Two Limitations in Legal Knowledge Base Constructing and Formalizing Law

Réka Markovich*
Computer Science and Communications Research Unit, University of Luxembourg
Department of Business Law, Budapest University of Technology and Economics
Department of Logic, Institute of Philosophy, Eötvös Loránd University
reka.markovich@uni.lu

Abstract

Extracting norms from legislative texts confronts us many tasks and requires decisions about approaches, methods, tools, and legal theoretical presuppositions. In this paper I present some examples from the Hungarian legislation showing how challenging sometimes the wording of these texts is from the viewpoint of norm extracting, then I present two limitations we need to face when dealing with this task. On the one hand, I argue that isomorphism cannot be upheld, but it is not necessary for having a faithful formalization. On the other, I argue that however appealing to base on constitutive norms is when formalizing—in order to avoid the necessity of normative reasoning, for instance—the consequential application of their theory in the approach leads us to a very controversial situation.

1 Introduction

In this paper some considerations of information extracting and formalizing law will be presented. The focus of this investigation is on some of the phenomena of law

I am grateful for the comments on my preliminary thoughts expanded in this paper to Trevor Bench-Capon, Marek Sergot, István Szakadát, Gábor Hamp and Csaba Oravecz.

*Support provided by the research project K-116191 of the Hungarian Scientific Research Fund is gratefully acknowledged. The research reported in this paper was supported by the Higher Education Excellence Program of the Ministry of Human Capacities in the frame of Artificial Intelligence research area of Budapest University of Technology and Economics (BME FIKP-MI/FM) and the Higher Education Institutional Excellence Grant Autonomous Vehicles, Automation, Normativity: Logical and Ethical Issues at ELTE BTK.

and legislative texts we face and need to deal with if we want to extract norms from the legislative texts and how these influence the methods we need to apply when constructing our knowledge base system and formal setup.[1] The examples used here come from Hungarian law,[2] but we have no reason to suppose that the phenomena their structure and linguistic surface confront us would be particular.

I have no intention to argue for the legal positivist approach in the sense of advertising the absolute primacy of written law, but I restrict this examination to the legislative texts and the linguistic-legal information that can be extracted from them—by human or machine tools (semi-automatized or automatized way). There are several reasons for doing so. One is the Hungarian—and, in general, the continental—legal system's tacit relying on this approach in the sense of requiring clear legislative texts free from ambiguities—as this is what people can access and according to which they are required to behave, and the law has to be understandable by those who are subjects of it. The decree regulating the legislative drafting in the Hungarian legal system (61/2009. (XII. 14.) decree of the Minister of Justice) says: "Legal regulation drafts should be drawn according to the rules of Hungarian language, in a clear, comprehensive, consistent way" (Section 2). The reason why the examination in this paper concerns not only the human understanding but the formal representation of laws too in order to prepare "machine understanding" is the amount of laws: at the time of writing these lines there are 7156 statues in force, the (parallel) processing of which is way beyond human capacity. We need to consider how (and whether) the (semi-)automated formalization of legislative texts can be managed: how (and whether) the (normative) content of a legislative text is clearly detectable and identifiable.

Obviously, such a topic involves considerations not only from legal theory, but also from computer science, natural language processing (hereafter, NLP) and logic. Nevertheless, in this paper I do not suggest or even consider tools or methodologies for NLP solutions for moving toward a (semi-)automated interpretation of legal texts, as it happened—among many others—in Camilleri et al. [11], Boella et al. [9], de Maat et al. [14], de Maat and Winkels [13] (for project results see Legivoc [18], Openlaws [42], EUCases [8], MIREL, and the most recent DAPRECO [3] aiming at translating

[1] This paper is not self-contained in the sense of referring many existing and often discussed issues without introducing or explaining them in detail. Therefore, the intended audience is the audience of this special issue having background in and experience in mining and reasoning with legal texts.

[2] The examples will be presented both in Hungarian and in—as faithfully to the original Hungarian texts' wording as possible—English translation.

the GDPR into a deontic logic specifically designed for natural language semantics in Robaldo and Sun [31]—also related to ProLeMAS), neither will consider specific formalisms—listing of which works would far exceed the limits and scope of this paper;[3] and from the knowledge base constructing issues I only will address those that directly relate to the topics discussed here.

The methodology of this paper is presenting real legislative examples to show that why and how seemingly basic steps of formalizing law and constracting legal knowledge base can be difficult, especially if we consider automated analysis—which, as I argued above, can easily made desirable by the amount of norms of a given legal system. The investigation of the examples will show why the popular and intuitive requirement of isomorphism cannot be upheld, and I will present arguments why it should not even be necessarily pursued. The examples and arguments above will also continuously reflect upon the semantics of legal norms and the differentiation between constitutive and regulative ones, trying to find solutions to the difficulties presented by the examples. This interaction between the examples and the arguments will lead us to another limitation: the limitation of extensively formalizing law in terms of constitutive norms.

2 Approach and Presuppositions

Let's suppose: we are looking for norms in legislative texts. In agreement with the approach of van Kralingen et al. [41] as "a norm must convey information to fulfil its function of communicating standards of behaviour; the way in which one is expected to behave must be clear from the norm", we are looking for the way this can be extracted from the legislative texts (presupposing that they contain this information as we think about them as sets of norms). The paper of van Kralingen et al. [41] refers to Brouwer [10] when listing five questions to be answered if we want to have a complete norm:
(1) Who is obliged or permitted to do something?
(2) Is there an obligation or a permission to do something or to leave something undone?
(3) What must be done or forborne?
(4) Where must something be done or forborne?
(5) When must something be done or forborne?
These elements more or less correspond to the factors most authors count with when

[3] By no means this implies the claim that one or another specific formalism could not be better in dealing with the issues presented in this paper, but here the focus is on preliminary considerations.

discussing norms. In von Wright [43], the "father of deontic logic" adds the authority to the elements above when talks about the parts or components of norms that are prescriptions—as von Wright differentiates between three type of norms: rules (rules of a game), directives (like technical norms of an instruction manual) and prescriptions (commands, permissions, and prohibitions, which are given or issued to agents concerning their conduct). "The laws of states are prescriptions", declares von Wright.

