International Journal of
Language & Law

Vogel et al., JLL 6 (2017): 90–100
www.languageandlaw.de

"Begin at the beginning"
— Lawyers and Linguists Together in Wonderland

Friedemann Vogel, Hanjo Hamann, Dieter Stein,
*Andreas Abegg, Łucja Biel, and Lawrence M. Solan**

Abstract
What do patterns in legal language tell us about power, policy and justice? This question was at the heart of a conference on "The Fabric of Language and Law: Discovering Patterns through Legal Corpus Linguistics", convened in March 2016 by the international research group "Computer Assisted Legal Linguistics" (CAL²) under the auspices of the Heidelberg Academy of Sciences. About forty scholars from Germany, Switzerland, Italy, Poland, Spain and the US brought together their different intellectual and disciplinary perspectives on computational linguistics and legal thinking. Concluding the conference, four legal linguistics experts – two native linguists, two native lawyers – discussed the perspectives and limitations of computer-assisted legal linguistics. Their debate, which this article faithfully reproduces, touches on some of the essential epistemological issues of interdisciplinary research and evidence-based policy, and marks the way forward for legal corpus linguistics.

Keywords
corpus linguistics, law and language, legal linguistics, fabric, pattern, CAL², panel discussion

Editorial (not reviewed), first published in The Winnower 2016, republished in JLL: 7 September 2017

* *Vogel:* Institute of Media Culture Science, University of Freiburg, Germany, friedemann.vogel@mkw.uni-freiburg.de; *Hamann*: Max Planck Institute for Research on Collective Goods, Bonn, Germany, hamann@coll.mpg.de; *Stein:* English Department, Heinrich Heine University, Düsseldorf, Germany, *Abegg*: Center for Public Commercial Law, ZHAW School of Management and Law, Winterthur, Switzerland; *Biel*: Institute of Applied Linguistics, University of Warsaw, Poland; *Solan*: Brooklyn Law School, Brooklyn NY, USA.

The State of the Art in Legal Corpus Linguistics

Faithful transcript of a panel discussion on 19th March 2016 in Heidelberg, chaired by a co-founder of the International Language and Law Association (ILLA), Dieter Stein. The text was edited sparingly for legibility, and typographically emphasizes Stein's *moderation remarks* to distinguish them from his debate contributions. The transcript is identical with its prior publication in the open access journal *The Winnower* (DOI: 10.15200/winn.148184.43176) and was republished here merely for reference.

Dieter Stein: *What had you not gotten to know, had you not come here for this meeting?*

Łucja Biel: I learned a lot about the fabric of law as such. I really like the concept of the fabric, because I think it nicely combines with the way we think of texts, the new way we think of texts, the new way we perceive texts.

I also learned that there are different views as to what a pattern is. This is very meaningful because there are different types of patterns and in particular you can see differences between people who work in different disciplines. For example lawyers, like Larry Solan, work at a very high conceptual level, with the understanding of a pattern as a rule. Perhaps lawyers will be more interested in the conceptual structure and how it is patterned. Linguists and people who work with language acquisition, like Ruth or María, had a somewhat different conception of a pattern, working with n-grams for example, trying to extract multiword terms. It was also very interesting to see that computational linguists, like Giulia, can think of patterns as the depth of embedding and the complexity of embedding. So you have all these very different perspectives and we should think of some common ground to integrate all those views and all those levels of patterns, at different levels of language organisation.

That leads me to a question: Is there any universality of such patterns between languages? Because it was very useful to see how we work with different languages, and what kind of patterns we have in different languages, and I think there is some common ground between the languages. We have to deal with how to compare patterns between languages so that we make those comparisons meaningful and methodologically balanced. Each corpus has a different composition, a different structure and design objectives, and this also makes it difficult to a certain extent to compare between languages.

