Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

What Counts as Zero-Shot? A Systematic Review of Target-Language Data Use in Structured ABSA Transfer

Botir Elov

elov@navoiy-uni.uz

Alisher Navoʻi Tashkent State University of Uzbek Language and Literature

Ruhillo Alaev

National University of Uzbekistan named after Mirzo Ulugbek

Zilola Xusainova

Alisher Navoʻi Tashkent State University of Uzbek Language and Literature

Research Article

Keywords: Aspect-based sentiment analysis, Cross-lingual transfer, Zero-shot evaluation, Target-language data, Structured prediction, Systematic review

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11746069/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

What Counts as Zero-Shot? A Systematic Review of Target-Language Data Use in Structured ABSA Transfer

Botir Elov[a,*], **Ruhillo Alaev**[b], **Zilola Xusainova**[a]

[a] Department of Computer Linguistics and Digital Technology, Alisher Navoʻi Tashkent State University of Uzbek Language and Literature, Tashkent, Uzbekistan

[b] Department of Information Security, National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan

*** Corresponding author:** Botir Elov, elov@navoiy-uni.uz.
Correspondence address: 103 Yusuf Xos Khojib Street, Yakkasaroy District, Tashkent 100070, Uzbekistan. ORCID: 0000-0001-5032-6648

Abstract

In structured cross-lingual aspect-based sentiment analysis (ABSA), the label zero-shot can leave the scope of target-language data access unspecified. We examined two restrictions, one on human-annotated target-language task examples and one on any task-specific target-language material beyond the current evaluation input. A structured search of the Association for Computational Linguistics (ACL) Anthology on 15 November 2026 and supplementary discovery yielded 114 unique bibliographic records. Of 38 prioritized reports, 20 had provisional eligibility support, nine were contextual and nine remained unresolved. Independent screening and reconciliation are pending. We traced resource provenance, supervision and stage of access in 16 selected configurations from nine reports. The evidence supported absence on both checks in four configurations. For six, it supported absence of human-annotated target examples and documented other target-language material. For four, it documented target material and left use of human annotations unresolved. Both checks were unresolved in two. Five within-study contrasts documented score differences under aligned evaluations. Resource access could change alongside training volume or inference operations, and different scoring protocols limited comparison across studies. For each experimental arm, we propose specifying the resources withheld and the stages covered by the zero-shot claim, and linking that claim to the data, selection procedure and scorer.

Keywords Aspect-based sentiment analysis · Cross-lingual transfer · Zero-shot evaluation · Target-language data · Structured prediction · Systematic review

1 Introduction

1.1 Motivation and problem statement

Aspect-based sentiment analysis (ABSA) identifies opinions about particular aspects of a text. In structured ABSA, the evaluation unit is a linked prediction. An aspect–polarity pair links an aspect term to its sentiment. Other formulations add an aspect category (Wu et al. 2025b) or include both a category and an opinion expression (Cai et al. 2021). When annotated examples are scarce in a target language, cross-lingual transfer can use supervision from a source language to learn these outputs. The resulting evaluation measures how well the procedure recovers the linked elements in the target language under a particular data-access condition (Šmíd and Král 2025).

The term zero-shot can qualify different parts of that condition. One interpretation rules out human-annotated target-language task examples. A stricter interpretation rules out task-specific target-language material used to train, adapt, select, or condition the system beyond its current evaluation input. For example, texts translated from the source corpus may appear in the target language and carry labels projected from the source. Independently collected target-language texts may also shape a model through adaptation on unlabeled data. These procedures can both operate without human-annotated target-task examples while exposing the model to different target-language evidence.

Zhang et al. (2021) use zero-shot for their baseline in which a model fine-tuned on labeled source-language data predicts directly on target-language inputs. The same study evaluates transfer through label projection, aspect code-switching (ACS), and distillation on unlabeled target-language texts. Wu et al. (2025b) report zero-shot cross-lingual transfer after English fine-tuning and separately use zero-shot to describe prompting without demonstrations. Fehle et al. (2026) use zero-resource for several strategies, including training on non-target languages, training on code-switched examples, and training on translations into the target language. The scope of each term is therefore part of the evidence to extract.

Exposure during model development also differs from ordinary test-time prediction and scoring. Target-language validation can inform model selection. Access to a collection of target test instances for adaptation creates a different procedure from predicting on each held-out input with a fixed system. General multilingual pretraining introduces a further source of exposure beyond the task-specific procedure. For languages with scarce ABSA annotations, these details determine what a reported score can tell readers about a model's performance under limited target resources.

1.2 Scope and contribution

Hua et al. (2024) synthesize ABSA domains, task combinations, and methodological trends. Santin and Rezende (2026) focus on adaptation across domains. Šmíd and Král (2025) survey cross-lingual ABSA tasks, datasets, and transfer approaches, while Fatemian et al. (2026) review ABSA transfer-learning techniques, including uses of translated or unlabeled target data. We build on this literature by tracing what each reported experiment reveals about target-language data access.

We focus on structured ABSA experiments with an identifiable source-to-target transfer setting and an evaluation of linked outputs in the target language. The unit of analysis is the reported experimental configuration. We retain each author's label and record the provenance and supervision of target-language resources, when and where they were accessed, and the source passage or artifact that supports each decision. The review assesses separately whether human-annotated target-task examples were used and whether any task-specific target-language material informed the procedure. Settings with insufficient documentation remain unresolved.

The synthesis connects each zero-shot label to the documented resources and their role in the experiment. We examine performance differences in within-study contrasts where the reported task, test split, metric, and other experimental conditions support comparison. The synthesis also identifies the information needed to interpret or verify a zero-shot claim, including details of translation and label projection, target-language validation, prompt construction, and access to evaluation data.

1.3 Research questions

RQ1. How do studies define and label zero-shot transfer in structured ABSA?

RQ2. Which target-language resources are used at each stage, and what are their provenance and supervision?

RQ3. What performance differences are reported when target-language data use varies across comparable experimental arms?

RQ4. What evidence is available to verify these labels, and which reporting gaps leave the experimental setting unresolved?

