should she not, would be

¹http://www.gate.ac.uk

²http://www.isocat.org/datcat/DC-3155

³http://www.isocat.org/rest/dcs/365

⁴http://www.isocat.org/datcat/DC-3384

⁵http://www.isocat.org/datcat/DC-3356



Figure 1: Overview of the Hohfeld Exploration Workflow

(a) "distance contract" means any contract concerning financial services concluded between a supplier and a consumer under an organised distance sales or service-provision scheme run by the supplier, who, for the purpose of that contract, makes exclusive use of one or more means of distance communication up to and including the time at which the contract is concluded;



Figure 2: Term definitions in European directives

in violation of the directive provision. This subset of extracted Duties formed the basis of our evaluation.

This criterion discards a number of duty candidates which, according to less strict criteria, would fall within our target set. For instance, in "The consumer shall return to the supplier any sums and/or property he has received from the supplier without any undue delay and no later than within 30 calendar days", the consumer is clearly bound to an action, and the duty is therefore taken into account. In contrast and a less direct fashion, this is also true of the manufacturer in "The labelling of pyrotechnic articles for vehicles shall include the name of the manufacturer or, where the manufacturer is not established in the Community, the name of the importer, the name and type of the article and the safety instructions." Yet, this example is not included in the evaluation because the manufacturer is only an implied duty bearer and is not syntactically realised with respect to the action expressed by the predicate "include". Moreover, "The Commission shall make available the information communicated by Member States to the European Parliament and the Council and shall ensure that it is also available to consumers and suppliers who request it." does not impose any duty on either consumer or supplier, and this duty instance is not taken into the evaluation.

Applying this criterion to the directives left us with 56 duties (13 in 2002L0065, 43 in 2007L0023) Of these 56, 20 were evaluated.

Figure 3 shows the annotation structure of a Hohfeldian *Duty*. The customisable selection of information to be displayed by ticking the desired annotation types in the right hand pane is a good way to hide confusing information from the scholar in order to facilitate judgment. In this case, only Hohfeldian annotations are shown, whereas linguistic and terminological information obtained from the linguistic analysis is hidden (but readily available if required).

Figure 4 shows the performance of our system for the recognition of the conceptual element *Duty* in these two directive texts.

One immediate observation is that the performance is not the same across directives. This means that the grammar rules created for *Duty* recognition are not sufficient for directive 2002L0065. This in turn leads to the assumption that duties are expressed by more complex linguistic means in this directive.

4.2. Duty Roles

Within the illustrated Hohfeld *Duty* span, four annotation types have been added by our pipeline that are indicative of Hohfeldian roles (see section 2. above): *ActionVerb* ("return"), *DutyBearer* ("consumer"), *DutyCounterPart* ("supplier") and *DutyObject* ("sum"). In addition, "sums" has also been erroneously annotated as *DutyCounterPart*.

According to our workflow (figure 1), it is then the turn of the scholar to closely inspect these results in the GATE interface. This tool allows the addition, deletion or change of annotations. Using this interface, the *DutyCounterPart* was deleted for "sums", and the span of the *DutyObject* was enlarged to cover the whole phrase "sums and/or property he has received from the supplier".

The performance of the system on Hohfeld role detection is displayed in figure 5. Several issues need to be mentioned, since they have an impact on the scores.

First, our present pipeline does not yet cover complex structures such as subcategorised finite and infinite clauses that can be interpreted as *DutyObjects* with their own internal structure. This accounts for 50% of the missing *DutyObject* instances in our evaluation set.

Secondly, *DutyCounterPart* performs less well than the other roles. The main reasons for this are that there were very few explicitly mentioned and the grammar rules considerably overgenerate regarding *DutyCounterPart* role assignment. In further workflow iterations, this should be restricted.

Thirdly, there are many implicit role instances in the texts. In order to detect these we will need to combine several sources of information from the wider context of the senfrom the day on which the supplier receives the notification of withdrawal

5. The consumer shall return to the supplier any sums and/or property he has received from the supplier without any undue delay and no later than within 30 calendar days. This period shall begin from the day on which the consumer dispatches the notification of withdrawal.