Let's restrict our investigation to the character or deontic status (whether something is obligatory, permitted or prohibited), the action (what is obligatory, permitted or prohibited) and the agent (the subject or addressee of the norm: for whom it is obligatory, permitted or prohibited the given action). The importance of agents is often tacitly underestimated: it is pretty general to discuss deontic logic in an impersonal way, that is, discuss obligations without agents, but in law there is always an addressee of the given norm. And if we take the expectation in van Kralingen et al. [41] above seriously accepting that a norm must convey how *one* is expected to behave, extracting who that 'one' is a fundamental requirement of the whole process. Let's start with the character or deontic status, though, as it is reasonable to suppose that this component is the most salient of the norms—and as we will see, the question of the actions and agents come with it anyway.

3 Language of Legislative Texts: Deontic Character and Linguistic Modalities in Law

The language use in legislation is more formal and bound than other registers of natural language (whether necessarily or not is often discussed, see the Plain English Movement, Benson [7] vs. for instance Crump [12]), even more than other registers in legal language (for instance compared to the language of judgements, contracts or explanations).[4] As we saw above, the legislator itself has formed requirements concerning the legislative language use in a decree on legislative drafting. There are a lot of features of the legal language use in Hungarian (for instance the high proportion of nominal structures) discussed often by linguists and lawyers (this is, of course, is not a specifically Hungarian issue, for English see e.g. Mellinkoff [28], Tiersma and Solan [37], for a comparative study to English, French, German, and Latin see Mattila [27]), but now we only consider those that are related to the expression of linguistic modality as these are by which we can identify the deontic character of a

[4]Most of the comparative studies of a recent Hungarian project on this topic are only available in Hungarian, see: Szabó [35], but one of them will be available in English soon: Ződi [44].

norm: what is obligatory, what is forbidden and what is permitted.
In the Hungarian language, modalities can be expressed in different ways (not only in Hungarian, of course, for a foundational entry see Kratzer [23]): the most typical ways are participles, auxiliary verbs with infinitives, and suffixes at the end of the verbs. For instance, *obligatory* can be expressed with saying that something is 'kötelező' (which is the literal translation of 'obligatory'), or 'szükséges' (which is the literal translation of 'necessary')—in its deontic reading, with using the auxiliary verb 'kell' (which means what 'must' or 'shall' means in deontic sense), or with the derivational suffix '-andó, -endő' at the end of the verb denoting the action (for instance, the translation of the word 'fizetendő' is 'is to be paid'). *Permission* can be expressed with the adverb/adjective 'megengedett' (which is the literal translation of 'permitted') or using the adjective 'szabad' (literally meaning 'free to'), but most frequently happens with the inflectional suffix '-hat, -het' attached to the verb which describes the action, or with the derivational suffix '-ható, -hető' attached to the verb (in the predicate place of the sentence) with a—tacit—copula: in this case the translation of the word e.g. 'fizethető' 'it may be paid' or 'it can be paid' with the deontic reading of 'can'. That is, in the Hungarian legal language, these suffixes play the role that in English is played by modal verbs 'may' and 'can' in their deontic reading. These suffixes also have other readings (just like the word 'szükséges', that is, 'necessary', obviously has), but in a legislative text the deontic context is given—or, as Kratzer [23] calls it, the conversational background is bound: it is *literally* is 'what the law provides' as we are reading *the law itself*. *Forbidden* can be expressed with the adjective 'tilos' (which is the closest version to the translation of 'forbidden'), with the negation of *permitted*, that is, any of the indicated possibilities above with a declarative use of 'no' ('nem' in Hungarian).

3.1 False Friends and Missing Modalities in Legislative Texts

Sometimes, though, other ways of expressing modalities can be faced with: ones that might mislead the reader, both human and computer ones.

In Markovich [25] the language of the Hungarian Criminal Code (and its legal theoretical consequence) is discussed: as a criminal code lists the actions a society (a legislator) considers the less desired, one would expect to meet the linguistic signs of *forbidden* several times—but would disappoint as the only deontic modality that can be detected in this legislative text is *obligatory*. The dominant form of sections in the Hungarian Criminal Code's special part (listing the felonies) is the following:

Example (1) "Aki mást megöl, bűntett miatt öt évtől tizenöt évig terjedő szabad-

ságvesztéssel büntetendő." (Any person who kills another human is to be punished for five to fifteen years of imprisonment due to having committed a felony.) [Act C of 2012 on the Criminal Code]

As we see, the command to the judge to punish the perpetrator of a forbidden act is not *attached* to the prohibition: it *is* itself the expression of the act being forbidden. Of course, the human interpretator understands that this realizes what we usually refer to as 'forbidden', but the automated processing needs some input in order to properly classify these norms—which, at least in Hungarian—could be the string 'büntetendő' (is to be punished) as these norms are very regular: whenever the computer finds this word, the declarative form verb of the given sentence will provide the action to which we need to assign the deontic status of being forbidden.

A more surprising set of examples can be found in the Hungarian Highway Code. There are several sentences in it like the following ones:

Example(s) (2)
"A fényjelző készülék folyamatos zöld fény jelzésnél kiegészítő hangjelzést is adhat" (The light-signalling device may additionally beep while green.)
"A fény kibocsátására alkalmas jelzőtáblán a jelzőtábla fehér és fekete jelzései egymással felcserélt színekkel is megjelenhetnek." (The white and black markings on the light signal board may appear with commuted colors.)
"A jelzőtábla alatt kiegészítő tábla adhat útmutatást a jelzőtábla jelzésének értelmezésére." Under the traffic sign, an additional sign may give guidance how to interpret the traffic sign.
"Az út mellett vagy közelében lévő egyes létesítményekről kék vagy barna alapszín? jelzőtáblák adhatnak tájékoztatást." (Information about the facilities passed by the road may be given by blue or brown signposts.)
"Az (1) bekezdésben említett jelzőtáblák alatt elhelyezett kiegészítő táblán nyíl jelezheti, hogy a tilalom hatálya a táblától kezdődően vagy a tábláig áll fenn." (Additional signpost with an arrow under the traffic signs mentioned in (1) may indicate whether the prohibition starts or ends with the traffic sign.)

