Auditory motorcycle recognition by experts

Heinrich Wack, Gea deJong-Lendle, Roland Kehrein, Kjartan Beier, Ramona Kaul, Paula Rinke, Lukas Cohrs, Laura Schmidt

Universität Marburg, Germany

One of the regular tasks of a forensic phonetician is a detailed transcription of an audio recording. In a typical transcript the verbal content of the disputed material is transcribed in detail but the (possible) identification of non-verbal audio-material may be equally important in solving a criminal case. For example, it may be useful to know, whether the door of a car is opened/closed (indicating the number of possible speakers on the recording has now been reduced/increased), whether the noise in the background comes from windscreen wipers (indicating the caller is sitting in a car at the time of the call and it is raining) or whether the sound on the anti-theft camera is the sound of a particular motorbike driving away. In this experiment, we raise the question: How well are people able to identify the noise of a motorcycle, if they only hear it starting and driving away?

As the organisers of the 47th ADMV Zweirad-Rallye-Meisterschaft "1000km Langstrecken-fahrt durch Deutschland" were still looking for small exercises that their drivers could carry out on route, it was decided to offer a short motor recognition task that could also serve as a pilot experiment for the research question asked. As this rally is rated as one of the hardest rallies in Germany, technically and physically, it is assumed, that the participants can be considered motor cycle "experts". (Non-experts will be tested in a subsequent experiment.) On the 22th of June 2019, a total of 61 motorcyclists (58 male, 3 female, between the ages of 26 – 77 years) were tested. 58 listeners also drove their engine (43 large-engine motorcycles and 15 small-engine mopeds), the remaining 3 listeners were passengers in a large-engine two-seater.

The stimuli were recorded using a ZOOM H4nPro Handy Recorder (Stereo, 44.1KHz, 32-Bit). The drivers of the motorcycles/mopeds of interest were asked to start the engine and subsequently drive away in a normal pace. The recording environment consisted of a quiet road in a rural area away from traffic noise. Background noise only consisted of fairly quiet high-frequency bird noises: it was assumed that these affected the perception of the low-frequency engine noises only marginally. As a large number of listeners had to be tested in a short period of time and as queues should be avoided, the duration of the exercise was kept short; 5 engines were selected and stimuli were created varying from 14-21 sec. in duration including the starting and driving noise. Subsequently all samples were placed in one file with 5 seconds and an announcement ("Aufnahme eins", etc) between each sample. Each engine sample was normalised separately, in order to present the quieter engines with a similar intensity as the louder engines. The selected engines should create a representative sample of those that currently exist or have existed in the past 40 years throughout Germany; 2 or 4 stroke and 2 or 4 cylinder engine types were included and those typical for former East- and West-Germany.

The listeners were asked to listen to 5 stimuli, that consisted of a particular motorbike sound and to select the correct bike from a selection of 12 motorbike pictures. Between each stimuli the listener had 5 seconds to decide. The entire test-recording lasted around 2:00min. Results: The analysis showed that over half of the participants (i.e. 52%) performed very well, identifying 4 or all of the engine noises they had heard correctly. None of the listeners had a score of 0% and only 3% of the listeners identified only 1 engine correctly.

Based on the results of a pilot experiment with 61 motor bikers, a preliminary conclusion is that over half of the people with experience with motor engines can perform auditory motor cycle recognition quite well. However, being an expert does not necessarily mean, that you have the motor identification skills: around 20% of the experts show rather poor recognition scores.

Identifying a fake non-native speaker competence in authorship profiling

Eilika Fobbe

Bundeskriminalamt, Wiesbaden, Germany

Authorship profiling is a typical task in forensic linguistics mostly applied in the early stages of a criminal investigation. Its primary purpose is to gain background information about the anonymous author. By analyzing texts (mostly threatening letters, ransom notes or extortion letters) on their different levels of language, authorship profiling tries to extract from them socio-biographical information, thus helping the investigators to determine their further course of action.

This presentation focusses on the writing competence revealed by the author and is grounded upon findings from different research fields of Linguistics, such as Sociolinguistics, Text Linguistics, or Language Acquisition. The central hypothesis is that the acquisition of written language competence and the level of its mastering depend on inflicting social conditions such as a person's (local) origin, educational background, mother tongue, profession, and social relationships. Because the development of writing skills does not end with adolescence but is a lifelong continuing process, one might expect that a person's writing competence will reflect his/her experiences linguistically.

