# Language Intelligence - Session 3

#### **Time:** 21 September 2023 - 15:45 to 18:05

#### **Chair:** Arthur Wetzel, Member of the Board of Advisors, LT-Innovate

## **Talks**

### Banishing Terminology Chaos: The Terminology Database to the Rescue!

You have a terminology list for your department but it does not comply with the content of other departments? So your terminology is inconsistent in front of the customer?

Let berns language consulting and a customer text representative show you how they tackled the topic together!

In this presentation we show how we unified several differently formatted terminology lists into one data set and how to clean up the data set in a next step. Can we manage to lay a foundation for high-quality terminology management applicable to the entire brand?

The first step was to collect and analyze the inventory data and to examine the existing terminology according to defined criteria. After this, automatic statistical term extraction was used to identify or verify potential term candidates. These candidates were then adjusted and standardized, considering the existing metadata. Manual evaluation was still necessary to ensure the accuracy and term worthiness by taking into account the importance of standardization in creating a consistent and usable database. This included ensuring that the terms were spelled consistently and that synonyms were linked correctly.

One of the challenges faced during the project was the standardization of different forms of original data. Some of the same concepts and terms were repeated in different list, while other concepts contradicted each other. Also the possibility of including more departments should be retained.

In conclusion, the creation of a terminology database for Quickterm was a complex process that required careful computer-aided terminology work. Despite the challenges faced, the team successfully created a reliable and useful database that will benefit the company in various ways, e.g. by improving recognition of vehicle functions and their understandability throughout different customer touchpoints.

| Sophia AckermannComputational Linguist, berns language consulting | Ira RotzlerHead of department term, text, translation, tooling, Porsche AG |
| --- | --- |

### Leveraging Technology to Build a Robust and Scalable Ecosystem for Global Content Operations

Language technology is shaking the practices and standards that have been driving global content operations for the past decades. More than ever organizations need an ecosystem to enable their teams to leverage agile workflows while integrating existing and new technologies in a cost and time effective way. Yet moving from systems and tools to a unified ecosystem covering content supply chains is a journey involving a number of stakeholders in various departments, functions and disciplines. During this session leaders in global content operations will share experiences and best practices on:

* How to integrate and manage technology components to ensure engagement, adoption and collaboration across content supply chains
* How to create a framework and a roadmap to balance language technology with human in the loop for multilingual content
* How to implement the conversion of content assets (text, video, audio) into language data feeding, training and improving AI
* How to create, capture, measure and increase value with language technology

| Bruno HerrmannMember of the Board of Advisors, LT-Innovate[LinkedIn profile](https://www.linkedin.com/in/brunoherrmann/) | Esther CurielGlobalization Manager, Zoetes[LinkedIn profile](https://www.linkedin.com/in/esther-curiel/) | Laura RossiProgram Manager Machine Translations, Medtronic[LinkedIn profile](https://www.linkedin.com/in/laura-rossi-55a2913/) |
| --- | --- | --- |

### Language Intelligence to Guide Medical Research

Because of methodological shortcomings, many clinical studies are of limited scientific value. Suitable methods guidance for health researchers is available but, due to highly heterogeneous terminology and lack of a specialized database, hard to find even for experienced researchers. To make published guidance more accessible and more widely used, an international group of researchers created the Library of Guidance for Health Scientists (LIGHTS) as a new searchable public database for methods guidance: https://lights.science/

We will briefly describe how we set up LIGHTS by tailoring Karakun's HIBU platform to the medical research domain and the given use case. Several search features allow users to efficiently find methods guidance relevant for their clinical studies, most notably: a set of targeted domain-specific filters which use a growing controlled vocabulary which harmonizes the terminology across the included papers. The solution also benefits the LIGHTS team through a lightweight toolset for adding new guidance articles to the search index and converting and validating metadata.

In a final section we will discuss how automation can support the LIGHTS team in the selection and indexing process to make this initiative viable in the long run. Currently, methodology experts carefully screen candidate papers following a set of transparent criteria and assign index terms manually. Because of the huge amounts of candidate papers, we currently explore ML/LT approaches to ranking and pre-filtering the candidate lists and suggest index terms to increase the speed and efficiency of the screening and indexing process.

| Holger KeibelHead of Language Analytics & Search Solutions, Karakune[LinkedIn profile](https://www.linkedin.com/in/dr-holger-keibel-bba57b1a/) | Stefan SchandelmaierSenior Researcher, Universitätsspital Basel[LinkedIn profile](https://www.linkedin.com/in/stefan-schandelmaier-4371ab195/) |
| --- | --- |

### Automated ESG Knowledge Extraction from Dow Jones - Factiva Global News

Presentation covers the technology used to automate Knowledge extraction from Text. This novel technology blends neural language models, semantic tech, rule systems, linguistic theory to achieve reliable performance. Specifically, the discussion will focus on the integration with Dow Jones/Factiva, involving the extraction of facts buried in news articles, newsletters, reports, etc. about the subject area of ESG Environmental..) through the application of ESG taxonomy and ontology. Extracted facts are output as RDF triples and ingested into a Semantic Knowledge Graph stored a triple store which supports BI & reporting over these facts through standard graph (SPARQL) queries.

| Prasad YalamanchiCEO, Lead Semantics[LinkedIn profile](https://www.linkedin.com/in/leadsemantics/) |
| --- |

### Plain X: A 4-in-1 Tool for Content Adaptation

plain X is a new, innovative offering content adaptation platform, introduced in late 2022. The software is based on several years of research in HLT (Human Language Technologies) by the two partners Priberam (Lisbon) and Deutsche Welle, Germany's international broadcaster, which publishes articles, audio and video in 30 languages.

The software is a 4-in-1 solution enabling editors to create audio or video transcriptions (speech-to-text or upload of existing transcripts), translations, subtitles and (synthetic) voice-overs. User can define which engine should be used for specific language pairs to achieve best results. Prior to the now public launch DW editors extensively tested the tool for the first deployment. By now, hundreds of editors use the platform daily.

Three main aspects make plain X noteworthy: Firstly, the platform is engine-agnostic. Users can test and choose which engine best suits their specific language pair. Connecting to the newest and best AI engine for language work is crucial, specifically today, with the swift development of generative AI software. As an example: Users can use language engines from Google, Facebook or Azure, but for example, they can connect to Open Ai Whisper for fast transcriptions even for long audios or videos. Secondly, plain X provides optimized workflows for content adaptation work - from upload and download from content repositories to optional keyboard shortcuts, which can be used to complete recurring tasks quickly. Thirdly, plain X emphasizes the "human in the loop". For example, each content item can be assigned to a reviewer as an option towards higher quality.

The talk will provide a behind-the-scenes insight into the development of this tool so far and a walkthrough of key features. Given the ever-increasing

| Mirko LorenzInnovation Manager, Deutsche Welle[LinkedIn profile](https://www.linkedin.com/in/mirko-lorenz-b2861299/) |
| --- |