Article 8

Payment by card



Figure 3: Example of a Hohfeld structure

Directive	Correct	Spurious	Missing	Precision	Recall	F-measure
2002L0065	4	3	9	57%	31%	40%
2007L0023	39	4	12	91%	76%	73%
Overall	43	7	21	86%	67%	75%

Figure 4: Performance scores for the recognition of Duty

tence that expresses the *Duty*, such as syntactic relations, role slot fillers derived from thematic role information associated with verbal semantics and perhaps pragmatic rules as well.

5. Related Work

In (Biagioli et al., 2005; Francesconi, 2009), excellent results are claimed for the identification of provision types and the semantic roles on 209 provisions using a xmLegesExtractor⁶, which is a suite of NLP tools for automatic analysis of Italian legal texts. The provision types are classified and their attribute-value construction is treated as a frame. The NLP tools apply chunking, dependency relations, and specialised grammar to associate the linguistic structure with the semantic roles in the frames. The approach and results have not been independently reproduced and validated; it cannot apply to English, which would require a substantive revision of the tools; it is unclear the extent to which the materials and tools are openly available for experimentation and development. Our approach shares core similarities, though for English and using with a widely accessible, open source platform.

In (Wyner and Peters, 2011), there is a GATE-based approach to rule annotation and extraction from US regulations, primarily for the components of the rules such as premise, conclusion, modality, action, and exception. It is not specifically tied to a Hohfeldian analysis or ontology. Semantic roles of syntatic constituents were not identified, nor associated with Hohfeldian roles.

At the other end of the analysis spectrum, (de Matt and Winkels, 2010) richly classify legal provisions using a machine learning approach and in Dutch. However, Hohfeldian provisions, roles, or ontological structures are not considered.

Semantic Web language analysis tools such as FRED⁷ (Presutti et al., 2012) and T2R (Hassanzadeh et al., 2013) generate RDF/OWL ontologies and/or linked data triples from natural language sentences. They are based on existing tools such as the Stanford parser, C&C/Boxer or the baseline analysis of the Penn Treebank. Our initial analysis is also based on Stanford, using dependencies. In previous work, it was found that the Stanford parser is highly limited in its treatment of legal texts (Wyner and Peters, 2011; Wyner et al., 2012). In our view, these parsers are not yet sufficient for detailed analysis of Hohfeldian structures. Given the complexities of the legal language used in directives we foresee the necessity of further refinement and coverage extension in our workflow.

6. Conclusion and future work

We have presented a flexible methodology that enables an incremental semi-automated analysis of Hohfeldian structures in European directives. Expert feedback is taken into account for each iteration of automated analysis, creating grammar rules to accommodate the expert's requirements. This cyclic workflow, as illustrated in figure 1, will incrementally increase the quality and coverage of the automatic extraction process, and increasingly reduce the amount of manual work required of the scholar.

The results reported in this paper are also to be interpreted as incremental. The sample size of two documents is small, but indicative of the feasibility of our approach. We will increase the number of directives in our analysis, in order to obtain a distributionally more representative overview of the linguistic and conceptual issues involved in the analysis. As a next step, we will concentrate on the most significant extension of our linguistic coverage, and start with the inclusion of the surface syntactic representations of Duty elements by means of subcategorised finite and infinite clauses as noted in section 4.2.

In our view, the analysis of particular Hohfeldian elements, e.g. Duties and Rights, can be extended to other Hohfeldian entities, which have very similar constructions.

The evolving combination of linguistic analysis and ontology-based inference will increasingly blur the distinction between explicit and implicit knowledge. It is clear that to capture a range of implicit roles, further knowledge will have to be integrated on top of ontological inference, e.g. relations between entities from external knowledge sources, and the pragmatics of the text.