The anonymous author's writing skills are evaluated against the norms of written standard language. The analysis covers language use, style, text structure, errors, and misspellings. The expectations of the norm may be met or not; in either case, this will lead to hypotheses on the author's social background and will call for an explanation. Usually, the results should not be contradictive but should form a consistent linguistic profile. If there are contradictory findings, they emerge from linguistic (and at times content-related) inconsistencies. These inconsistencies point to a possible authorial strategy that seeks to hide potential linguistic traces by manipulating the author's genuine writing competence. One of the strategies applied is the manipulation of one's native-speaker competence. If a person's written language shows specific morpho-syntactic and lexical errors as well as misspellings this would not only imply lacking writing skills but a lacking language competence as well. A linguistically untrained person could infer from this that the perpetrator in question must possibly have a foreigner's background. However, even if it seems clear that the presented writing competence is a faked one, still a scientific explanation has to be provided to support this impression. Errors and mistakes result from many different reasons; therefore, it is crucial to determine which mistakes are potential candidates of second language acquisition, and which of them are not. A sound determination would have to prove first that some errors only occur with non-native speakers and others only with native ones. It is well documented that some types of errors are common among second language learners, e.g., errors in word order or the wrong application of masculine and feminine articles in languages that have the category of grammatical gender. The fact that it cannot be ruled out that these errors happen to native speakers as well (then representing rather mistakes) makes a different approach to be more promising. If we consider the perpetrator's strategy to distract from his language competence and to point to somebody else's his primary aim is actually to obfuscate belonging to his group. The group of non-native speakers the perpetrator pretends to be part of is therefore of much lesser importance and, in many cases, lacks any detailed description. Only in very few cases, the perpetrator has an actual person in mind when imitating a non-native speaker (Fobbe 2014). The language the author uses for his disguise is the realization of a simplified register, also called foreigner talk. Simplified

registers are varieties of language that refer to the cognitive or linguistic capabilities of the recipient. They help to preserve a fragile communication situation from breaking down by reducing the language's complexity when one of the participants has not the necessary language skills (Fischer 2016, Ferguson 1981). Simplified registers show modified features on all linguistic levels (e.g., articulation, syntactical structure, and lexicon), but only some of the features are simplified, while others (even complex ones) remain intact. Applied to a written text this specific register will evoke an impression of linguistic inconsistency in the reader. Research suggests that foreigner talk and inter-language differ from one another because they originate from different sources, although there are some shared phenomena in both varieties. A learner's inter-language is shaped by the learner's conditions and will evolve with the learner's progress while foreigner talk depends strongly on the structure of the language it simplifies. Studies have found its realizations among native speakers to be very similar both in feature selection and feature variation. Making use of the simplified register in the context of authorship profiling to detect deception will help to evaluate identified inconsistencies in writing. If the writing competence of the perpetrator in question displays modified and correct language features in combination as they occur in foreigner talk, then this can support the deception hypothesis.

The presenter starts with an introduction to authorship profiling, then provides an overview of deception strategies by focusing on the manipulation of written language competence. The talk then presents typical characteristics of the German variety of FT and explains to some detail selected examples from the database of forensic texts from the BKA.

Freedom of speech or the creativity or verbal assault? Forensic linguistics in press trials in France - a summary

Dominique Lagorgette

Université Savoie Mont Blanc / LLSETI

Based on a corpus of linguistic reports produced at trials in France (2004-2018), this paper will describe how discourse analysis applied to forensic linguistics is received by magistrates and lawyers, and which perspectives exist for this discipline and its research. It will focus on speech acts linked to freedom of speech as ruled by the 29/07/1881 law, which makes a distinction between "injure" and "diffamation".

Catching predators: a forensic linguistic profile of child sex offenders' texts

Benjamin Duncan

National University, San Diego

The increasing use of social media, text messaging, and anonymous online chat forums has simultaneously advanced the ability for child sex predators to plan and perpetrate their crimes. Electronic devices have expanded our social and communicative abilities, but sadly, it has concurrently allowed predators a direct, yet concealed, pathway to groom and sexually exploit their victims. With increasing regularity, forensic linguists are called to provide expert testimony in legal cases involving disputes over the alleged authorship of online texts. This study uses two open software natural language processing tools, WordNet (Miller, 1995) and Wordsmith Tools, v.22 (Lexical Analysis Software, 2018), to compare a corpus of online texts between convicted child sex offenders and undercover police impersonating children (n=100,000 tokens) against a control corpus of online texts between two child authors (n=100,000 tokens). The purpose of the study is to identify idiolects associated with potential child sex offenders. Features at the lexical-, grammatical-, and discourse-levels are described and separately assessed to determine their validity in profiling online predators.