These 'may'-s (the '-hat', '-het' suffixes in Hungarian) might seem permission at first sight. We, of course, know that 'may', just like the given Hungarian suffixes, can be used with other modal tastes—e.g. alethic or metaphysical, but before facing these examples we did not expect to meet with them in legislative texts since, as it was mentioned above, here the context is bound, or, as Kratzer [23] says, the conversational background is given since the resolving deontic reading can paraphrased

as 'in view of what law provides'—which is tacitly the case as we are reading the law itself. But the modalities in examples (2) do not really make sense as permissions: the legislator does not give a permission to the light-signalling device to beep. These rather seem alethic possibilities. Does this mean that we need to introduce a new modality into our formalization? I don't think so. *If* we consider descriptive propositions, these possibilities can be expressed with a disjunction: the light-signalling device beeps or it does not. But what is the deontic character of this rule then? Well, it is an obligation to people using public roads to consider a light-signalling device as such—both if it beeps and if it does not. So we found a conjunction of obligations where the first impression—and the trained machine's result would have—suggested permission. Not an easy tension to resolve with automatic tools.

It also happens that there is no linguistic sign of any modality in a sentence. In Kratzer [23] we find an example of a sentence missing any linguistic sign of a modality: the explanation Kratzer gives to "The car goes twenty miles an hour" is that "the modality may be inherent in the verb" (p. 639). She does not mention cases, though, where the deontic modality is present inherently in a verb. In legislative texts, actually, we often find seemingly declarative sentences lacking any linguistic sign of modality. The reader probably says promptly: of course, those are constitutive rules. Indeed, in case of constitutive rules, the lack of 'shall' or other phenomenon expressing normativity is not surprising, this is part of the description we got in Searle [33] describing constitutive rules as mostly expressed in non-imperative, 'counts-as' rules. We also can take a step back in history and check what an earlier author whom Searle also leans on: Rawls [30] describes practice rules as definitive ones, which definitive feature is so strong that if we don't follow these rules, we don't engage in the given action they define.

If we want to explain the using of indicative mood with this distinction, we need to understood broadly the word 'definition', at least this is how it seems from the Hungarian legislative language use. Let's see the following examples:

Example (3) "Szünetel a biztosítás a fizetés nélküli szabadság ideje alatt." (The insurance intermits during the unpaid leave.) [Act LXXX of 1997 on Social Security]

We might say that the definition of insurance as an institution can be paraphrased in a way it contains this aspects of intermitting. Maybe, we can also say something similar in case of the following rule:

Example (4) "Ha leszármazó és szülő nincs vagy nem örökölhet, az örökhagyó háza-

társa egyedül örököl." (If there is no descendant or parent, or if they are excluded from succession, the surviving spouse inherits the entire estate.) [Act V of 2013 on the Civil Code]

What does this mean in terms of the norm elements we are looking for? Just to start with the first element we look for: what is the deontic character of these norms? Is there one at all? Constitutive norms are usually something which help us avoid dealing with deontic notions—and therefore deontic logic or normative reasoning. But should we really lean on this option when constructing our knowledge base and formalizing law? Actually, Searle never said that there is no obligation in counts-as norms (what is more, speaks about *deontic power* in their case)—(even if this is typically not the case) they *might* be expressed imperative. (And there are several critiques of the theories of the whole differentiation between constitutive and regulative norms, see for instance Giddens [16], Lagerspetz [24], Garcia [15], Tuomela [39], Tummolini and Castelfranchi [38]). For instance rules listing the required elements of an official document to count as an official document is a paradigmatic case what we usually consider a constitutive rule. Still, the phrasing of the following sentence suggests an obligation:

Example (5) "A meghatalmazás képviseleti jogot létesítő egyoldalú jognyilatkozat. A meghatalmazást a képviselőhöz, az érdekelt hatósághoz, bírósághoz vagy ahhoz a személyhez kell intézni, akihez a meghatalmazás alapján a képviselő jognyilatkozatot jogosult tenni." (A power of attorney is a unilateral act granting the right of representation. The power of attorney shall be addressed to the agent, the competent authority or court, or any person to whom the agent is authorized to make a legal statement.)[Act V of 2013 on the Civil Code]

In von Wright [43] we also find a clear claim about the laws of the state being *prescriptions*, per se, but let's see what the strictly legal approach says.

Legislation as such is subject of a discipline called *legistics* or *legistica* taught at law schools. Hungarian legistics textbooks—obviously—often discuss Hungarian legislative language use, both in descriptive and prescriptive way (discussing how the language should be used when phrasing legal norms). Considering our topic, we can find the following in Tóth [40]: "In legal norms indicative mood is dominant. Thus the norm text shuns expressions like 'should', 'ought', 'necessary', since the indicative mood is imperative for the authorities, government agencies (e.g. rules of competence, authorizing rules, rules of procedure); while, less often, to put emphasis, norms expressing obligation literally too may occur—e.g. when the obliged agents

are directly citizens, business organizations."[5] Unfortunately, the parsing tools for the Hungarian language does not provide the possibility to check the correlation between the addressees (personal scope) and the linguistic mood of expressing obligations, the claim is still worth to deeper analysis. On the one hand it corresponds nicely with the foundations of the regulative–constitutive rules distinction: as it is often referred, the activities defined by constitutive rules are logically dependent on the rules constituting them, so it is not strange if agents created by the law conducting activities created by the law get their commands from rules having the same linguistic features as the ones created them and their activities. On the other hand, this claim of Tóth [40] says this indicative mood is imperative to them, that is, sounds constitutive or not, these rules bear normativity, they convey imperatives. This is confirmed by another textbook of legistics: Tamás [36] says that "in Hungarian legislative texts the indicative mood means imperatives or obligation. (...) The predicate in norms' text is always a command, even if it has no linguistic sign."[6] This means that we need to look for and identify deontic character in all the laws' sentences. The correlation indicated above might help is, but there are cases, where this correlation between the official addressee and the phrasing does not hold, like in the following example:

