# BABELSTREET

# Shining a Light on Consumer Feedback

Spun out of the MIT Media Lab in 2010, Luminoso quickly drew the attention of major consumer brands. Its flagship product, Luminoso Analytics, digests large volumes of text-based customer feedback, such as online reviews, surveys, and customer service interactions. Instead of just finding key words, Luminoso identifies key concepts, ideas, thoughts, and sentiments that drive consumer choices. Luminoso empowers its clients to understand, measure, and act on consumer feedback across any number of channels.

## The Challenge

Every time Luminoso added a new human language to their portfolio, finding a reliable linguistic analyzer in that language was the first step.

For Luminoso's algorithms to tease out ideas, thoughts, and sentiment from unstructured text, the raw text needed to be enriched and tagged with exacting accuracy. This process involves stripping out extraneous aspects of the text to reveal finely tuned messages within.

Careful analysis of the words using Natural Language Processing (NLP) involves:

- Finding word boundaries, also known as tokenization
- Tagging parts of speech
- Finding the lemma of each word

"Whenever someone asks us, does your system work in language x? The answer is yes, it does

| Tokenization | Part of Speech Tagging | Lemmatization |
| --- | --- | --- |
| 貴社 の 記者 が 汽車 で 帰社 した。 | Pronoun Verb Adjective Noun | am / are / is → to be |

to the extent that you can put your text into our system and we'll see how it does," Lance Nathan, former Senior Linguistics Developer at Luminoso said. "The question really isn't does it work in language x, but how well does it work in language x."

Getting the quality results expected by Luminoso users is easier when the linguistic enrichment is of higher quality.

Luminoso started by using open source solutions for linguistic analysis, but in some cases, the results were simply not accurate enough.

In early experiments, Luminoso used stemming, which is one of the most basic tools to remove suffixes from words. However the error rate was unacceptable. More sophisticated tools made more sophisticated mistakes.

## The Solution

Luminoso implemented the base linguistic capabilities of Babel Street Text Analytics for features including part-of-speech tagging, language identification, and lemmatization.

In particular, Text Analytics excelled at finding lexicalized phrases. For example, it knows that "nouveaux riches" is a unit that should be singularized to "nouveau riche."

Text Analytics' ability to identify ambiguous parts of speech with greater accuracy meant that using Text Analytics increased Luminoso's robustness in the face of spelling errors such as missing accents in French and Spanish.

Luminoso continued to run a battery of tests for every new language to determine whether to stay with Text Analytics. This was done by comparing a language corpus that was already tagged with parts-of-speech with the results obtained by

## Luminoso in Action

A multinational consumer goods company was designing a new variant of its most popular men's personal care product. The question was, what features actually matter most to their consumers?

Thousands of testers filled out a survey about the new product variant, which asked them to rate attributes on a scale of 1 to 5, and to leave free-form comments. These comments were more text than could be accurately processed manually.

The company chose Luminoso to analyze the free-form comments, uncovering distinct product feature discussions and measuring each one for positive and negative feedback. Luminoso also looked at reviewer rating tiers separately to see what company and product attributes drove both high and low review scores.

Surprisingly, the color of the item was mentioned in comments as a significant factor in the product experience. This key feature would not have been discovered through guided queries alone.

running it through Text Analytics. Text Analytics consistently achieved a 97–98% match, so there was never a need to replace it.

## The Impact

Text Analytics' accuracy across multiple languages and the ease of a unified API for all languages are very valuable to Luminoso.

"I needed to be convinced we are doing at least an adequate job in any new language," Nathan said. "Our English is the gold standard. The new language has to be at least 'pretty good' to 'very good' or at a level approaching English."

"Using Text Analytics gets [the results] to a point that I wouldn't be embarrassed to show them to a native speaker of the language," Nathan said.

The feedback from Luminoso customers was loud and clear. In less than four years since implementing Text Analytics, Luminoso went from startup to a portfolio of 60 to 70 customers, including Fortune 1000 companies.

Text Analytics excels at finding lexicalized phrases. For example, it knows that "nouveaux riches" is a unit that should be singularized to "nouveau riche."

Babel Street is the trusted technology partner for the world's most advanced identity intelligence and risk operations. The Babel Street Insights platform delivers advanced AI and data analytics solutions to close the Risk–Confidence Gap.

Babel Street provides unmatched, analysis–ready data regardless of language, proactive risk identification, 360-degree insights, high–speed automation, and seamless integration into existing systems. We empower government and commercial organizations to transform high–stakes identity and risk operations into a strategic advantage.

Learn more at **babelstreet.com.**

BABELSTREET