2 Background and Related Reviews

2.1 Structured ABSA tasks and transfer direction

A structured ABSA task specifies which sentiment elements must be recovered jointly from a text. We use a for the aspect term, c for the aspect category, o for the opinion expression, and s for sentiment polarity. An aspect term is a textual mention of what is being evaluated. The category is a label from the benchmark's predefined inventory. The opinion expression is the textual evidence of the evaluation, while polarity gives its direction. The SemEval-2016 shared task includes an opinion tuple comprising a category, an opinion target expression and a polarity. In that formulation, the opinion target expression names the object of the evaluation (Pontiki et al. 2016).

Joint aspect-term and polarity prediction returns pairs (a, s). Zhang et al. (2021) evaluate a pair by checking its aspect boundary and polarity together. Aspect sentiment triplet extraction (ASTE) links an aspect term, an opinion expression and their polarity (a, o, s), the three elements specified by Peng et al. (2020) and targeted in multilingual transfer by Sheng et al. (2025) and Dewangan and Maurya (2026b). Target-aspect-sentiment detection (TASD) pairs the term with a category and polarity (a, c, s) (Wan et al. 2020). Wu et al. (2025b) benchmark both term–polarity pairs, called unified ABSA (UABSA), and TASD triplets. Aspect-category-opinion-sentiment (ACOS) extraction links all four elements (a, c, o, s). Aspect sentiment quad prediction (ASQP) evaluates the linked quadruple. The ACOS annotation rules allow a designated null value for an implicit aspect term or opinion expression (Cai et al. 2021).

The quadruple (a, c, o, s) illustrates which fields each formulation preserves. Its projections are (a, s) for joint aspect–polarity prediction, (a, o, s) for ASTE and (a, c, s) for TASD. Each benchmark defines its category inventory, span conventions and rules for implicit elements.

Joint scores also depend on how a predicted structure is matched to a reference. M-ABSA micro-averages over complete matches of the task's pairs or triplets (Wu et al. 2025b). MultiAspectEmo reports exact and partial matching for end-to-end ABSA (Szołomicka and Kocoń 2022). The shared task on dimensional ABSA represents sentiment by valence and arousal and defines a continuous F1 score for its structured predictions (Yu et al. 2026). For each experiment, the output schema, benchmark version, test split and matching rule identify what the score measures.

We describe transfer at the level of a donor language or a set of donor languages and a distinct target-language evaluation set. A multilingual donor pool can support source-only supervision when it excludes target-language labeled task data, as in the non-target training settings of Fehle et al. (2026). Supervised target-language fine-tuning defines a different experimental condition. For example, Wang et al. (2026) examine bilingual pretraining followed by language-specific supervised fine-tuning. Both donor and target domains can also vary. We therefore record language direction, domain and supervision separately.

The origin of the target test text is an additional property of the evaluation. M-ABSA uses machine-translated texts with human quality checks on a sample (Wu et al. 2025b). LASQ contains Uzbek and Uyghur app reviews collected separately in their original languages (Yusufu et al. 2026). Parallel translations can preserve content across languages. Independently collected texts capture language-specific usage and may differ in topic or domain. The benchmark's provenance and the resources available to a specific experimental arm are recorded separately.

2.2 Operational definitions of zero-shot transfer

The review retains each designation, including zero-shot and zero-resource, together with its meaning in the source. We then apply two checks to the reported configuration. The first tests for the absence of human-annotated target-language ABSA examples in training, adaptation, model or prompt selection, and inference. The second tests for the absence of task-specific target-language material in those stages beyond the current input being evaluated. This second check also covers translations of donor text into the target language, target-language instructions, independent unlabeled target text and machine-produced annotations. The checks are coded separately. An experiment that satisfies the second check ordinarily also satisfies the first.

For example, Zhang et al. (2021, §§2.5 and 3.1) use English validation across their configurations. In their distillation arm, the teacher labels target-language training sentences whose benchmark labels were withheld. This procedure uses task-specific target text and teacher labels, with human gold annotations restricted to the source during development. LACA also draws on unlabeled target text and then trains with pseudo-labeled data (Šmíd et al. 2025b). Dewangan and Maurya (2026b) use code-switched prompts, pseudo-triplets and contrastive distillation for ASTE. The provenance and supervision fields distinguish these resources from human-verified target-language task annotations. When humans review translated data, the coding records whether they checked the translation, verified its ABSA annotation, or performed both steps.

The point of access matters when the method uses target materials after training. Fehle et al. (2026, §3.4.3) translate prompt templates into target languages and vary whether annotated demonstrations are supplied. Localized instructions and demonstrations therefore receive separate entries. The test-time code-switching method of Sheng et al. (2025) transforms the input currently being predicted and combines outputs from its variants. The constrained decoder of Šmíd et al. (2025a, §3.3) also derives span candidates from the current input. We record such input processing and any additional resources it uses separately. Examples of additional resources include a lexicon fitted on target data, a retrieval index built from target examples and unlabeled test items used to adapt shared parameters.

Fig. 1 places each resource at the stage where it is used. General multilingual pretraining is recorded with checkpoint provenance. A development set used to choose a checkpoint, threshold or prompt contributes to model selection. Fixed predictions are compared with reference annotations during scoring. If the resulting scores feed back into model or prompt selection, the configuration is coded as test-informed selection. For each check we record supported absence, documented use, or unclear, with a source locator.

A documented use at one stage establishes exposure. A claim of absence requires documentation across the applicable stages.