I also learned that we have the growing availability of corpora, the growing resources. If you read any literature on the use of corpora and legal language or in particular legal translation, the first thing you learn is that it is so difficult to work with corpora, because there are scarcely any resources. Now I learned that we have a lot of resources right now, and perhaps, we have to communicate to the people who might be interested in working with those resources, who might have good ideas how to use those resources.

It was also very interesting to see how people with different academic interests approach corpora and what they do with the data, to see how corpora can be used.

Larry Solan: One thing I noticed is: The first two speakers were Americans and we immediately started talking about Big Data and cases. 'Here is what happened in this case, here is what happened in that case.' And then everybody else is working in a civil law environment, and almost everybody immediately started talking about legislation. Of course, the Americans and the Brits and the Canadians and the Australians have plenty of legislation also, we do not even have much common law left in the United States. Almost everything's statutory, not everything of course, but almost everything. And yet, we orient ourselves around the judges as opposed to orienting ourselves around the parliament and the congress.

Getting back to our question: At the end of the last paper, when the question was, how universal is this, are the German judges, and who are they going to quote? The answer is: They are going to cite the German legal literature, in German, that is what they are interested in. Picking up on your comments, there really is a kind of duality here, there are certain tools that are available universally, depending upon what language you are trying to work in. Those are the data. Then the corpus tools seem to be pretty well developed, I mean, everybody who wanted to do something technically, came here and talked about what they did, and more or less accomplished what they wanted to, in terms of organising corpora. There are difficulties, and there are soft spots, and Google is a terrific thing to criticize for just the right reasons, but generally speaking, there seems to me to be a great deal of success: In the kinds of tools, in the kinds of analyses people do – with some level of statistical sophistication, which probably should be higher –, and then the data are growing. It would be great to have just a resource bank where everybody can know where everything is, it probably could be collected in a couple of months, if a grad student wanted to do that.

So all of that is good, but then what use you put it to, some of that is universal, you could talk about information that reveals the hidden underbelly of comported rule of law values generally. But my guess is, it is not going to really work that way. My guess is that individual legal systems will be using it, using these tools towards either internal advances from within the system, such as in the United States deciding cases in the way Stephen and I were talking about, or improving legislation, as others were talking about when investigating the relationship between the supranational system and the individual countries. This is very important because Europe is so concerned about such things. You actually gave somebody some good news, we do not hear that too much these days.

Then there are many other tools that could be used in a domain specific manner. I really doubt, other than intellectuals in international law or comparative law, that there is going to be an enormous interest in something that is specifically about Swiss legislation. If you decided to give a talk in the United States, there is a group of international comparative law people, they come to it, they find it fascinating. Similarly, without universalising the problems, people always like to see pathologies in the United States system of justice because it feels good. But generally speaking, I think the

usefulness of these tools are likely, from what I have learned today, to be more domain specific. The tools themselves look like there is a big sophisticated international community of people who just know a lot about this stuff. That was really revealing and quite exciting for me!

Andreas Abegg: It was indeed very interesting to see the different projects and the creative ideas on empirical linguistics and law. With this very new method of empirical linguistics at hand, it is of great value to exchange on possible fields of application, to exchange on what approaches work (or do not work). It might be fruitful to establish a network or a site to collect and share ideas and to know about the current and finished projects. Such a network or site might also help to enter into new collaborations.

Furthermore, the different approaches by scholars from common law and continental law were of great interest to me. Common law lawyers do not find it difficult to immediately connect an empirical linguistic method to their case law. However, as continental law scholars, we cannot just concentrate on case law, but we always have to connect to legal principles which guide continental law. From our history, our path-dependency, we are much more into scholastic deduction. This makes it more difficult to use an empirical method. Therefore, because we do not have this immediate access to empiricism, there is a need for continental scholars to work on a theory that links empirical linguistics to the legal methods.

Dieter Stein: Maybe I will myself provide one or two remarks: For me, this is still kind of a new field, and if you have a new field, you have a situation where things are meandering a little bit until they really fall into place. I believe, this is the time now to establish some sort of a meta-theory of what we are doing.