⁶http://www.xmleges.org/eng/

⁷http://wit.istc.cnr.it/stlab-tools/fred

Role	Correct	Spurious	Missing	Precision	Recall	F-measure
DutyBearer	8	23	4	35%	67%	46%
DutyCounterPart	3	14	0	21%	100%	35%
DutyObject	5	8	9	63%	36%	46%
ActionVerb	19	1	0	95%	100%	97%

Figure 5: Performance scores for the recognition of Duty Roles

Finally, once a sufficiently populated ontology, based on textual sources, is developed, a key phase of evaluation will be the extent to which legal professionals concur with the inferences from the ontology, for in such a way, the Hohfeldian analysis approach will be put to an empirical test.

7. Bibliographical References

- Biagioli, C., Francesconi, E., Passerini, A., Montemagni, S., and Soria, C. (2005). Automatic semantics extraction in law documents. In *Proceedings of the 10th International Conference on Artificial Intelligence and Law*, ICAIL '05, pages 133–140, New York, NY, USA. ACM.
- Burdick, A., Drucker, J., Lunenfeld, P., Pressner, T., and Schnapp, J. (2012). Digital humanities. In *Jan Baetens Cambridge, Mass.: MIT Press.*
- Cunningham, H., Maynard, D., Bontcheva, K., and Tablan, V. (2002). Gate: A framework and graphical development environment for robust nlp tools and applications. In *In Proceedings ACL02*, pages 168–175.
- de Matt, E. and Winkels, R. (2010). Automated classification of norms in sources of law. In Proceedings of the workshop on Semantic Processing of Legal Texts, volume 6036 of Lecture Notes in Computer Science, pages 170–191. Springer.
- Farrar, S. and Langendoen, T. (2003). A linguistic ontology for the semantic web. In *GLOT International 7 (3)*, pages 97–100.
- Francesconi, E. (2009). An approach to legal rules modelling and automatic learning. In *JURIX*, pages 59–68.
- Francesconi, E. (2015). Semantic model for legal resources: Annotation and reasoning over normative provisions. Semantic Web Journal, Special Issue on Semantic Web for the legal domain, pages 1–11. To appear.
- Hassanzadeh, K., Reformat, M., Pedrycz, W., Jamal, I., and Berezowski, J. (2013). T2r: System for converting textual documents into rdf triples. In Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) - Volume 03, WI-IAT '13, pages 221–228, Washington, DC, USA. IEEE Computer Society.
- Hoekstra, R., Breuker, J., Di Bello, M., and Boer, A. (2009). Lkif core: Principled ontology development for the legal domain. In *Law, ontologies and the semantic* web, pages 21–52, Amsterdam.
- Hohfeld, W. (1913). Fundamental legal conceptions as applied in judicial reasoning. In *Yale Law Journal 23(1)*.
- Kipper, K., Korhonen, A., Ryant, N., and Palmer, M. (2006). Extensive classifications of english verbs. In *Proceedings of the 12th EURALEX International Congress. Turin, Italy.*

- Klein, D. and Manning, C. (2003). Accurate unlexicalized parsing. In Proceedings of the 41st Meeting of the Association for Computational Linguistics, pages 423–430.
- Presutti, V., Draicchio, F., and Gangemi, A. (2012). Knowledge extraction based on discourse representation theory and linguistic frames. In *Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management*, EKAW'12, pages 114–129, Berlin, Heidelberg. Springer-Verlag.
- Raz, J. (1980). The concept of a legal system: an introuction to the theory of legal systems. Oxford University Press.
- Sartor, G. (2006). Fundamental legal concepts: a formal and teleological characterisation. *Artificial Intelligence and Law*, 14(1):101–142.
- Wyner, A. and Peters, W. (2011). On rule extraction from regulations. In Katie Atkinson, editor, *Legal Knowledge* and Information Systems - JURIX 2011: The Twenty-Fourth Annual Conference, pages 113–122. IOS Press.
- Wyner, A., Bos, J., Basile, V., and Quaresma, P. (2012). An empirical approach to the semantic representation of law. In *Proceedings of 25th International Conference* on Legal Knowledge and Information Systems (JURIX 2012), pages 177–180, Amsterdam, The Netherlands. IOS Press.