"What else can you do to pass…?"

A pragmatics-based approach to *quid-pro-quo* sexual harassment

Victoria Guillén-Nieto

University of Alicante

EU laws define sexual harassment as a "Form of gender-based violence encompassing acts of unwanted physical, verbal or non-verbal conduct of a sexual nature, which have a purpose or effect of violating the victim's dignity and creating an intimidating, hostile, degrading, humiliating or offensive environment". (European Commission 2010; Council of Europe 2011). The European Union Agency for Fundamental Rights (FRA) (2015) has recognised sexual harassment as discrimination on the grounds of sex and as a breach of the principle of equal treatment between men and women in the workplace. This paper focusses on a specific type of sexual harassment: "Quid-pro-quo sexual harassment". This refers to situations in the workplace in which an individual who holds power offers or merely hints that s/he will provide advantages or withhold disadvantages in return for the target victim's satisfaction of a sexual demand.

Sexual harassment is a very common but generally underestimated type of gender-based violence. It is, in effect, typified as a minor crime (misdemeanour) in the laws of modern societies all over the world despite the severe consequences it may have for the target victim. At times, neither perpetrators nor victims, especially in patriarchal societies driven by sex-based ideologies of discrimination, are able to recognise and label incidents as sexual harassment. Other times, when the conduct is identified as such the victim of sexual harassment is often reluctant to report it to the hierarchy at work or to file a complaint due to embarrassment, fear of not being believed, retaliation or lack of trust in the organisations. And when the victim dares to report, proving an episode of sexual harassment before the Administration or in a court trial is not an easy task because unlike felony sexual assault, sexual harassment leaves no ostensive trace in the target victim. On reviewing the literature, we found that while sexual harassment is well documented in the fields of psychology, sociology and law (Maas, Cadinu & Galdi 2013: 341-358), it has not yet received sufficient attention as a scientific object of study in linguistics; in fact, Cotterill's (2007) *The language of sexual crime* and Shuy's (2012) *The language of sexual misconduct cases* are two of the few references existing on the topic.

Drawing on the premise that perpetrators of quid-pro-quo sexual harassment are bound to commit the crime indirectly for the major reason that they want to avoid accountability, we discuss two questions: Can linguistics help us detect indicators of quid-pro-quo sexual harassment? If so, what type of evidence can the linguist collect in the language of the perpetrator and of the victim to show that there are indicators of sexual harassment? Our ultimate goal is to propose a pragmatics-based approach to the analysis of cases involving quid-pro-quo sexual harassment. This methodology comprises the following linguistic framework: (a) speech situation, e.g. the social context of interaction, the sequence of expected behaviours for a given situation (behavioural scripts and schemas), (b) discourse structure (turn-taking system, move structure, speech acts and topic control), (c) discourse strategies, e.g. indirectness, and (d) referencing and ambiguity. Our model of analysis is illustrated through a paradigmatic case study involving quid-pro-quo sexual harassment in a campus university in 2016. When the victim reported the episode to the Administration she provided a 20-minute audio digital recording of the incident. The

transcription of this piece of physical evidence turned to be a key element in identifying and labelling the episode as quid-pro-quo sexual harassment.

"I just got called by a lawyer who wants me to help in her case": seven ethical issues of concern in serving as an expert witness in litigation

Janet Ainsworth

University of Seattle

Lawyers as advocates for their clients obviously play an important role in litigation—a role strongly mediated by a complex and legally-binding set of ethical rules, the Rules for Professional Conduct. Violations of these rules can result in serious sanctions, up to and including disbarment. Expert witnesses, too, play an important role in litigation, because accurate fact-finding often requires scientifically-based information critical to reaching the right verdict in the case. However, lacking an enforceable set of professional rules of ethics, scientific experts often are insufficiently attuned to the ethical issues presented during the unfolding of the litigation process. This presentation will highlight seven common aspects of the legal process that potentially raise ethical issues for expert witnesses: 1) the ethical issues involved in being retained by an attorney for a party, 2) the ethical issues involved in turning down participation in a case, 3) the ethical issues involved in expert witness compensation, 4) the ethical issues involved in analyzing a case, including confirmation bias and motivation bias on the part of the expert, 5) the ethical issues involved in preparing to testify under oath, 6) the ethical issues involved in drafting expert reports, and 7) the ethical issues involved in communications during the trial. The role of the attorney is to single-mindedly serve the interests of their client; the role of the expert witness is to single-mindedly adhere to and faithfully represent their science. The intersection of these roles is the source of ethical challenges to the expert witness; awareness the potential for ethical issues is key to fulfilling the promise of science to the legal system—to fairly and responsibly bring to bear evidence-based science in the service of justice.