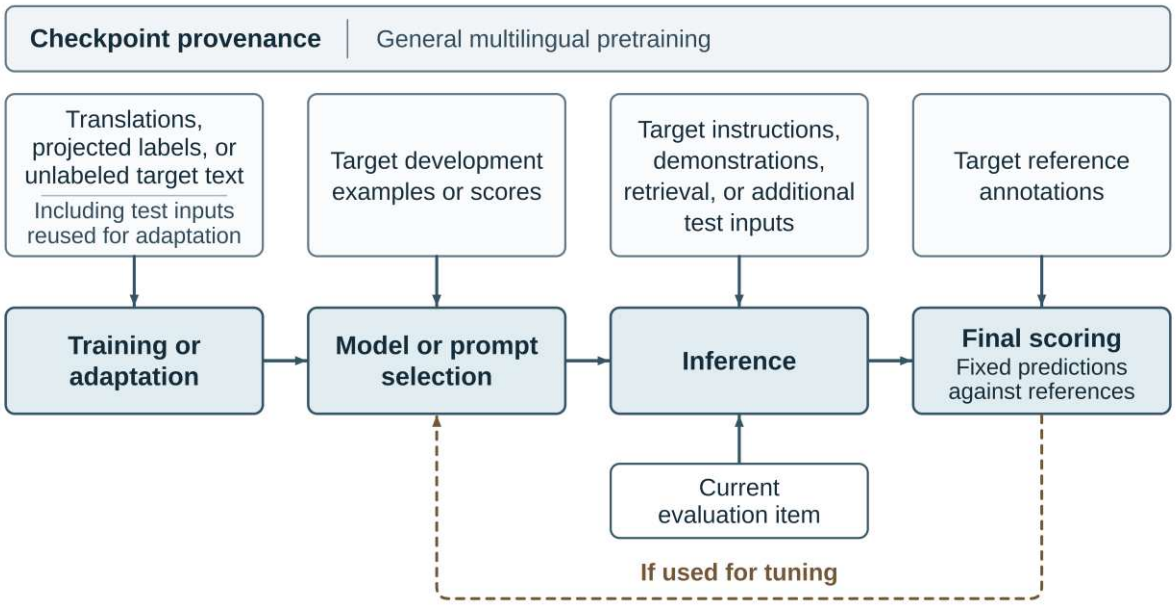


Fig. 1 Stages at which target-language data can enter structured aspect-based sentiment analysis, with checkpoint provenance, current-item inference and reference scoring shown separately (OpenAI Codex assisted with vector-layout code)

2.3 Relationship to existing reviews

Table 1 compares the reviews that define the immediate scope of this study. Hua et al. (2024) provide evidence about ABSA domains, task combinations and methods. Their main review excludes papers whose focus is transferring existing ABSA methods between languages, while their targeted discussion covers zero-shot in-context learning as a prompting condition. Ahmad et al. (2025) organize the literature around aspect extraction and aspect sentiment analysis and document the modeling families and datasets used for these tasks. Zhang et al. (2024) give a focused treatment of quadruple extraction with implicit elements.

The closest review of the transfer setting is Šmíd and Král (2025), which synthesizes cross-lingual ABSA tasks, benchmarks and methods based on multilingual models and translation. Fatemian et al. (2026) also discuss translated training data and distillation with unlabeled target text. Santin and Rezende (2026) study domain shift and explicitly place cross-lingual transfer outside that definition. These scopes support treating the language direction, domain and available supervision as separate descriptors.

The present study organizes the evidence by reported experimental configuration. It links each author's terminology to the documented source, supervision, stage and split of target-language access. This record supports comparison of differently named configurations that use the same kinds of resources, and identifies settings with the same label that require different target-language inputs. Performance contrasts are interpreted using the task and scorer definitions and the documented experimental conditions.

Table 1 Scope of related ABSA reviews

Review	Scope and analytical focus	Relevance to the present review
Hua et al. (2024)	Study and dataset domains, ABSA task combinations and method trends	Main screening excludes transfer of existing ABSA methods between languages. Targeted review distinguishes prompting without examples
Zhang et al. (2024)	Quadruple extraction with implicit aspects and opinions	Establishes why the output schema and implicit-element treatment need to be recorded
Ahmad et al. (2025)	Aspect extraction and aspect sentiment analysis, organized by methods and datasets	Provides a technical taxonomy and identifies the datasets on which methods are evaluated
Šmíd and Král (2025)	Cross-lingual ABSA tasks, datasets and transfer methods	Closest language-transfer comparator. Discusses multilingual models, translated data and zero-shot settings
Santin and Rezende (2026)	Cross-domain ABSA, its methods and benchmark evaluation	Defines domain shift separately from cross-lingual transfer
Fatemian et al. (2026)	Transfer-learning methods, datasets and performance across ABSA settings	Discusses translated-data adaptation and distillation on unlabeled target data

3 Systematic Review Methodology

3.1 Review design and protocol

We designed a systematic review and an experiment-level audit of target-language data use in structured aspect-based sentiment analysis. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guide reporting of the review and selection flow (Page et al. 2021). PRISMA-S guides reporting of information sources, executed strategies, limits, dates and deduplication (Rethlefsen et al. 2021). The evidence comes from published experiments and their associated materials.

The starting materials were a working protocol (version 0.1, 15 November 2026), a purposive pilot and a bibliographic register prepared for a broader audit of cross-lingual ABSA benchmarks. The pilot recorded decisions for seven reports and extracted six reported results from four primary studies. The current review narrows the task criterion to linked outputs that include sentiment and adds separate checks of human-annotated target examples and all task-specific target-language material. We retain the earlier record identifiers and document the scope changes and affected fields. The date of the narrower amendment was not available in the decision files. The available documentation identifies a working protocol. A public registration identifier was unavailable in the review records.

The documented work to date consists of the initial search, preliminary screening and selected evidence extraction. The earlier evidence check covered 21 provisionally eligible reports and 17 candidates. It recorded 23 working inclusions, eight contextual reports and seven decisions awaiting evidence. The corresponding 52 outcome rows were selected to examine benchmark and protocol differences. These records inform the current extraction form and the identification of candidate configurations. The narrower eligibility and exposure checks apply to each configuration before it contributes to the current review.

ChatGPT and Codex (OpenAI) assisted with bibliographic preparation, provisional screening, evidence extraction and coding, manuscript drafting and editing, and figure code. The three diagrams use review-defined labels and recorded data. They were prepared as scripted vector drawings and exported with Inkscape and Poppler. The working records provide source locators for checking the outputs and identify decisions still awaiting verification or independent review.