Example (6) "...a vállalkozás egyértelműen és jól látható módon felhívja a fogyasztó figyelmét a 11.§-ban meghatározott információkra" (...the company calls the consumer's attention to the information detailed in section 11 in an univoque and visible way) [45/2014 (II. 26.) Government decree on Detailed Rules of Contracts between Consumer and Company]

Here the obliged agent is a company, that is, not an authority, the mood is still indicative representing no modality. We might be appealed to conclude that all indicative mood predicates cover obligations in legislative texts. For these cases the NLP considerations can be made up with saying that in legislative texts all the verbs in indicative mood—except for the ones in the antecedents of conditional rules[7]—should be detected as obligatory. That is, in the case of example (6) finding 'felhívja'

[5]"A jogi normákban egyeduralkodó a kijelentő mód. Ezért a norma szövege kerüli a "kell", "köteles", "szükséges" kifejezést, mert a hatóságok, állami szervek számára a kijelentő mód imperatív (pl. hatásköri szabályok, felhatalmazó szabályok, eljárási szabályok), míg nyomatékosításként ritkábban előfordulhat a kötelezést nyelvileg is kifejező norma, (...) például akkor, ha a kötelezettek közvetlenül magányszemélyek, gazdálkodó szervezetek."

[6]"Magyar nyelvű normaszövegekben a kijelentő mód felszólítást vagy kötelezést jelent. (...) A normatív állítmány mindig rendelkezés, akkor is, ha nyelvileg nem az."

[7]About the automated identifiability of the antecedent of conditional sentences in legislative texts in Hungarian see Markovich et al. [26]

('calls') should be translated to a formal representation of 'it is obligatory to call'. But this solution does not help in examples above: in example (3) the verb is 'intermits' and the deontic content as 'it is obligatory to intermit' sounds strange, the 'it is obligatory to inherit' even more does so. The reason is that inheriting and intermitting are not actions to which oblige someone—or something—would make sense: the spouse has no influence on whether (s)he inherits, the insurance has even less whether it intermits. The principle along which this problem might be solved in NLP and ((semi-)automated) formalization is that instead of acts, we put propositions in the argument of the deontic operators at this level: 'it is obligatory that the insurance intermit' and 'it is obligatory that the spouse inherit' makes much more sense. The satisfaction is temporary, though, as this solution does not provide some crucial information of compliance: whose obligation is that to make things so? In example (6) we see the agent, but in example (3) and (4) the real agent is not in the given sentence or section, and the extraction of this information sometimes is not easy to the human reader either: it requires some legal knowledge where to look for the answer. That is, constructing the knowledge base requires that legal knowledge. Before going further with this direction, note that we already lose something: isomorphism.

4 Isomorphism and its limitation

At the end of the eighties, the term 'isomorphism' has been introduced into the discussion of legal knowledge based systems and formalizing law. In Karpf [22] we find five conditions that have been listed in Bench-Capon and Coenen [5] as the following:
(i) Each legal source is presented separately.
(ii) The representation preserves the structure of each legal source.
(iii) The representation preserves the traditional mutual relation, references and connections between the legal sources.
(iv) The representation of the legal sources and their mutual relations (...) is separate from all other parts of the model, notably representation of queries and facts management.
(v) If procedural law is part of the domain of the model then the law module will have representation of material as well as procedural rules and it is demanded that the whole system functions in accordance with and in the order following the procedural rules.

Practically, (ii) is what matters the most, the short or narrow references at least

refer to isomorphism mainly covering that: Bench-Capon and Coenen [5] sum up presenting the rules above with saying that "the important demand made by isomorphism is that there is a clear correspondence between items to be found in the source material and items to be found in the knowledge base. (...) Where one to one correspondence is not achieved, however, it is important to relax the constraint only so that one source item corresponds to several knowledge base items and maintain the prohibition on a single knowledge base item capturing the material from several source items." The definition we see in Prakken and Schrickx [29] is quite similar: "the situation that one source unit is formalised in one knowledge base unit. By source unit we mean the smallest identifiable unit of the source from which a norm can be extracted. In general this will be a section or a subsection of a code."

Bench-Capon and Coenen [5] convincingly argue that following these rules greatly help to satisfy the concerns of well engineered knowledge systems presented in Jackson [19]: verification, validation, and maintenance. They argue that isomorphic formalization also helps the user: "Many of the problems with experts systems come from a mismatch between the rule based conceptualisation of the expert system and the conceptualisation of the user." (Here the authors do not discuss the potential users of expert systems but Bench-Capon does so in [4].) We come back later to the question of users and mismatches between concept structures.

But not only software engineering considerations serve arguments for isomorphism. In Gordon [17] and Bench-Capon and Gordon [6] we find legal theoretical considerations on legistics about reasons behind the typical structure of laws providing the general rule first and then a number of specific exceptions. According to the authors, this structure reflects on a need coming from the addressees' side. As Gordon [17] says: "This [structure] facilitates the normative function of the law; the law would have little effect on social behaviour if its rules were so convoluted that persons could only with great difficulty, if at all, predict the legal consequences of their actions." According to Bench-Capon and Gordon [6], this presentation structure of general rules and exceptions helps achieve "ease of application, ease of understanding, and the possibility of allocating the burden of proof"—points that can be upheld with keeping on isomorphism when formalizing. In Prakken and Schrickx [29] a few years before, the structure above is yet presented as a threat to, or a limitation of isomorphism, which, however, can be overcome (building up formalization upon nonmonotonic reasoning or conflict solving metarules), as it is presented in the paper.