Using textometry and semantic forms to characterize documents and help with authorship attribution: a digital semiotic analysis

Julien Longhi

Cergy-Pontoise University, France

Institut universitaire de France (IUF)

In *Distinction*, Pierre Bourdieu (1979) aims to describe the process of entering the social space, and the question of the passage from abstract situations to objective structures, which are nevertheless provisionally reified. From a linguistic point of view, discourse is a field that semiotizes one's understanding of the social world, and in the same way that "talking of habitus, it means including in the object the knowledge that agents who are part of the object, have of the object, and the contribution that knowledge brings to the reality of the object", talking about discourse will include knowledge that authors have about the social world and the semiotic contribution they bring to their discourse. To sum up and expand on Bourdieu's social structure, we see that, for authors, there is a unity between the perceived social world, their action in the social world, and the expression of the social world. This can be shown by analyzing the "construction of semantic forms as a micro-developmental process, which involves more or less stable 'phases'" with, for example, "linguistic motifs that may be viewed as unstable germs of form" (Visetti and Rosenthal 2006). To do that, textometry offers an instrumented approach for dealing with corpora, articulating quantitative syntheses and analyses including text (Lebart and Salem 1994). This approach highlighted the similarities and differences observed in a corpus according to the representation dimensions considered (lexical, grammatical, phonetic, prosodic, etc.) (Pincemin 2012). By focusing on corpora related to criminal activities accessed by the presenter through professional collaborations with the Judiciary Pole of the French Gendarmerie (PJGN), we will show the relevance of textometry in forensic linguistic analysis, because it involves the same process of collecting, analyzing, interpreting and reporting linguistic evidence in a forensic context. This paper will present different methods and tools that can assist the expert linguist with authorship attribution tasks.

The defenses of infringements of reputation right in China ---Perspective of the expert linguist

Luping Zhang

China University of Political Science and Law

Defamation law is not sufficiently clear. If there is an area of law in need of linguistic clarification, this is defamation law. This presentation will provide an illustration of this problem based on real life empirical data in a defamation case. The most basic postulate is that language (behavior) is to be described in sociolinguistic terms: it depends on who says what to whom, when, in which context with which intention and to which end. The paper will firstly analyze the position of the auxiliary expert in Chinese civil actions from the relationship between auxiliary experts and litigants, auxiliary experts and courts, as well as auxiliary experts and examiners. The paper will then introduce the two major defenses of infringements of reputation right, namely defense of truth and defense of honest opinion stipulated in Common law and also embodied in China's judicial practice, with the discussion of a famous defamation case, i.e. Beijing Qihu Technology co., Ltd. v. Wang Xiaochuan case.

Can corpus linguistic analysis make a science of legal interpretation?

Lawrence M. Solan

Brooklyn Law School

Corpus linguistic analysis has become a popular tool in the realm of legal interpretation. Whether the document in question be the U.S. Constitution, a statute or a contract, linguistic experts and some legal analysists have been turning to corpora to determine how disputed terms are used and understood in everyday discourse. In this presentation, I will point out the extent to which the use of corpus analysis requires the exercise of considerable judgment about the corpus to use, about the searches to conduct, about how to interpret the results, and about the actual goals of corpus analysis. Some of these judgments are linguistic, but for the most part they stem from having to decide the utility of drawing legal inferences from the prevalence of one usage over another in a corpus. For example, while the law defaults to the "ordinary meaning" of terms in a law or contract, this practice is defeasible. The corpus can tell us nothing about whether to use the default principle in a particular case, or whether to adopt a different sense of a term because the context suggests that the drafters had a broader or narrower meaning in mind. Moreover, laws sometimes contain terms in a rich linguistic context. Corpus analysis tends to ignore this context in favor of determining the contexts (by means of collocates) that typically occur with the term in question. Making the task more difficult is the fact that the absence of a particular usage in a corpus may be a linguistic fact (people don't speak that way) or a fact about the world (people don't talk much about that, but if they did, they might use the language in question). The presentation concludes that corpus analysis can be a useful tool in legal interpretation, but that it does successfully convert a practice of individual judgment into a science, as is sometimes claimed.