3.2 Information sources and search strategy

The initial search was executed on 15 November 2026 (Asia/Tashkent). The main structured search used the ACL Anthology BibTeX export with abstracts. The export contained 139,573 entries. After the removal of 3,032 proceedings containers and 222 books, 136,319 publication records were searchable. The search used titles and available abstracts. The 45,815 records without an abstract were searched by title. The search covered the export as a whole. The log preserves the export URL, server timestamp and SHA-256 digest.

The matcher used Unicode Normalization Form Compatibility Composition (NFKC), case-insensitive matching, normalized hyphens and whitespace, and removal of LaTeX braces. Whole-word boundaries limited incidental matches to short forms. Each block used OR between terms. The executed ACL Anthology query was (B1 AND B2) OR (B3 AND B2). The three term blocks were as follows.

B1 — ABSA "aspect-based sentiment analysis", "aspect based sentiment analysis", "aspect-level sentiment analysis", "aspect level sentiment analysis", "aspect-based opinion mining", ABSA

B2 — Transfer "cross-lingual", crosslingual, "cross lingual", multilingual, "cross-language", "language transfer", "translate-train", "translate-test", "zero-shot", "code-switching", "low-resource"

B3 — Structured tasks "aspect sentiment triplet extraction", "aspect sentiment quadruplet extraction", "aspect sentiment quadruple prediction", "aspect-category-opinion-sentiment", "aspect opinion pair extraction", "target aspect sentiment detection", "unified aspect-based sentiment analysis", ASTE, ASQP, ASQE, ACOS, AOPE, TASD, UABSA

B1 AND B2 matched 55 records and B3 AND B2 matched eight. Six appeared in both sets, yielding 57 ACL Anthology records. Supplementary identification contributed 122 occurrences from the pilot (7), publisher and citation trails (13), earlier literature and citations (19), targeted 2024–2026 searches (27), journal and technical publisher pages (24), and indexed pages and citation trails (32). These figures count occurrences supplied to the register. Online Resource 1, Sections S1–S3, gives the executed searches and counts. The Search Log and Screening worksheets in Online Resource 2 identify the source route and query or citation trail when available.

Two title queries through the Crossref application programming interface (API) provided a supplementary metadata check. The exact query.title values were “cross-lingual aspect sentiment triplet extraction” and “multilingual aspect sentiment quadruple extraction zero-shot transfer”. Both used until-pub-date:2026-11-15 and rows=60. Together they returned 120 ranked records with 119 distinct digital object identifiers (DOIs). Eighteen DOIs linked to entries already in the register. The Crossref output was retained as a separate search log. Two further Crossref queries returned HTTP 429. Scopus and Web of Science Core Collection advanced searches were unavailable, and ACM Digital Library and IEEE Xplore records were identified through indexed pages, DOIs and citations. Direct exports from those four interfaces are absent from the executed search. The final search log will date and document any later search expansion separately.

3.3 Eligibility criteria

Table 2 gives the eligibility rules for the current review. A report contributes to the main review if it contains at least one eligible configuration with an identifiable donor-to-target transfer direction and a reported quantitative evaluation of linked ABSA outputs in the target language. The configuration must be described as zero-shot or zero-resource, or be designed to transfer without human-annotated target-task examples. The latter route covers documented source-only transfer, source-derived translation and unsupervised target adaptation. When a report uses a zero-shot label and the procedure raises doubts about that label, the configuration remains eligible for the terminology audit.

An eligible donor may be one language or a documented multilingual donor set distinct from the target in the transfer condition. Target-supervised, few-shot and monolingual target arms are retained as linked controls when their report includes an eligible transfer arm. General multilingual pretraining by itself establishes checkpoint provenance. Direction is determined from the task-specific transfer setup. A documented development-only evaluation can inform the description of data exposure. The performance contrast analysis uses the held-out target test split.

Table 2 Eligibility criteria for the zero-shot transfer review

Criterion	Eligible evidence	Context or exclusion
Task and outcome	Text-based ABSA with at least two jointly evaluated elements including categorical polarity or a dimensional sentiment value	Standalone extraction, polarity given a gold aspect, or isolated element scores
Transfer and evaluation	Identifiable donor language or donor set, distinct target language, and a quantitative linked-output evaluation in that target	Cross-domain-only designs and multilingual evaluations with an unidentified transfer direction
Resource setting	Reported zero-shot or zero-resource claim, or a configuration designed to transfer without human-annotated target-task examples	Target-supervised arms are linked controls when the study contains an eligible transfer arm
Publication	Peer-reviewed original journal or conference report. No publication-language restriction	Reviews, benchmark descriptions without directed experiments, and preprints without a peer-reviewed report are retained as contextual or supplementary leads

3.4 Study selection and units of analysis

The 57 ACL Anthology records and 122 supplementary occurrences yielded 179 entries for deduplication. DOIs were normalized first. Title and publication year were used when a DOI was unavailable, and metadata were checked when linking publication versions. The log documents 65 merges, comprising 60 DOI matches, four title–year matches and one preprint–journal match. The resulting register contains 114 unique bibliographic records. Its preliminary decisions were 21 potentially eligible, 17 candidates, 36 contextual and 40 out of scope.

We accounted for all 114 registry records in the disposition-based analysis. Appendix A and Table 7 cite each record by author and year and explain its role in the review. Every retained search record is traceable to either the References or the publication-type notes in Appendix A. The outcome synthesis uses the selected experimental outcomes relevant to each research question.

A record is a bibliographic entry. A report is a document describing a study and can have multiple versions. A study can span several reports. Separate published reports within a study family remain

traceable as reports. When a conference paper and its extension share an experiment, the configuration is counted once and both evidence locators are retained. The main bibliographic citation uses the version of record. Supplementary versions can substantiate implementation details when their relationship to the reported experiment is clear.

The protocol calls for title and abstract screening followed by full-text assessment. An unavailable full text is recorded separately from an assessed report. A full-text exclusion receives one primary reason in the decision log, selected in the following order: ineligible publication type, topic outside text-based ABSA, no qualifying linked output, no identifiable transfer direction, no quantitative target-language evaluation, or no eligible transfer configuration. An unresolved decision retains the reason and evidence needed to settle it.