There are other limitations brought up in the literature, though. Bench-Capon and Coenen [5] refer to Sergot et al. [34] as latently raising objections against isomophism.

It is the critique of isomorphism which is latent in Sergot et al.[34] as it is not even mentioned literally since the paper is about following a top down development of logic programming when formalizing the Indian pension rules, but the criticism of the isomorphic approach can be read clearly from it. As Bench-Capon and Coenen [5] rephrase it: "the isomorphism approach is very well if the legislation is itself well stuctured. In such a case, the structure of the problem, the structure of the legislation, and the structure of an isomorphic knowledge base would all be in harmony. It is, however, often the case that the legislation is not well structured. Often repeated amendments and 'patching' mean that the legislation is itself a complete mess, and fails to reflect the real structure of the domain. In such a case, basing the structure of the knowledge base on legislation would lead to a poorly structured knowledge base, which fails to correspond to the 'real world' problem." With Sergot et al. [34] own words: "in common with many other examples of legislation and regulations, especially those that refer to periods of time, the Pension Rules are imprecise and very casual about many of the key concepts. They are certainly not precise enough to be formulated directly as an executable program, and they are arguably not precise enough to be applied by a human agent either." At their reaction, Bench-Capon and Coenen [5] put a light on the question what is meant by 'poorly structured'—a point to which we come back soon—, and argue by turning the objection into a good argument *for* isomorphism: a poorly structured legislative text very likely becomes subject of amendment, which requires maintenance of our knowledge base—which something that can be efficiently done if we previously followed the isomorphic approach.

At this point, accepting Bench-Capon and Coenen [5]'s reply (and original arguments above), one might be convinced that isomorphism is an obviously advantageous approach when formalizing legal rules, one definitely to pursue, and this whole problem urging us to give it up only came up because we wanted to extract real (regulative) norms from constitutive ones, so we should not do so; examples (3) and (4) refer to actions of official agents/government organizations, so if we would like to help people to comply with law, these can be kept as they are: in descriptive form about what happens in law. In total agreement with the first part of this conclusion, getting rid of the intention to change constitutive into regulative norms won't solve the problem: we encounter it in case of clearly regulative rules, too. Let's have a look at an example looking for a norm in clearly regulative rules.

Rules on advertising (Hungarian Act XLVIII of 2008 on advertising and parts of the Act LVII of 1997 on Fair Competition—which parts earlier were parts of the Act XLVIII of 2008) are full of rules like the following ones:

Example(s) (7)
"Advertising is prohibited for goods whose production or marketing is forbidden."
"Advertisements inciting violent, or personal or public security threatening behavior, are prohibited."
"Comparative advertising may not injure the reputation of another company or the name, merchandise, brand name and other marking of such company."

From the first two, we cannot identify any subject (addressee), from the third, linguistically, we can: the advertising itself. But the advertising or advertisement is clearly not an agent in the sense of being capable of conducting actions, therefore one whose behaviour could be ruled or influenced by imposing a prohibition. Therefore, the advertising (or advertisement) itself cannot be the addressee. But then who is it? Who should see to it that a comparative advertisement does not injure the reputation of other companies? There are several general laws in a society—a phenomenon because of which the harmfulness of the tendency of forgetting about the addressee is not so conspicuous—, the criminal code or the civil code are ones concerning all members of a given society. Is this Act on advertising is one of those? Obviously not, as most people of a given society have absolutely no influence on what happens in advertising. Fortunately, the—overwhelming majority of—legislative texts provide their personal scope in one of the first sections, that is, they denote the agents who shall comply with them. In case of the Act on advertising this section declaring the personal scope does not provide the final solution yet since there are three agents denoted: the advertiser (the company whose product/service the advertisement is about), the advertising service providers (practically the agencies), and the publishers of advertising (the TV and radio companies, the publishers, etc.). From this we could conclude that there are three agents on whom the duties above are imposed. This is not far from the truth, but the situation is a bit more complicated: after checking the beginning of the Act in order to find the agents and put them into the norms we would like to extract, we need to check some sections at the end of the Act, too. After providing the rules in the above form, the legislator put a subtitle (which is one of the possible units of Hungarian legislative texts), the 'Vested responsibility'. This part tells us which addressee will be liable for the violation of a given prohibition above (listed protractedly in the Act)—which practically means that the given agent is the one which is obliged to obey the rule (sometimes the three addressee have joint and several liability—which is a specific problem in terms of formalization, one which we do not discuss here, though—, some rules need to be complied with by one or another of the agents (aligning the reasonable aspect of which phase can be influenced of creating and publishing an advertisement)). If we

want to have proper complete norms containing the addressee too (which would be nice considering the information need to be extracted for compliance), we need to get that information from another parts of the Act, not the one where the "main rule" can be found.

We, of course, could say that, in order to keep isomorphism, we formalize them separately when the prohibitions above (let's call (any of) them Rule Nr. N):
Rule Nr. N: '*it is obligatory that action A is executed / that C is the case*'
and then the "rules" in the section on vested responsibility in a way that:
'*it is obligatory to subject S to comply with the rule Nr. N*', that is,
'*it is obligatory to subject S to fulfil the obligation in rule Nr. N*'
but this result is rather redundant.

But then what could we do with the laws similar to the Act on advertising (of which there are a lot) in "ruling" the agents in separate norms? Annotating the knowledge base slots with the locution in the source (legislative) text could serve as a solution. Do we lose isomorphism? It depends on how strictly we interpret its definition. As it is mentioned above, in Prakken and Schrickx [29] we see one saying that we need to consider one source unit (whose correspondence with one knowledge base unit should be upheld) as the smallest identifiable unit of the source from which a norm can be extracted. Well, in this case it is the whole Act itself—a "structural unit" that authors would scarcely accept as such: representing the whole Act in one unit of the knowledge base would completely contradict to what we consider as one unit. We, therefore, seem to be in the need of loosen up the requirement. A requirement of an *algorithmizable correspondence* would still be considerable as substantial help in pursuing the software engineering concerns of verification, validation, and maintenance.

