

Harnessing court data using NLP and spoken language technology

Constantin Orăsan and Hadeel Saadany

C.Orasan@surrey.ac.uk and Hadeel.Saadany@surrey.ac.uk



Centre for Translation Studies
UNIVERSITY OF SURREY

Objective 1: Improve Automatic Transcription of Courtroom Hearings



Amazon Web Service Transcription gives **errors** with legal terminology/jargon, named entities, numbers

We minimize the transcription errors by building a **Custom Language Model (CLM)** trained on legal text and edited transcriptions. Our CLM results correct most errors of AWS generic model and reduce Word Error Rate (WER)

BASE MODEL

Legal jargon "my learned" → with my **lonely** Junior Miss Eleanor Harris. I appear on behalf of **Mrs Agbaje**, the appellate um, and then your friend Mr Timothy Scott, Queen's counsel, and Mr Peter Mitchell appear on behalf of Mr M. Party, the respondent to the appeals. If I may, I will have dropped the Family Law convention of calling the party's husband and wife. The principal issue arising on this appeal is, in what circumstances is it appropriate for the English court to make a further financial order when a foreign court has already divorced the spouses and made a financial order? At first blush, it may seem unusual that courts in two different countries can, at different times make financial orders arising out of the same cause of action, namely, **the party's divorced**. The power in the English court to make a further financial order derives from Part three of the Matrimonial and Family Proceedings Act 1984 and it was intended by **its frame** as the law commissioners to remit financial hardship arising from two distinct circumstances. I suppose I should make it absolutely plain that this is the work of my predecessor. Yes, the Law Commission. I had no hand in part three at all. Yes, we saw the authors that listed in the report. The two circumstances that the law commissioners were addressing were these firstly, when a foreign court had made no financial order at all and secondly, where the foreign Court had made a financial order. But that order was **inadequate** and this appeal is concerned only with the second category. It is a condition precedent to the exercise of powers under Part three that the wife has to show in the circumstances of her case that a serious injustice has arisen. The serious injustice that this wife relies upon is that after a marriage of 32 years which produced five **Children of the founded** and where the assets were around £700,000 the Nigerian court in June of 2005 awarded her a lump sum of £21,000 and a life interest in the house on Tin Can Island in Lagos. This very modest award produced a very significant disparity or discrepancy in the allocation of the assets between the parties

Grammar mistakes "parties divorced" →

Names of people "Mrs Agbaje" →

Legal Terms "its frame" →

Vocabulary "children of the family" →

CLM MODEL Correcting Mistakes

Legal jargon → with my **learned** Junior Miss Eleanor Harris. I appear on behalf of **Mrs Agbaje**. The appellate, um, my learned friend, Mr Timothy Scott, Queen's counsel, and Mr Peter Mitchell appear on behalf of Mr M. Party, the respondent to the appeal. If I may, I will have dropped the Family Law Convention of calling the parties husband and wife. The principal issue arising on this appeal is in what circumstances is it appropriate for the English court to make a further financial order when a foreign court has already divorced the spouses and made a financial order? At first blush, it may seem unusual that courts in two different countries can at different times make financial orders arising out of the same cause of action, namely, **the parties divorced**. The power in the English court to make a further financial order derives from Part three of the Matrimonial and Family Proceedings Act 1984 and it was intended by **its framers** as the law commissioners, to remit financial hardship arising from two distinct circumstances. I suppose I should make it absolutely plain that this is the work of my predecessor. Yes, the Law Commission. I had no hand in part three at all. Yes, we saw the authors listed in the report. The two circumstances that the law commissioners were addressing were these firstly, when a foreign court had made no financial order at all and secondly, where the foreign Court had made a financial order. But that order was **inadequate** and this appeal is concerned only with the second category. It is a condition precedent to the exercise of powers under Part three that the wife has to show in the circumstances of her case that a serious injustice has arisen. The serious injustice that this wife relies upon is that after a marriage of 32 years which produced five **Children of the family** and where the assets were around £700,000 the Nigerian court in June of 2005 awarded her a lump sum of £21,000 and a life interest in the house on Tin Can Island in Lagos. This very modest award produced a very significant disparity or discrepancy in the allocation of the assets between the parties"