The protocol specifies two independent reviewers for final screening and reconciliation by discussion, with third-person adjudication if needed. The current registry has provisional decisions and empty independent-review and reconciliation fields. Agreement will be calculated from completed paired decisions before reconciliation. The PRISMA flow will use the final verified counts for identification, retrieval and inclusion, with the numbers of reports and studies stated separately.

3.5 Data extraction and coding

The primary unit for the audit is the experimental configuration. We link the study, report, configuration, resource and outcome to stable identifiers. A configuration is defined by a donor set, target language, task, benchmark and split, model or checkpoint, and experimental treatment. Changing a treatment or evaluation condition creates another configuration. Repeated seeds and different metrics from the same configuration are retained as outcomes. A resource entry is linked to its configuration and stage of use so that a single input can be followed across stages without inflating the number of configurations.

Table 3 specifies the extraction fields. For each target-language resource, the form records its origin, supervision, amount, split, purpose and stage. Origin distinguishes source-derived translation and code-switching from independently collected target material. Supervision distinguishes unlabeled text, projected annotations, teacher or other pseudo-labels, human-annotated target examples and mixed cases. The form records human language review and human ABSA label verification separately. The applicable dataset and artifact versions and exact source locators support each coding decision.

The review preserves the authors' terminology and applies the two checks from Sect. 2.2 to each eligible configuration. Each check is coded as supported absence, documented use or unclear. Documented use at one stage establishes exposure. Supported absence requires evidence covering all applicable stages. A field can also be not applicable when a stage is absent, or not reported when the source omits an applicable detail. These field-level states inform the overall judgment. For instance, source-language fine-tuning with direct target inference and target-language distillation both occur in the same report, with different documented resources (Zhang et al. 2021, Sects. 2.5 and 3.1–3.3).

Table 3 Core variables for experiment-level extraction

Variable	Entry to record
Study and report	Stable identifiers, DOI, publication version, study family, decision and locator
Configuration and data	Model and checkpoint, treatment, donor set and target language, domain, linked output schema, dataset version, split IDs and test text provenance
Author terminology	Verbatim transfer label and definition with a page, section, table, supplement or code locator

Variable	Entry to record
Resource and supervision	Text, instructions, examples, lexicon or other material. Record provenance, label type, quantity and unit, human involvement, and evidence
Stage and access	Checkpoint provenance, task adaptation, model or prompt selection, inference or scoring. Record split, purpose and whether access is instance-wise or uses a collection
Exposure checks	Status for human-annotated target ABSA examples and all task-specific target-language resources, with applicable stages and evidence locators
Outcome	Benchmark scope, metric and scale, matching, normalization, aggregation, scorer version, score and reported uncertainty
Evidence and verification	Paper or artifact locator, access date, repository revision, agreement between sources, and separate execution or recalculation status

3.6 Evidence appraisal and synthesis

The evidence appraisal examines task and output definition, target-resource documentation, split provenance and access, scoring, and reproducible artifact identification. It records evidence for each domain and flags unresolved items such as test-informed model selection, query or corpus overlap, and uncertain translation or alignment treatment. An artifact link, an accessed file, a verified version match and a reproduced numerical result are distinct evidence states. If the available implementation and the paper disagree, both locators are retained. The score stays on its original scale and any unverified interpretation remains unresolved.

For RQ1 and RQ2, the worked synthesis relates reported terminology to documented resource use. The reported denominator is the number of selected configurations with explicit coding, including unresolved values. A cross-tabulation reports unclear judgments alongside the resolved categories. Resource types and stages permit multiple entries per configuration. For each type and stage, repeated references to the same configuration are counted once. Any study-level summaries use unique study identifiers.

For RQ3, quantitative contrasts are drawn within studies from arms evaluated on the same target test items and reference annotations, with compatible donor and target settings, output schema and implicit-element treatment, scoring and aggregation, and sufficiently aligned model, prompting and selection conditions. The coding records which input or treatment changes and any accompanying changes in training data volume or model budget. When several factors vary together, the contrast is interpreted as a joint change. For comparable arms, the contrast is the arm score minus the reference score. Units are retained or converted only when the scores differ solely in scale. Uncertainty is reported when available for the relevant contrast. Separate arm-level intervals are recorded as such.

Categorical exact F1, partial or weighted F1 and dimensional continuous F1 are synthesized within their own scoring protocols. Heterogeneous and unresolved contrasts are described narratively, with the reason for limited comparability recorded. The pilot comparison of two M-ABSA configurations illustrates the distinction between term–polarity pairs and term–category–polarity triplets on the same benchmark (Wu et al. 2025b, Table 5). The broader pilot and working audit transcribed published scores. The present protocol calls for source-linked descriptive contrasts and narrative synthesis.

For RQ4, the outcome is the extent to which each label and resource classification can be verified from the cited source or artifact. The synthesis reports documentation gaps and conflicting evidence, with the unit and denominator explicit for each summary. Online Resource 1 documents the search, protocol

amendments and provisional selection. Online Resource 2 links the screening register to the extraction codebook, configuration, resource and outcome records, and contrast decisions.

4 Results

4.1 Study selection and evidence coverage

The search and supplementary identification produced 179 record occurrences. Deduplication merged 65, leaving 114 records (Fig. 2). The preliminary register prioritized 38 for additional evidence checks, comprising 21 potentially eligible reports and 17 candidates. The remaining 76 records were classified as contextual (36) or outside scope (40) at that stage. The earlier evidence check of the 38 reports yielded 23 working inclusions, eight contextual reports and seven decisions awaiting evidence, with 52 selected outcome rows.