One could say that the Act on advertising is a good example of the poorly structured legislative texts Sergot et al. [34] talk about. It is important to emphasize: this is not the case. Of course, it depends on, as Bench-Capon and Coenen [5] note, what we call 'poor structure', but in the case of the original example of being poorly structured, the Indian pension rules, the authors found that the same word was used to refer to slightly different things, sometimes even in the same sentence, while in other cases the key concepts were really casually phrased. In the Act on advertising there is no such problem. There is nothing 'poor' from a legal point of view or, specifically, from the viewpoint of legistics in this Act's structure. It is, actually, a well structured law. What is behind the structure of this Act then? From legal theoretical point of view we can call the attention to the fact that advertising law is an area of law aiming

at the protection of consumers and fair competition. The legislator's view reflects on this aim and the viewpoint of the consumer/other companies, thus, instead of concentrating on the obligations of each concerned agent, the legislator ends up in listing the situations that should not be realized (as for the consumer it does not make a difference which agent is responsible for a given undesirable situation).

We need to distinguish two phenomena Sergot et al. [34] talk about, at least we need to reflect more precisely their relationship. The—clearly problematic—practice of handling concepts and terminology poorly is a sufficient condition of not having matching with the real world problem structure, but far from being necessary. The legal (more precisely, legislative) mapping of the world *is not isomorphic* with the world itself. In Ződi [44] we find an—experiment-supported—analysis of the comprehensibility of legal texts. The author's conclusion is that changing the often scolded legal language would not solve the problem that people find it difficult to understand the legal (not just legislative) texts. Offering a classification of three kinds of pragmatic situations where laymen meet legal texts, Ződi points out that the difficulties in understanding legal texts should be investigated together with these situations and not just the syntactical and semantical features of legal texts. He emphasizes: "the texts of legal sources are not organized around everyday problems—they follow a different logic. (...) The texts of legal sources are mainly organized around theoretical legal 'fields', and try to avoid redundancies. Therefore, even for a very simple contractual problem the answer will lie in many places. (...) One is involved in a car accident: there are the rules of the traffic code, rules of the penal law (code), rules of obligatory third-party insurance, and rules for the whole procedure, including the usage of forms, and so on. And even if texts are found, circumstances are fixed, and proper interpretation is in place, the question still remains: What follows from all these rules? What should I do? Where should I go? What should I write down, fill out, submit? Who should I inform, call? And so on."

In an *ideal case*, a knowledge base system can mediate between the two: *faithfully* represents the law (where faithfulness, as we saw, does not necessarily mean isomorphism, especially not verbatim as Johnson and Mead [20] advocate), and is built up in a way that with normative reasoning can help people find answers to their question in a real life problem. Expecting that the law itself is structured in a way that its verbatim formalizing (if it makes sense at all in terms of feasibility) provides all the answers to real life problems would be reasonable if we thought the legislative texts themselves as doing so. But according to Ződi [44]—and the experience—this is not the case. This mediating task is exactly something because of which building legal knowledge base systems would provide a real contribution. This requires legal

theoretical knowledge (and of course knowledge of the given legal domain) when constructing. Again, this is the ideal case, presupposing that we want a system for information extraction useful to laymen (too). It can be the case that our target audience is different: specific types of adjudicators. About this case, see Bench-Capon [4].

One might feel at this point that the requirement of extracting (regulative) norms demands too many sacrifices, so we should be satisfied with constitutive norms in their "normal, descriptive form", what is more, regulative norms should be rephrased as constitutive ones, too, in order to avoid any complication. Unfortunately, from a legal theoretical or philosophical point of view, consequentially leaning on constitutive rules also has a strong limitation.

5 Dealing with Constitutive Norms in Law and Its Limitation

As it has been mentioned earlier in this paper, there is a view advocating that there is no real difference between these two types of the norms. It might mean that what we call regulative norms, can be rephrased in the way 'constitutive ones' are phrased, that is, exactly in the other way around than we have pursued in this paper so far.

The possibility is appealing: we don't need to deal with normative reasoning or deontic logic, we have simple descriptive propositions. We cannot even argue with calling attention to the presence of deontic mo dalities since Searle only said that constitutive rules are *mostly* phrased in indicative mood, did not say that this happens always: the fact that we find rules using 'ought to' and 'obligatory' in the law of contracts supports this thesis: we all know that if we infringe some of this rule, what comes is not a punishment, but the nullity of the contract—which is actually not a contract then. And the solution of resolving regulative norms into constitutive norms seems to be easy: an obliging norm can be translated into a conditional sentence saying that if the given action is not conducted, then it is a violation.[8] Saying so we already have a paradigmatic constitutive rule: not performing the given action counts as a violation. If it is so easy to lean on constitutive rules when formalizing law, what is the problem?

Let's see a new example. The Hungarian traffic rules are a set of norms—we can call

[8]This approach is used when defining specifications for instance in LegalRuleML, see Athan et al. [2] and this thought was the basis of the Andersonian-Kangerian reduction of deontic logic, see Anderson [1] and Kanger [21].

their set as the Highway Code, in Hungarian we use the acronym KRESZ—that look like the classical examples of constitutive rules: the rules of a game. Some rules are general during "playing" (participating in public road use), but the obliging rules are mostly given in a very similar way it happens in board games: if you land on this and that square, you have to do this and that—if you see this and that sign (an octagon with the string 'STOP', down triangle with red edge, circle with red edge and a number in it, etc.), you have to (or should not) do this and that (stop, give priority, go with a specific speed, etc.). This and that (things in the world) count as this and that (pedestrian way, public road, traffic sign). Even the topographical feature comes, and given situations result in a situation when you miss a (or more) turn(s): your driving license is taken. For a moment we might become uncertain if we thought to 'walking' and 'going by car' as actions that we do anyway, that is, are not defined by these rules. Sure, we can do these: in our home or at our land, we can walk and use our car in any way we would like to, but the whole system of public road use is *defined* by these rules. In Hungary, traffic rules are contained in the 1/1975 (II.5.) joint decree of Minister of Traffic and Minister of Internal Affairs on public road traffic rules, and this statute—not like others—does not contain anything about the consequences of infringing the rules it sets. In Roman law this was called 'lex imperfecta': a law imposing no consequence on its breaching cannot fulfil its role, it is not a proper law. It of course does not mean that in Hungary, breaching traffic rules have no consequences: it has, but these consequences are handled in another law (mostly in the Act on misdemeanors). This structure strengthens the feeling that the traffic rules are just a specific set of rules of a game—the game called public road use.