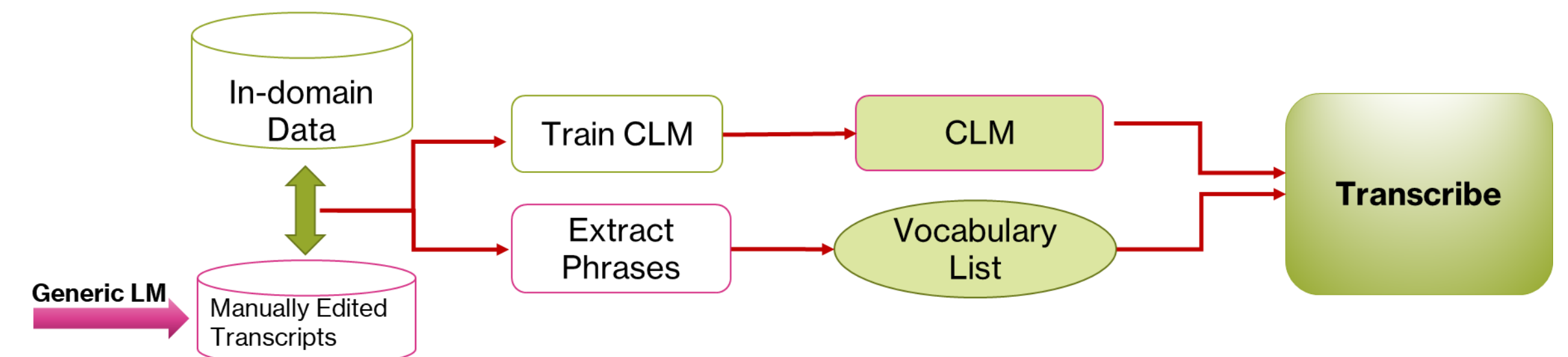
Grammar →

Names →

Legal terms →

Vocabulary →

Our pipeline for Improving Transcription Quality



1. Employ **NLP tools** to **extract a vocabulary** list with most common bigrams in the edited transcripts.
2. Employ Blackstone library to create a **list of legal terms**.
3. Scraped a large dataset of judgements (approx. 25k documents) from the UK Tribunal Decisions (<https://www.judiciary.uk>) to use for training the CLM.
4. Use the extracted **list of legal terms** and most common entities to **build CLM** using different training datasets (supreme court judgements and edited transcripts).

The objective and design expectations

- Transcripts (very long 10+ hrs.) can be used as a tool for students/academics/lawyers to learn about a particular law.
- **We semantically link time spans of transcription to sections/paragraphs in the judgement.**
- This can help understand how the judgment is reached.



Progress so far

- **GPT3 semantic search** spots the gold-standard timestamp in a list of top 20 most relevant.
- Manual analysis of results shows that the 20 extracted timestamps are relevant to the judgement and demonstrates good linking.

Our Method

We treat the judgement sections as a **Query** and the **Transcriptions** as potential answers. We rank spans using the **cosine similarity between the GPT3 sentence embeddings** of the judgement and transcription spans

Results

Judgement:
The complaint, very moderately advanced by Mr. Geekie QC, is that a "starting point" of undesirability, placing the burden upon the person wishing to cross-examine a child to show some "particular justification" for doing so, gives insufficient weight to the **Convention rights** of all concerned. All the parties in care proceedings are entitled to a fair hearing in the determination of their civil rights and obligations the parents who stand to lose their children if allegations of abuse are made out, the children who stand to lose their parents if allegations of abuse are made out, but also stand to suffer abuse or further abuse if they are left at home because those allegations cannot be proved. And it is not only their **article 6 rights** which are in play. The **civil rights** in issue are also **Convention rights** in themselves the right to respect for the family lives of the parents and their children but also the right to respect for the private lives of the children, which include their rights to be protected from attacks upon their physical and psychological integrity. Even a "**stranger**" child, whose future is not in issue in the proceedings

Timestamps:
00:01:31.440,00:03:19.240,spk_2 (The Judge),

Legal Point → Looking at what we say should be the starting point. Can I just consider, the issue of the welfare of the child and in the context and this feeds directly into my paragraph 47 where I say the welfare, of the child is irrelevant but not paramount consideration it **links into Mr Geekie's point** about whether a **non subject** child, so a child who is simply a witness within proceedings has Article 8 rights engaged, so not as confident as **Mr Geekie** on that. But what I would say is that if the court is considering a child as a witness and in no other capacity, we would say that that **child's welfare is relevant** and that probably actually answers the Article 8 point if that's engaged. But **when we say welfare of a child is relevant, we mean for any child witness, whether a party or no**. The way in which we had formulated the factors to be put into the balancing exercise came as is clear from paragraph 45 of the Medway case from the judgement of the Lady Justice Smith, where she went on to look at the various factors, she's put them in a much more generalized form.

Correct link →

Correct link →

Correct link →

Next steps

- Create a large gold standard for training and testing a linking model.
- Experiment with more NLP methods to improve linking accuracy
- Produce a smaller list of timestamps linked to judgement
- Gain a better understanding of how the users would use such a system
- Create a pipeline for cleaning, segmenting and linking the Courtroom hearings and the Transcriptions produced by our CLM model