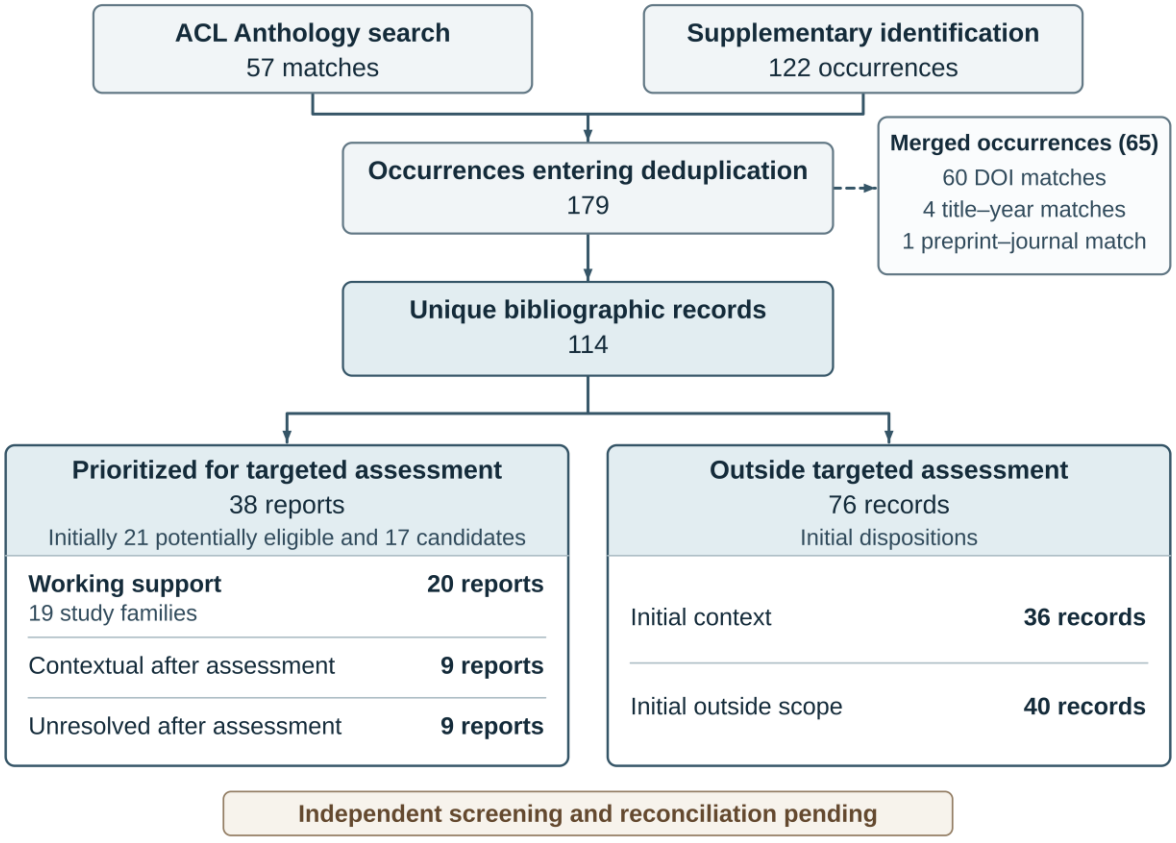


Fig. 2 Record identification and provisional study selection, showing deduplication, targeted assessment and initial dispositions (OpenAI Codex assisted with vector-layout code)

Reapplying the directional transfer and linked-outcome criteria to the 38 prioritized reports yielded working support for 20 reports. Nine were contextual and nine remained unresolved. Two reports moved from working support to unresolved because their qualifying resource settings required verification. One also lacked a verified primary result table broken down by language. One previously retained report used target-language labels for cross-lingual knowledge fusion and monolingual fine-tuning and moved to the contextual group (Wang et al. 2026, Sects. 3.4–4.4 and Tables 2–4). The 20 supported reports represent 19 working study families after linking the conference paper and its expanded study under the registry's version rule (records ABSA-006 and ABSA-009). Independent decisions, reconciliation and final PRISMA selection counts remain pending.

The detailed synthesis uses 16 configurations checked against the cited sources from nine of the supported reports. We selected them to cover direct donor transfer, source-derived target representations, teacher supervision, current-input transformations and access to multiple test items. Within this subset, the coded directions are English to Spanish, Czech, French, German or Chinese. Outputs include aspect–polarity pairs, aspect-term–opinion–polarity triplets, aspect-term–category–polarity triplets, categorical quadruplets and dimensional quadruplets. The cases include native review benchmarks and a parallel benchmark with sampled human quality checks. The configuration counts describe this subset. The broader evidence check includes additional languages and settings, such as English-to-Chinese categorical triplet extraction (Liu et al. 2024) and seven-language benchmarking (Fehle et al. 2026).

The other reports with provisional eligibility support broaden the set of designs. They document single and pooled donors, Polish-to-English pair transfer on translated test text, English-to-Chinese triplet transfer, and language-specific linked baselines (Van Thin et al. 2023; Szołomicka and Kocoń 2022; Liu et al. 2024; Wu et al. 2025a). Other reports distinguish projected donor labels, bilingual training, donor-only fine-tuning, and few-shot target controls (Šmíd et al. 2026d; Fehle et al. 2026; Šmíd et al. 2025a; Li and Zhang 2025; Šmíd et al. 2026c; Zhao et al. 2024; Lin et al. 2023). The linked ICAART reports retain both publication locators while treating overlapping experiments as one configuration.

The context and scope flags document recurring sources of screening ambiguity. The dimensional shared task and the structured sentiment shared task define useful evaluation resources (Lee et al. 2026a; Yu et al. 2026; Barnes et al. 2022). Their reports range from regression conditional on a given aspect to general opinion graphs and linked dimensional outcomes (Affan et al. 2026; Pfister et al. 2022; Dharpure and Rusnachenko 2026; Hellwig et al. 2026b). The registry also includes domain transfer, zero-shot prediction for supplied aspects, and directed extraction of aspect spans alone (Shahana et al. 2025; Shen et al. 2025; Wang et al. 2024; Jebbara and Cimiano 2019; Wang and Pan 2018). These cases inform the task and transfer boundaries. The nine unresolved reports require, for example, the original linked score table, clarification of the donor language or a verifiable account of target-language resources (Dewangan and Maurya 2026a; Zhao et al. 2025; Nour et al. 2025; Lin et al. 2024; Lin et al. 2014). Table 7 gives the record-level basis for all five working dispositions.