The language of love: romance fiction in the BNC

Hans van Halteren

University of Nimwegen

It is generally accepted that the form of a text is influenced by various factors, foremost among which are: a) who wrote it (*author*), b) what it is about (*topic*), and c) which form was chosen to optimally reach the intended effect in the intended target audience (*communicative situation*, more or less also covered by the terms *genre/register*). In Forensic Linguistics, it is obviously *authorship recognition* that is our main interest, and we have found that both genre/register and topic are strong confounding factors. In Natural Language Processing (NLP), *topic detection* is the most studied aspect. *Genre recognition* has drawn less attention so far, and is often limited to high level distinctions, with broad classes like fiction and academic writing. We would like to advance all three classification tasks by mapping which classification features are bound to which of the three factors, and to which degree. Before addressing this ultimate question, however, we first need some more clarity regarding genre/register. Specifically, we would like to employ much more fine-grained groupings, such as romance fiction, detective, science fiction, or medical texts, legal texts, history books. This means we have to determine whether such labels lead to groups of texts which have enough internal consistency, and enough contrast with other groups, so that they can be recognized. In this talk, we present a pilot study concerning romance fiction, a genre for which we most expect to find clear stylistic properties. Using texts from the British National Corpus, we look at various stylistic features, and investigate how these differ between romance fiction and other texts, but also how they vary within the genre, e.g. between authors.

Child abuse: the linguistics of reliability in non-adult witnesses

Martina Nicklaus and Dieter Stein

Universität Düsseldorf

The issue of veracity of verbal evidence, of reports and narratives is at the center of many cases in forensic linguistics. The issue is the more in the center of interest where evidence given verbally is, apart from external and psychological evidence, the only or the nearly only evidence, such as is the case in child abuse cases. The main instrument in evaluating linguistically given evidence in Germany is CBCA (Criteria Based Content Analysis). However, the impression from a recent survey of work on CBCA clearly points to the necessity to include more sharply defined criteria to increase validity. What we wish to suggest in this paper is to not only include more psychological criteria but to also look to include technical linguistic knowledge. We will argue that linguistic knowledge should be included in veracity evaluation in two basic ways:

1. Although psychological evaluation makes use of facts of language use, there can be an issue of the linguistic-theoretical underpinning of psychological categories of analysis. For instance, modern discourse-based linguistic theory of the narrative may provide insights in how stories are generated from memory that a psychological evaluation might profitably make use of. One of the issues here is whether the non-adult narrative accessing of her/his experience is as much as possible free from external factors not related to her/his reporting the experience, such as evaluations of her current situation or possible interests in framing the story, and to what extent the narrative really factually reports what has happened.

2. Intertwined with and likely not always readily separable from 1: the interpretation of linguistic forms themselves, to the extent that a scholarly linguistic analysis, different from a psychological evaluation, has been carried out at all, may profit from more modern and discourse-pragmatics-based technical knowledge in terms of cognitive linguistics. For instance, it is too simple to state that consistency in tenses is an across-the-board diagnostic of consistency of the narrative, and therefore a diagnostic of reliability. Or the functional analysis of word order options, restarts or self-correction strategies, as they are realized in the text may yield interpretable evidence about how cognitive evidence from memory is managed. Similarly, the use of so-called "particles" ("well", "now" in English) and what they tell us about how speakers go about the underlying knowledge structure needs to be exploited for forensic purposes in a linguistically informed way. Modern studies of conversational phenomena with adults from the sociolinguistic point of view lead us to question assumptions about the validity of judgments about the reliability of repeats of stories or the role of the interviewer in re- or co-creating narratives, in all age groups.

For all these linguistic parameters, replete with interpretable information they may be in the case of adults, care must be taken of how the developmental stage of a child or adolescent allows a direct methodological "translation" from what we know about adult handling of discourse markers. The paper discusses some of the linguistic parameters used in CBCA on the basis of ten concrete abuse cases. In all cases the veracity of the victim's statements was not confirmed by the psychological expert. The available data consists of free, experience-based narrative parts as well as offence-related parts. The transcripts therefore allow to compare individual linguistic baselines with the corresponding individual linguistic lying behaviour.