You see, we have a bottom-up aspect here, we have many people, having wonderful work on corpora. But then arises the issue: What are we doing with this corpora? So what comes first? Do we construct the corpora first or do we first ask our questions and then construct the corpora? This is kind of a top-down perspective. And I would like to see a matching of those two perspectives. That is what I believe may still be something that we need to work on.

The second aspect to me is: Of course I was intrigued by the way legal language – language of law – just does not exist. You have a number of legal genres that are pretty much separate and these appear to me to be separate also in different countries. I was much intrigued by the work on evaluative elements in judgements by Stanisław. I would imagine this is not the same in all countries. What I would like to see is the theoretical instrument of genre, sharpened and applied to the analysis of legal language.

DOI: 10.14762/jll.2017.090

Open Debate: The Future of Legal Corpus Linguistics

Dieter Stein: *Let me now open the floor for discussion: Everyone can discuss, everyone can chip in, the audience is invited to comment.*

Larry Solan: I would like to respond to Andreas first and to you, Dieter. [...] I agree with both of you: This really requires nothing other than collaboration. I was teaching at a university, a few years ago, as visiting professor. Maybe ten years ago, they got their fMRI machine, somebody gave millions of dollars and they get and haul this thing in. It is very expensive to keep up and everything. They had no idea what to do with it. So they put signs up, 'Anybody have an experiment that you want to do with brain-imaging? That's great!', and then some people would sign up. Now it is really pretty sophisticated. One thing that the field is crying out for, is a collaboration between the identification of issues in law.

They could be very practical issues: How can we draft statutes that you can read? In the United States that is not much of a concern, we do not care whether you can read them. We care about whether they are precise and that there is not going to be ambiguities, but we do not care whether they are comprehensible. That is what lawyers are for, to spend their time reading them.

Or they could be at a high level of theory, they could be quite abstract, but it seems to me that most of the research is in the service of improving various aspects of the legal system. It could start with basic research, it does not have to have practical ramifications for the first generations. There is nothing wrong with that. When research funders require immediate gratification through practical consequences, that is a bad thing sometimes because it stifles basic research values. But it seems to me that this is really a direction.

Now, to the extent that this is work that just happens to be in the law – because language for special purposes and corpus research generally is something that people are interested in and law is just a nice domain to do it in, because there is learning within the linguistics – nothing what I am saying really applies. But to the extent that linguists sometimes are frustrated by the lack of attention they get from the legal community, it is not easy to start with a perspective that the target community is going to be impressed with initially – unless you work with them initially. Then it becomes that kind of collaboration you want.

That is really the only thing I can think of, as a direction, that seemed not altogether missing here, but I think that people are craving more of it even in their own work.

Lucja Biel: Drawing on what Larry has just said: We have to think of ways how to increase the uptake of corpora by the legal community, because right now we know how to use it for research. We have some applications for teaching students for example, we have corpora for training translators. However, there is still a problem with the uptake of corpora among the professionals, especially in the legal field. I think it would be in-

teresting to invite more lawyers to collaborations, to see what they need, and what they expect, so that our research can be more meaningful to them.

Andreas Abegg: I very much support this. I found it fascinating to work with a linguist, because he knew the tools or methods and I thought about the relevant questions that could be asked. Such a collaboration can be a very fruitful, very creative work, in course of which many new research questions may be discovered.

Hanjo Hamann: To relate to that, yesterday's second presentation, by Stephen, nicely told the story of how he basically brought corpus linguistics into the courts, which is a good illustration of that: 'I told a judge that this [corpus linguistics] is there, I told him how to do it.' – and this basically set off an entire avalanche of work in legal practice. I take it there are at least two people here who will attend the conference at BYU in April on "Corpus Linguistics and the Law", and as far as I know, the roster of participants contains a lot of lawyers who probably haven't done a lot of corpus linguistics. My hope is that this will influence legal scholars and judges in the US. For us Europeans it is easy to take it once Americans have taken it up, because then our research institutions will gratefully fund things. All the originality that comes from Europe aside, there is still a heavy dependency on role models, I guess. Americans are often role models in what they do, so we have always looked to them after the Second World War, because their research is often ahead by ten or twenty years. So I think one path this will go through is through American lawyers taking it up in their court decisions, and German and European lawyers see that and transfer it to their domain. In Europe we have to try to inspire judges and lawyers, which is not as easy, because they are not as open to social science matters in general and linguistics in particular.