The worked extract in Online Resource 2 contains 16 arm-level outcomes, including 15 test and one development outcome. The separate Prior Outcomes worksheet preserves 52 historical extraction rows for traceability. The Contrasts worksheet distinguishes the five selected within-study test contrasts from eight contextual or historical comparisons.

4.2 Reported zero-shot definitions and coded settings

The evidence supported absence on both zero-shot checks in four of the 16 configurations. In six it supported absence of human-annotated target-task examples and documented other task-specific target material. In four it documented target material and left use or provenance of human annotations unresolved. Both checks remained unresolved in two. Coding concerns documented access outside routine reference scoring. Table 4 lists the source labels and evidence locators.

A literal ZERO-SHOT label accompanies source-only training and source-language validation in LACA and the sequence-labelling study (Šmíd et al. 2025b, 2026b). The direct-transfer arms labelled CL by Sheng et al. (2025) and Original by Šmíd et al. (2026a) satisfy the same two checks in the selected configurations. Dewangan and Maurya (2026b) call CLAOCS-TX zero-shot and report that target benchmark gold labels were used for evaluation. Their teacher prompt also contains two French input–output demonstrations with ASTE triplets. The annotation provenance of those examples is unspecified, so the human-annotation check for the French configuration remains unclear.

Table 4 uses H for human-annotated target ABSA examples and T for any task-specific target-language material. A indicates supported absence, U documented use, and ? insufficient documentation. Configuration identifiers also index Fig. 3. Pair denotes a jointly evaluated aspect span and its polarity. The same report can contribute several configurations if the training or inference procedure changes.

Table 4 Author terminology and documented target-language data use

Arm	Author label	Target-language resources	Checks	Primary source locator
Z1 Zhang et al. (2021), EN–ES pair	ACS	Donor translations and bilingual variants in training	H: A T: U	§§2.2–2.4, 3.1, Table 2
Z2 Zhang et al. (2021), EN–ES pair	ACS-DISTILL-M	Translations and target train text with teacher scores	H: A T: U	§§2.5, 3.1–3.3, Table 2
S1 Sheng et al. (2025), EN–ES ASTE	CL	Direct inference after English fine-tuning	H: A T: A	§§2, 3.1–3.4, Table 2
S2 Sheng et al. (2025), EN–ES ASTE	CSW	Donor-derived bilingual training, including alignment	H: A T: U	§§2.1–2.2, 3.3, Table 2
S3 Sheng et al. (2025), EN–ES ASTE	CSW +tta	Bilingual training and current-input variants at test time	H: A T: U	§§2.2–2.3, Tables 2–3
L1 Šmíd et al. (2025b), EN–ES pair	ZERO-SHOT	English fine-tuning and model selection	H: A T: A	§§4.1–4.4, Table 2
L2 Šmíd et al. (2025b), EN–ES pair	+LACA LLaMA70	Unlabeled target text, teacher predictions and rewritten examples in adaptation	H: A T: U	§§3.3–3.4, 4, Table 2
C1 Šmíd et al. (2026a), EN–CS ASQP	Original	English training, selection and task instructions	H: A T: A	§§4–4.3, Table 5
C2 Šmíd et al. (2026a), EN–CS ASQP	Translated	English plus aligned Czech translations of donor data. The ASQP translation demonstration provenance is unresolved	H: ? T: U	§§3.3, 4.3, Fig. 4, Table 5
E1 Šmíd et al. (2026b), EN–ES pair	ZERO-SHOT	English fine-tuning and model selection	H: A T: A	§§4.1–4.4, Table 1
E2 Šmíd et al. (2026b), EN–ES pair	ACS	English, translations and bilingual variants in training	H: A T: U	§§3.3, 4.1–4.4, Table 1

Arm	Author label	Target-language resources	Checks	Primary source locator
D1 Dewangan and Maurya (2026b), EN–FR ASTE	Zero-shot ASTE	Unlabeled target adaptation text with pool unspecified, teacher pseudo-triplets and two French demonstrations of unspecified annotation provenance	H: ? T: U	§§3–4, Table 2, App. A Fig. 5
M1 Wu et al. (2026a), EN–ES pair	MSMO	Donor translations and bilingual variants used in training	H: ? T: U	§§2.1, 4.1, Table 3
M2 Wu et al. (2026a), EN–ES pair	MSMO-DISTILL-M	Reported student adaptation on target test inputs with teacher soft labels. Run-level overlap unverified	H: ? T: U	§§3.1–3.3, 4.1, Fig. 5, Table 3
W1 Wu et al. (2025b), EN–DE TASD	English-centric zero-shot transfer	Parallel target test with sampled human quality checks. Selection inputs unclear	H: ? T: ?	§§4–6.2, Table 5(a)
G1 Gazetas et al. (2026), EN–ZH DimASQP	ENG-Model	Translation and back-mapping of current dev item. Selection unclear	H: ? T: ?	§§3.3, 5.4, Table 5

4.3 Target-language resources by experimental stage

Fig. 3 traces where the 16 configurations access information. Donor translations and bilingual variants enter training in the ACS studies and the test-time code-switching study (Zhang et al. 2021; Sheng et al. 2025; Šmíd et al. 2026b). In a distinct route, Zhang et al. remove the labels from target training sentences and distil predictions onto those sentences. LACA uses unlabeled target text to obtain teacher predictions, then uses the resulting annotations to produce target training examples. The Czech translated arm adds donor-derived translations and aligned structures to the English training examples and uses English validation. Its exact ASQP translation prompt and the annotation provenance of any Czech demonstrations remain unverified (Šmíd et al. 2026a, Sect. 4.3 and Fig. 4).

At inference, the selected test-time code-switching (TT-CSW) arm transforms the current target sentence and combines predictions from its variants (Sheng et al. 2025). The dimensional aspect sentiment quad prediction (DimASQP) example translates the current target development item into English and maps the predicted elements back (Gazetas et al. 2026). The MSMO distillation procedure describes obtaining teacher scores on target test items and using them to train the student (Wu et al. 2026a, p. 18728, Fig. 5 and accompanying text). The paper describes teacher predictions as the training signal. The correspondence between this description and the executed adaptation pool remains unverified, and the paper does not document access to test reference labels for training.