But we should be cautious with this conclusion. This constitutive rules approach of law has an unwanted consequence: as it has been already referred above, constitutive rules have a definition in Rawls [30] according to which if we do not follow these rules, we do not engage with the action itself. But having said that, with breaching the traffic rules, we getting out of the scope of the given law (in case of KRESZ the decree above)—but then what serves as a basis to the policeman to impose a fine? Paying money is something that has nothing to do with the "game rules": we work, earn and spend money independently from the public road use, the word 'money' does not even appear in the text of the decree. It's clearly something that is out of the frame imposed by the traffic rules, we are still exposed to it by breaching them. It is because the system of law is not something discerptible from our life like football or chess is. It provides the structure of our life and—especially—our society in many ways, so thinking about its rules as merely constitutive rules just because there are institutions created by it comes with a controversial consequence. I don't

claim that constitutive rules as such make no sense in law: they do—especially the approach we find in Ross [32] about constitutive rules as a manageable and effective technique of presentation of a system of norms as we can connect a set of concrete circumstances to a set of legal consequences by them. But there are always legal consequences.

6 Conclusion

Pursuing extracting norms from legislative text is a reasonable task, just like the presupposition that it can be conducted. As we saw, at least in the Hungarian legislation, there are challenging examples of identifying and extracting the deontic character of the legal norms we would like to have in our knowledge base systems, especially if we would like to use (semi-)automatized methods. This task becomes even more challenging with that, as I argued, it is not reasonable to take isomorphism as a general requirement that should be met, as it not always can be met, even in the case of well structured legislative texts. The strength of this conclusion, of course, might vary according to the goal pursued by the formal representation, but in the case of extracting norms in a way it serves a legal knowledge base which faithfully represents the law and, at the same time, which might be subject of normative reasoning helping for example compliance of agents, being faithful probably won't (cannot) be equal with insisting isomorphism—even in the case of well structured norms. This good structure, though—having a labelled section providing the agents for instance—gives the opportunity to have another requirement somewhat looser than isomorphism, the algorithmizable correspondence using solutions like annotating the knowledge base slots with the locution in the source (legislative) text. Clusters of different—more specific—technical solutions might be developed according (and corresponding) to different legislative techniques. I also argued in this paper that leaning on constitutive norms—especially in their Rawlsian definition—when formalizing law brings very problematic result as breaching a rule would end up in getting out of the scope of the law, that is, per definitionem not being liable for breaching it.

References

[1] Alan Ross Anderson. A reduction of deontic logic to alethic modal logic. *Mind*, 67(265):100–103, 1958.

[2] Tara Athan, Harold Boley, Guido Governatori, Monica Palmirani, Adrian Paschke, and Adam Wyner. Oasis legalruleml. In *Proceedings of the Fourteenth International*

Conference on Artificial Intelligence and Law, ICAIL '13, pages 3–12, New York, NY, USA, 2013. ACM.

[3] Cesare Bartolini, Andra Giurgiu, Gabriele Lenzini, and Livio Robaldo. Towards legal compliance by correlating standards and laws with a semi-automated methodology. In *BNCAI*, volume 765 of *Communications in Computer and Information Science*, pages 47–62. Springer, 2016.

[4] Trevor J. M. Bench-Capon. Deep models, normative reasoning and legal expert systems. In *Proceedings of the Second International Conference on Artificial Intelligence and Law, ICAIL '89, Vancouver, BC, Canada, June 13-16, 1989*, pages 37–45, 1989.

[5] Trevor J. M. Bench-Capon and Frans Coenen. Isomorphism and legal knowledge based systems. *Artif. Intell. Law*, 1(1):65–86, 1992.

[6] Trevor J. M. Bench-Capon and Thomas F. Gordon. Isomorphism and argumentation. In *The 12th International Conference on Artificial Intelligence and Law, Proceedings of the Conference, June 8-12, 2009, Barcelona, Spain*, pages 11–20, 2009.

[7] Robert W. Benson. The end of legalese: The game is over. *New York University Review of Law & Social Change*, 13:519, 1985.

[8] Guido Boella, Luigi Di Caro, Michele Graziadei, Loredana Cupi, Carlo Emilio Salaroglio, Llio Humphreys, Hristo Konstantinov, Kornel Marko, Livio Robaldo, Claudio Ruffini, Kiril Simov, Andrea Violato, and Veli Stroetmann. Linking legal open data: Breaking the accessibility and language barrier in european legislation and case law. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law*, ICAIL '15, pages 171–175, New York, NY, USA, 2015. ACM.

[9] Guido Boella, Luigi Di Caro, and Livio Robaldo. Semantic relation extraction from legislative text using generalized syntactic dependencies and support vector machines. In *Proceedings of the 7th International Conference on Theory, Practice, and Applications of Rules on the Web*, RuleML'13, pages 218–225, Berlin, Heidelberg, 2013. Springer-Verlag.

[10] P.W. Brouwer. *Samenhang in Recht: een analytische studie*. PhD thesis, University of Twente, 6 1990.