The Polish Corpus of Suicide Notes: research results and their potential usage for forensic linguistics purposes

Monika Zaśko-ZIelińska

University of Wroclaw

Linguistic text analysis for forensic purposes requires reference to language data which is collected within specific speech genres. At the present time, this material is stored and managed by reason of corpus linguistics methods. One-genre corpora are the resources which enable us to reveal the state of language awareness for different groups of users when their social profile completes the language data. The description of language practice within genre studies does not aim to constitute communication models, but rather to discover their diversity (e.g. theory of discursive moves) and pragmatic determinants. Therefore, this approach in contemporary research causes that findings of genre studies or stylistics are a valuable contribution to forensic linguistics. They help to follow external data on the basis of text. This data (both social and cultural) has impact on language behavior (anthropological linguistics) and provides information on the sender, which we can compare with corpus resources.

This presentation will discuss suicide notes in the context of their analysis within forensic linguistics. The research is based solely on suicide notes which were followed by the suicide's death. Usually, there are two research questions connected with suicide notes in the viewpoint of forensic linguistics:

1. What information about the sender is provided by the suicide note analysis when we use corpus data?
2. What features can be considered a trace of inauthenticity of the suicide note?

The author would like to discuss these questions in context of Polish Corpus of Suicide Notes (PCSN) and the guidelines which are usually the base of giving opinions in Poland by experts in writing analysis, but not professional linguists.

The methodological approach to building one genre corpus for forensic linguistics has to refer to corpus linguistic methods, genre studies and pragmatics. A congeneric specialised corpus (PCSN) contains 614 genuine suicide notes written in 2008-2010, and obtained from prosecutor's offices from all over Poland. The legal system does not facilitate the process of including new documents to the corpus. The comparative material for the corpus consisted of texts collected in two sub-corpora with 79 simulated notes and 117 experimentally collected forged notes. The biggest collection of texts of this genre allowed quantitative and qualitative studies to be carried out. The quantitative analysis has to be supplemented by the qualitative viewpoint, since genuine and forged texts are created in different situations. Hence, various pragmatic contexts are considered in the final conclusions. As an example: the number of personal pronouns in the genuine suicide notes is linked with the sender focusing on himself due to his mental state and in the forged suicide notes it results from the fact that the sender does not know the recipient. The usage of quantitative method in this project was possible, as the corpus holds a set of annotations: structure layout (text structure, physical text division into text blocks and lines; text block layout, text alignment, indention; text formatting; text omissions), linguistic information (segmentation into tokens and language expressions of various complexity; semantic and pragmatic classification of text elements, e.g., salutation inside text, signature, envelope date expressed in different ways; proper names), meta-data

(information on the author; physical description - paper format, type, etc.; linguistic description - type of text and act of speech) (Marcińczuk, Zaśko-ZIelińska, Piasecki, 2011: 1-5).

The primary results of this project, which was aimed to describe a suicide note as a speech genre, provides essential pragmatic, structural and thematic data for interpretation of quantitative data. The characteristics of texts which significantly differentiate authentic and forged notes from the main corpus and sub-corpora can be used to recognize authentic suicide notes and forged texts. During the analysis of the text authenticity several features were considered: graphical features of writing (graphical errors, author's corrections, deletions, readability, highlights, figures, discontinuity of writing); spelling accuracy, punctuation; text structure, length of the text; the traces of communicative context (e.g. information about the sender-recipient relationship, relation to time and space), the illocutionary aim; thematic content; linguistic picture of suicide and stylistic feature of the text (e.g., spoken and written elements in the text).

The usage of the corpus for forensic purposes is determined by the manner in which the experts carry out the procedures in a given country. In Poland, the language expertise of the text is firstly outsourced to experts in the field of writing expertise. Linguists are on the expert lists of a few courts, although occasionally non-list linguists are also asked to provide the expertise of the text due to their special competence. As the corpus is not widely available and can be used only after receiving permission, we are aware that it is used only by researchers. It is not used to prepare expertise, although it is commonly accepted that corpus sources are the best material for this purpose. However, the set of guidelines applicable to experts in the field of writing examination does not suggest the use of corpus. These guidelines are a set of rules prepared on the basis of structuralist linguistics and they require the use of neither quantitative methods nor corpus data. In my opinion, this approach should change after lectures and workshops in the Central Laboratory of Criminalistics during which I indicated the disadvantages of the existing guidelines and referred to the claim stated in Kredens's paper (2016) in *Iustitia*, a journal for Polish judges, that it is necessary to develop standards of the linguistic analysis in Poland and to build corpora from the existing police text resources (e.g. files of anonymous letters). On the basis of empirical experience with the PCSN, the author presumes that the tools developed for this project can be applicable to other forensic text corpora.