Dieter Stein: You know, I have been trying to persuade our Düsseldorf lawyers to come and let me do service for them. Linguists are in a position where they are often talking to lawyers: "Why do you not come and love us?" The thing is, there is a wall between lawyers and linguists, and the name of this wall is ideology. It is an ideology of language. This is what I find very hard: To persuade lawyers to not pursue their ideology of what language is, what words are, what meanings are, and so on. I think there are two ways of handling this. One way is to try and educate lawyers. That is totally futile. Do not even try. The other way for us is to condescend and say: "Okay, we try to speak your language." That borders on prostitution in a way, does it not? [Addressing a lawyer in the audience:] What is your impression, as a lawyer: Am I misrepresenting you?

Ralph Christensen (Mannheim): No, no, you don't. You have to start the conversation. It started in Germany. Brothers Grimm were both lawyers and linguists. And now the lawyers have the power and the linguists are the ones who have the knowledge…

Dieter Stein: They have got the guns!

Ralph Christensen (Mannheim): … and we have to get these back together.

Friedemann Vogel: But I think this is a second problem. It is not only the differences of language ideology, but also law is connected to power. Linguists and social scientists explore what lawyers do, and claim they could make it better, maybe, and this is not only a question of methodology. The question is: Who can speak with whom about power? And law is power, it is the fundamental structure of sharing power and controlling power. So I would be interested what you think: If lawyers came to me and told me 'Nice work, but show us what you have. Here you have to be more normative. I will show you a better method.' and so on – I would not be all too happy about that.

Andreas Abegg: I am not very worried here, because there are so many different levels we could collaborate and benefit each other. If a linguist would come and describe with my corpus how the language developed and how the court used words, I would be fascinated. That would be a contribution to legal theory. It would probably not be taken up by the Federal Court, nor by an article, arguing how you should construe some kind of statute. But so what? It would still be very valuable.

But then again we have those topics where really both disciplines align, as Stephen has told us with the example of the grammatical interpretation. There we are very close. I could think about the use of words and patterns for example. That is a very direct use. We have both our competences aligned directly. And I think we have to try to be creative and then find other ways to collaborate, other topics. It is a fascinating time. I feel that everything is possible at the moment.

Friedemann Vogel: This is the question. Is really everything possible? Or is it – in a critical view – only a game, where we can play, play with legal texts, but with no impact on the practice, where power is made and reproduced? Look to Europe at the moment. We are in a really difficult situation, and nobody knows what the results will be in the next years. What could we do? Could we contribute in this situation with our perspective?

Dieter Stein: We are convinced we could and we should. In fact, we must. But they have to let us, you see.

Larry Solan: I have to say that, at least with respect to translation theory, the EC spends something between 500 and 600 million euros a year on translation. You talk to the translators and they feel like they are sitting in a room with no windows. The translators all feel oppressed. Everybody wants more efficient translation and everybody wants translation where you do not have too many legal problems. The truth is, at least with the regulations and directives, you really do not have that many legal problems. You do not have that many cases coming up where that is a big issue. You probably have them coming up in international courts and nobody notices it. That probably happens. You do not find them in the court of justice of the EU. You find eight cases a year, or something like that. That is not many for a big society like this.

