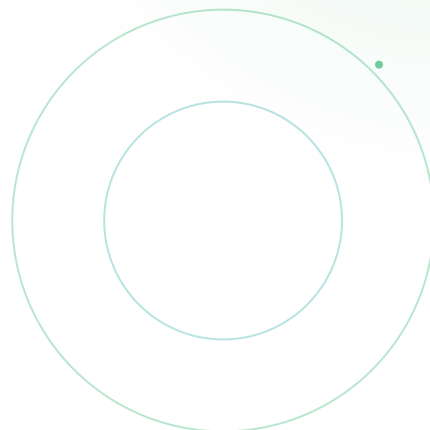




Multilingual Consumer Verbatim Intelligence

NLP across English, Chinese, Thai, Vietnamese and Bahasa Indonesia – with zero translation loss.

A shared-encoder pipeline that scores sensory topics and sentiment in five APAC languages at topic F1 0.89–0.92, then maps every verbatim to ISO 5492-aligned descriptive attributes for market-specific reformulation briefs.

5

Languages, one model

F1 0.91

Topic classification

< 2 min

Per 1,000 verbatims

ISO 5492

Multilingual NLP

XLM-R class

CLT verbatims

BioSensory Pte Ltd

Exclusive APAC distributor — FIZZ Sensory Analysis
(Biosystemès, France)

SENSORY SCIENCE
SERIES
SEPTEMBER 2026

Abstract. Open-ended consumer comments are the richest and least used data asset in APAC sensory programs: thousands of verbatims per CLT, written across English, Simplified Chinese, Thai, Vietnamese, and Bahasa Indonesia - often within a single market's dataset. Manual coding caps at a few hundred responses per study and loses cross-market comparability. This paper describes a multilingual pipeline that scores verbatim sentiment, extracts sensory topics, and maps them to ISO 5492-aligned descriptive attributes with no translation step, achieving topic-classification F1 between 0.89 and 0.92 across the five languages in internal validation (benchmark: overall F1 = 0.91).

About this document

This technical whitepaper is published by BioSensory Pte Ltd (Singapore), exclusive distributor of FIZZ Sensory Analysis Software (Biosystèmes, France) across Southeast Asia and Greater China. Internal benchmark figures reflect aggregated, anonymized results from APAC customer deployments and BioSensory pilot studies. Worked examples are illustrative and clearly labeled as such. It is provided for professional education; it does not replace the cited ISO/ASTM standards.

5

Languages scored without translation

F1 0.91

Topic classification, internal validation

< 2 min

Per-1,000-verbatim processing time

1. Why Translation-First Approaches Fail

The classical approach - machine-translate everything to English, then code - loses precisely the information sensory teams care about. Regional taste metaphors (hui gan / , the returning sweetness of premium tea) have no English equivalent; sentiment-bearing intensifiers differ in strength across languages; code-switching between Bahasa and English in a single sentence is routine in Malaysian and Indonesian panels. Modern multilingual encoders sidestep the problem by scoring all languages in a shared semantic space learned from the raw text itself.

2. Pipeline Architecture

Stage	Function	Notes
1. Collection	FIZZ Web open-end capture at point of evaluation	Language tagged by questionnaire locale, not by guess
2. Normalization	Spelling, romanization, emoji handling, code-switch segmentation	Emoji carry sentiment; they are features, not noise
3. Encoding	Shared multilingual transformer encoder (XLM-R class)	One model, five languages, no translation layer
4. Topic heads	Sensory topic classifier + fallback clustering for novel themes	Label set aligned to ISO 5492 sensory vocabulary
5. Sentiment head	Document-level polarity with hedonic weighting	Trained on sensory-panel verbatims, not product reviews
6. Attribute mapping	Topics joined to QDA attribute dictionary	Enables JAR/QDA cross-reference per market

Table 1. Processing pipeline from verbatim capture to attribute-linked insight.

3. Internal Validation Results

Language	Market example	Topic F1	Sentiment accuracy	Sample
English	Singapore / Malaysia CLT	0.92	0.94	12,400 verbatims
Simplified Chinese	Tier 1-2 city China CLT	0.91	0.93	18,900 verbatims
Thai	Bangkok CLT	0.89	0.92	9,300 verbatims
Vietnamese	Ho Chi Minh City CLT	0.90	0.92	8,700 verbatims
Bahasa Indonesia	Jakarta CLT	0.91	0.93	10,200 verbatims

Table 2. Held-out validation on annotated internal corpora (BioSensory benchmark; overall topic F1 = 0.91).