[11] John J. Camilleri, Normunds Gruzitis, and Gerardo Schneider. Extracting formal models from normative texts. *CoRR*, abs/1706.04997, 2017.

[12] David Crump. Against plain english: The case for a functional approach to legal document preparation. *Rutgers Law Journal*, 33:713, 2002.

[13] Emile de Maat and Radboud Winkels. A next step towards automated modelling of sources of law. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*, ICAIL '09, pages 31–39, New York, NY, USA, 2009. ACM.

[14] Emile de Maat, Radboud Winkels, and Tom van Engers. Automated detection of reference structures in law. In *Proceedings of the 2006 Conference on Legal Knowledge and Information Systems: JURIX 2006: The Nineteenth Annual Conference*, pages 41–50, Amsterdam, The Netherlands, The Netherlands, 2006. IOS Press.

[15] J. Garcia. Constitutive rules. *Philosophia*, 17(3):251–270, 1987.

[16] Anthony Giddens. *The constitution of society : outline of the theory of structuration.* Social and political theory from Polity Press. Polity Press, Cambridge, 1984.

[17] T. F. Gordon. Oblog-2: A hybrid knowledge representation system for defeasible reasoning. In *Proceedings of the 1st International Conference on Artificial Intelligence and Law*, ICAIL '87, pages 231–239, New York, NY, USA, 1987. ACM.

[18] Pierre Jouvelot Hughes-Jehan Vibert and Benoit Pin. Legivoc—connectings laws in a changing world. *Journal of Open Access to Law*, 1(1), 2013.

[19] M. A. Jackson. *Principles of Program Design.* Academic Press, Inc., Orlando, FL, USA, 1975.

[20] Peter Johnson and David Mead. Legislative knowledge base systems for public administration: Some practical issues. In *Proceedings of the Third International Conference on Artificial Intelligence and Law, ICAIL '91, Oxford, England, June 25-28, 1991*, pages 108–117, 1991.

[21] Stig Kanger. New foundations of ethical theory. In Risto Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, pages 36–58. D. Reidel, Dordrecht, 1971.

[22] Jorgen Karpf. Quality assurance of legal expert systems. In *Proceedings of the Third International Conference on "Logica, Informatica, Diritto"*, pages 411–440, 1989.

[23] Angelika Kratzer. Modality. In Arnim Stechow and Dieter Wundelrlich, editors, *Handbook of Semantics*, pages 639–650. de Gruyter, New York, 1991.

[24] Eerik Lagerspetz. *The opposite mirrors : an essay on the conventionalist theory of institutions.* Law and philosophy library ; v. 22. Kluwer, Dordrecht ; Boston, 1995.

[25] Réka Markovich. Order of Norms and Deontic Modality. *South American Journal of Logic*, 1:435?445, 2015.

[26] Réka Markovich, Gábor Hamp, and Syi. A kondicionálisok problémája jogszabályszövegekben [Conditionals in legislative texts]. In *Proceedings of Hungarian Natural Language Processing Conference 2014*, pages 295–302, Szeged, 2014. University of Szeged.

[27] Heikki Mattila. *Comparative Legal Linguistics.* Ashgate Publishing, Aldershot, 2006.

[28] David Mellinkoff. *The Language of the Law.* Resource Publications, Eugene, Oregon, 2004.

[29] Henry Prakken and Joost Schrickx. Isomorphic models for rules and exceptions in legislation. In *Legal knowledge based system JURIX 91: Model-based legal reasoning*, pages 17–27. The Foundation for Legal Knowledge System, 1991.

[30] John Rawls. Two concepts of rules. *The Philosophical Review*, 64, January 1955.

[31] L. Robaldo and X. Sun. Reified input/output logic: Combining input/output logic and reification to represent norms coming from existing legislation. *The Journal of Logic and Computation*, 7, 2017.

[32] Alf Ross. Tu-tu. *Harvard Law Review*, 70(5), March 1957.

[33] John R Searle. *The construction of social reality.* Penguin, London, 1996.

[34] M. J. Sergot, A. S. Kamble, and K. K. Bajaj. Indian central civil service pension rules: A case study in logic programming applied to regulations. In *Proceedings of the 3rd*

International Conference on Artificial Intelligence and Law, ICAIL '91, pages 118–127, New York, NY, USA, 1991. ACM.

[35] Miklós Szabó, editor. *A törvény szavai [Words of the Law]*. Bíbor Kiadó, Miskolc, 2018.

[36] András Tamás. *Legistica*. Szent Istvn Társulat, Budapest, 2009.

[37] Peter M. Tiersma and Lawrence M. Solan. *The Oxford Handbook of Language and Law*. Oxford University Press, 2012.

[38] Luca Tummolini and Cristiano Castelfranchi. The cognitive and behavioral mediation of institutions: Towards an account of institutional actions. *Cognitive Systems Research*, 7(2):307–323, 2006.

[39] Raimo Tuomela. *The philosophy of social practices : a collective acceptance view*. Cambridge University Press, Cambridge, 2002.

[40] Judit Tóth. *Jogalkotástan [The Lore of Legislation]*. Corvinus University, Budapest, 2011.

[41] Robert W. van Kralingen, Pepijn R. S. Visser, Trevor J. M. Bench-Capon, and H. Jaap van den Herik. A principled approach to developing legal knowledge systems. *Int. J. Hum.-Comput. Stud.*, 51(6):1127–1154, 1999.

[42] Radboud Winkels, editor. *The OpenLaws project: Big Open Legal Data*. Internationales Rechtsinformatik Symposion IRIS 2015: 26.-28. Februar, Universität Salzburg, Wien: Österreichische Computer Gesellschaft, 2015.

[43] G. H. von (Georg Henrik) Wright. *Norm and Action : a Logical Enquiry*. International library of philosophy and scientific method. Routledge and Kegan Paul, London, 1963.

[44] Zsolt Ződi. The limits of plain legal language – understanding the comprehensible style in law. *International Journal of Law in Context*, 15 (to appear), 2019.

Received 15 October 2018