So when you find both, the resources from a linguistically trained group and the corpus perspective, infiltrates only to some extent, but I am talking more generally. Society is feeling the need to have more sophistication with respect to language analysis. I think these people have a fair amount of influence on translation procedures, and there is serious debate about it and there is much less of a gap between conferences, legal translation within Europe and the people who consume it, namely the commission. There are always people from the European commission at these conferences, they are often the keynote speakers. There is a real collaboration in an area like that.

Then you get the statutory interpretation or the interpretation of contracts. The Americans have these weirdest traditions with their dictionaries. I remember once I was consulted on a contract case. The lawyer said: "So I have this linguist, he is a law professor, he gets…" – "I do not need your Henry Higgins telling me how to speak English!" So that is the power relationship that you have, and that happens. You really do not want linguists in court every time anybody is having a dispute about what a law means. You need to hire a bunch of linguists, and you probably do not. So you need to learn the right way to take care of exactly these unusual cases, but they are not totally rare, they are just once in a while.

To me, the challenge here is what you were talking about, Stephen. It is about replacing looking at six different dictionaries, which are just the luck of the draw given that these are all borderline cases of concept formation – who knows which one is going to hit the jackpot, for this side or that side –, and replacing it with the legal system taking lexicography seriously through the data that you have, which everybody in this room knows how to use well. It is about substituting good analysis of word usage for a snapshot that a lexicographer wrote when they are given an average of three lines per word and they plagiarize anyway. If that substitution succeeded, that would be really helpful.

It looks like with these conferences, like the one that BYU is running, there is some chance at least over there of it happening and conceivably, then coming over here and talking to lawyer groups. It can spread in a way that Hanjo suggested. At least that point can. The translation points can. I think assisting in legislation already is happening here. So it happens opportunistically. We need to identify the problems within the legal system in a way the legal system will find it welcoming. That seems to be going on to a greater or lesser extent, depending upon the project.

Dieter Stein: *I think it is interesting that this should develop into a discourse on power, ultimately.*

Friedemann Vogel: This is the point. Power is the point that is important for me. I wonder if it would be more than a symbol to ask or to create an internationally united European corpus of legal language. Do we need such? And how could we proceed?

Dieter Stein: *This would be a top-down interest, in fact.*

Friedemann Vogel: And it is obviously relevant. What do you think?

Larry Solan: That is a perfect example of a project that really needs collaboration from the beginning. Without it, it is a big risk. With it, it is something that is still a risk, but not as much anymore.

Dieter Stein: So this will be one of the take-home bottom lines from this conference that we could subscribe to. And the other is: My impression is that all the lawyers feel they are being, in a way, deconstructed if linguists talk to them. Don't you agree? […]

Stefan Höfler: If I may contradict you a little bit, I am not quite convinced. I am not sure if I buy into this argument about power. I am not sure if it is really a matter of power. I'd rather say that it is a matter of communication. In my experience, what is important is that we, as linguists, do not go to lawyers and tell them where their problem is. First, we have to listen to them and try to figure out where they think their problem is. And then we will see whether we can support them. That is not a power struggle, but a struggle for communication and for trying to understand each other's worlds. I personally think this is a much more fruitful way of looking at the situation.

Dieter Stein: Can I just support you? The godfather of one of my children is actually a lawyer. I was trying to discuss these issues with him. He replied: "*Was wollt ihr denn? Es läuft doch!* – What the hell do you want? It is all alright. We do not need you, really. What is wrong with us?"

Stefan Höfler: Let me look at the situation from my perspective of the problem, legislative drafting: If I go to a lawyer and if I tell them that they should really write clearer laws, because everybody should understand them, then obviously the lawyer would say: "Bugger of!" So that argument does not work.

Dieter Stein: That is a democratic Swiss concern: "*Populärdemokratie*" [direct democracy]. We do not care in Germany.

Stefan Höfler: But if I go to the same lawyer and I explain to him, how his own work will become easier, because a statute is written in a clearer way, then he will be much more open to my suggestions and that is the way I think we should go forward.

Dieter Stein: I think we all think that way.