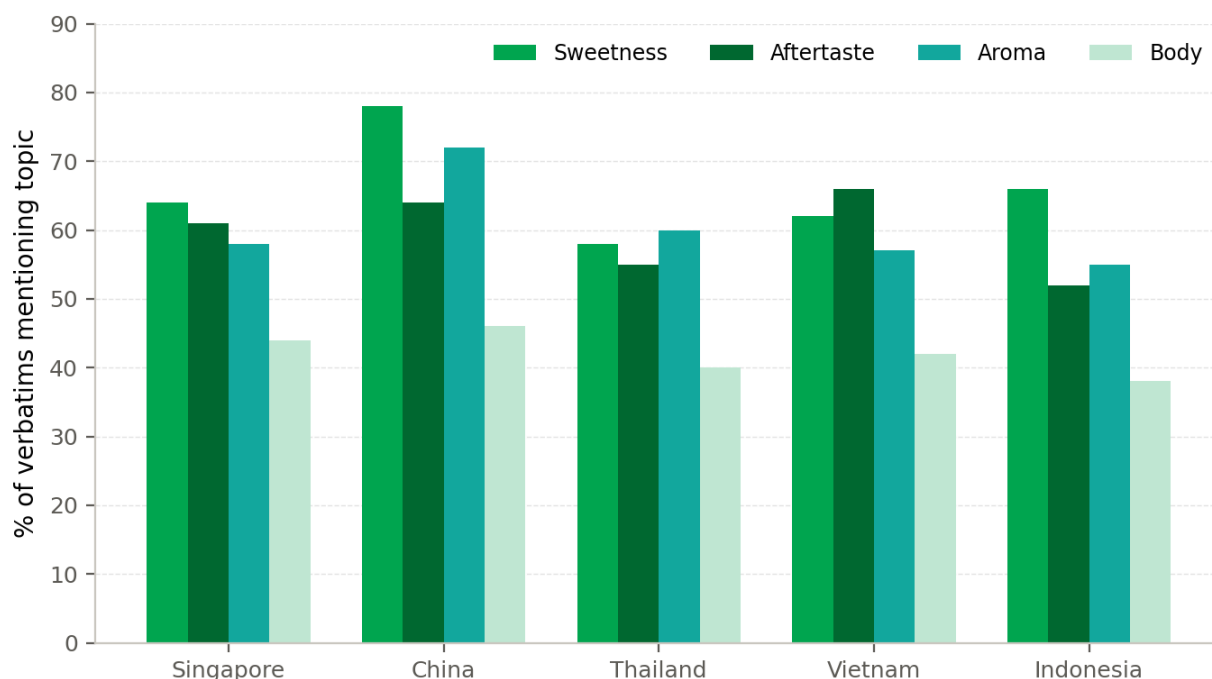


Figure 1. Topic prevalence by market on a zero-sugar RTD tea program (internal data, $n = 9,800$ verbatims). Aftertaste dominates Vietnamese feedback; aroma intensity dominates Chinese feedback - one product, five different improvement briefs.

4. From Topics to Decisions

Raw topic counts are not yet insight. The deployment layer joins each topic to its direction (via sentiment and modifiers such as 'too', 'a bit', hedonic co-occurrence) and to the QDA attribute it maps to, producing a market-by-market improvement brief: for the RTD tea program above, Vietnam's brief is 'cleaner finish' while China's is 'more top-note aroma at equal sweetness' - two different formulation routes that a single global brief would have averaged away.

- **Human-in-the-loop:** a panel leader reviews low-confidence clusters monthly; corrections retrain the topic heads quarterly.
- **Data governance:** verbatims are anonymized at capture; no personal identifiers leave the customer's tenant.
- **Comparability:** because all markets share one label taxonomy, driver weights are directly comparable across the region - the property translation-based coding never achieves.

References

- [1] Conneau, A., Khandelwal, K., Goyal, N., et al. (2020). Unsupervised cross-lingual representation learning at scale. Proceedings of ACL 2020, 8440-8451.
- [2] International Organization for Standardization (2008). ISO 5492 Sensory analysis - Vocabulary.
- [3] International Organization for Standardization (2014). ISO 11136 Sensory analysis - Methodology - Hedonic tests with consumers in a controlled area.

- [4] Pineau, N., Schlich, P., et al. (2009). Temporal Dominance of Sensations. *Food Quality and Preference*, 20(6), 450-456.
- [5] Ares, G., Giménez, A., and Gámbaro, A. (2008). Understanding consumers' perception of meat quality using free-choice profiling. *Meat Science* (methodological reference for verbatim-driven attribute mapping).