Victoria Guillén-Nieto (Alicante): I completely agree with you. I have been attending this conference and I think it is fascinating. As for the methods that are applied, and the way the corpora are build, it is really wonderful. But I can see some weaknesses.

The first weakness is that even if we joined together in the creation of an international legal corpus – which I am very much in favour of – what is the purpose? What is the purpose of gathering a corpus, what do we need this corpus for? Since we moved into this society of information and knowledge, the way we set our hypotheses is not

just academic. It has to be socially and professionally relevant. 'Begin at the beginning', the King said bravely, 'and when you come to the end, then stop.' This is Lewis Carroll in Alice in Wonderland. But I do think the "first thing" is to organise a group together with professionals and listen to them. And then, once we listen to them, we can really brainstorm and gather lots of ideas. And then we can focus on the sort of corpora we need to build and the sort of hypotheses we need to set up to make sure that apart from being academic, they are professionally and socially relevant. Otherwise the findings, whatever we do and however great and wonderful it is, will not be transferred to the society of information and knowledge.

If we establish this relevance, we can also get funds. I can tell you about an experience I had years ago, which had nothing to do with the legal language, but it had to do with intercultural pragmatics. Begin at the beginning, we organised a group, together with people who were in international business, we found out the red lights, we did research but at the same time, the findings of this research were transferred into the creation of a graphic adventure. The graphic adventure thing attracted 60,000 Euro, which is something that no one at the faculty of humanities at the university Alicante had dreamt of. It was relevant. It was academic, but it was socially and professionally relevant and we think: 'Begin at the beginning' is to group together with professionals. This will really help us to find the purpose, because we already do very well in the methods and in the building of corpora.

Dieter Stein: *I think here is another take-home message.*

Victoria Guillén-Nieto (Alicante): And I am very much in favour of organising this international team and constructing an international corpus…

Łucja Biel: … that will let us look for patterns above specific national languages. […]

Dieter Stein: It would be something like the successor to Eurotyp. Remember that, Eurotyp?

Hanjo Hamann: I agree with that, but I also want to put it in perspective, because I think the way corpus linguistics has proceeded in many cases is top-down, in some cases with questions. Then you assemble your corpus and then you throw the corpus away and it is forgotten. What I think is missing is an infrastructure and you cannot define exactly the task of an infrastructure because there are so many ways to use it. Whatever I think of the EUR-Lex collection of documents at the European level, this started out as an infrastructure to make legislation public and the EU transparent. And I think meanwhile it has found so many applications in corpus linguistics and law and other areas, which were never intended. It was always 'Oh, there is this material, what can we do with it?' – and suddenly you find a wealth of research questions that you can answer with that. And in that sense I think, even building a corpus without some specifications of questions, can be useful as an infrastructure.

What I find most challenging in our projects is that all databases that we as lawyers have are good as long as you 'do it the way they always did'. That is: I look for a single document, which I want to read. How can I get to it most quickly? And they [our current databases] are good for that. They are appropriate. But they are never set up in a way to look at a number of documents in a general frame 'from above'. Take as an example the things that I showed in our presentation about the quality of the juris data: I said there are something like eleven court decisions that have wrong page numbers. That is in a universe of 9,000 court decisions. Nobody would care if we did not do it from a bird's-eye perspective. But being able to change the focus from 'databases are just for retrieving documents that you can then read', to a perspective like 'databases are collections that you can look at on a microscopic level' is something that is missing entirely from current databases. That is a sort of infrastructure that we will need. For example, if EUR-Lex came with a KWIC, a keyword-in-context display with collocates and everything: That would help in so many ways, even if we do not have the research question yet to address with this. But just extending the infrastructure so we can do these things at all, I think, would be helpful.

Dieter Stein: *Thank you very much. I think this is a wonderful concluding statement with practical advice. Our time is nearly up. This is the time, as we were talking about power, to use my own personal power to thank you for putting up this wonderful conference, and thank you for your contributions.*

Have a safe journey home!