Online Resource 2 records 86 stage-level entries across the 16 configurations. The stage counts are 16 checkpoint-provenance, 20 training or adaptation, 18 model or prompt selection, 16 inference and 16 scoring entries. Resource and operation rows can share a configuration, so these counts document the structure of the evidence. Two rows have a numeric amount recorded in the workbook. The amounts refer to prompt examples. Comparable data volumes cannot be reconstructed across arms from the current extraction.

Target-language instructions and target demonstrations also require separate recording. Fehle et al. (2026, Sect. 3.4.3) localize their prompt templates and compare zero demonstrations with larger demonstration counts. A localized instruction is task-specific target-language material even at zero demonstrations. The standalone prompt condition lacks a specified task-specific donor and is contextual for this review. In Fig. 3, TR denotes translation or bilingual source-derived training data, PL teacher-supervised target training, PR target-language examples in a teacher prompt, and TEST reported use of target test inputs for training. SRC denotes donor-language selection, DIR direct prediction, and ITEM transformation of the current input. In the scoring column, TEST and DEV refer to the reference split. A dash indicates no documented extra target training input in the shown training condition. A question mark indicates unresolved selection provenance.

ID	Author label	Training or adaptation	Selection	Inference	Reference scoring
Z1	ACS	TR	SRC	DIR	TEST
Z2	ACS-DISTILL-M	TR+PL	SRC	DIR	TEST
S1	CL	—	SRC	DIR	TEST
S2	CSW	TR	SRC	DIR	TEST
S3	CSW +tta	TR	SRC	ITEM	TEST
L1	ZERO-SHOT	—	SRC	DIR	TEST
L2	+LACA LLaMA70	PL	SRC	DIR	TEST
C1	Original	—	SRC	DIR	TEST
C2	Translated	TR	SRC	DIR	TEST
E1	ZERO-SHOT	—	SRC	DIR	TEST
E2	ACS	TR	SRC	DIR	TEST
D1	Zero-shot ASTE	PL+PR	?	DIR	TEST
M1	MSMO	TR	?	DIR	TEST
M2	MSMO-DISTILL-M	TR+TEST	?	DIR	TEST
W1	English-centric zero-shot transfer	—	?	DIR	TEST
G1	ENG-Model	—	?	ITEM	DEV



Fig. 3 Resource access by stage for the 16 configurations in Table 4, with codes distinguishing training resources, selection provenance, current-item inference and reference splits (OpenAI Codex assisted with vector-layout code)

4.4 Performance under comparable conditions

Table 5 reports five descriptive test contrasts with the target dataset, task and score definition aligned within each pair. On the Czech ASQP test, Gemma 3 12B scored 4.96 exact micro-F1 points higher in the arm trained with translations. For the same reported comparison, scores were lower with translations for mT5 (41.68 to 40.07) and Orca 2 13B (41.90 to 36.62). Translated training also changes the number of examples, so the contrasts quantify the published training packages. In LACA the improvement of 4.41

points accompanies target adaptation and additional training. In TT-CSW the 1.8-point change accompanies extra input variants and inference operations while prior training remains code-switched.

Several available comparisons use a different unit or split. The M-ABSA German restaurant scores of 57.20 for TASD and 67.63 for UABSA compare triplets and pairs respectively, even though both use English-trained mT5 and the same parallel benchmark (Wu et al. 2025b, Table 5). Fig. 3 traces the TASD arm. The DimASQP study in Track A, Subtask 3, reports Chinese restaurant development cF1 of 0.26 with English training versus 0.36 with Chinese training examples translated into English and the same inference transformation (Gazetas et al. 2026, Table 5). The split, continuous scoring rule and translated-target supervision make this a development comparison in a separate evidence stratum.

Fehle et al. (2026, Table 3c) report Spanish TASD micro-F1 of 38.87 for DLO-mT5 trained on multiple non-target languages and 51.37 when using English-derived code-switching. The donor mix and training data also differ, so this remains a strategy comparison outside Table 5. For MSMO the reported 69.74 and 69.93 Spanish pair scores correspond to the base and multi-teacher arms using XLM-R (Wu et al. 2026a, Table 3). The paper reports target test inputs for adaptation in the latter arm. Which items entered the published run remains unresolved. We kept these scores descriptive. The reported differences across exact, overlap-weighted and continuous scoring were not pooled.

Table 5 Within-study performance contrasts by target-language data use

Study and direction	Shared test conditions	Change between arms	Metric and scores	Difference and evidence
Zhang et al. (2021), EN–ES	Aspect–polarity, SemEval restaurant test, XLM-R	ACS → ACS-DISTILL-M. Target train text, teacher scores and extra training	Exact micro-F1, 67.32 → 69.24	+1.92 points. Five-run means, paired interval unavailable. Table 2
Sheng et al. (2025), EN–ES	ASTE, OpeNERES test, mT5-base	CSW → CSW +tta. Input variants, alignment and output fusion	Overlap-weighted F1, 49.1 → 50.9	+1.8 points. Paired interval unavailable. Table 3
Šmíd et al. (2025b), EN–ES	Aspect–polarity, SemEval restaurant test, XLM-R	ZERO-SHOT → LACA LLaMA70. Unlabeled target text, pseudo-labeled examples and continued training	Exact micro-F1, 67.48 → 71.89	+4.41 points. Five-run means, paired interval unavailable. Table 2
Šmíd et al. (2026a), EN–CS	ASQP, Czech restaurant test, Gemma 3 12B	Original → Translated. Add English-derived Czech training examples	Exact micro-F1, 44.17 → 49.13	+4.96 points. Five-run means, paired interval unavailable. Table 5
Šmíd et al. (2026b), EN–ES	Aspect–polarity, SemEval restaurant test, mT5	ZERO-SHOT → ACS. Add aligned target translations and bilingual variants	Exact micro-F1, 62.30 → 70.84	+8.54 points. Five-run means, paired interval unavailable. Table 1 with arm details in Table 3

4.5 Reporting gaps and verification

Model-selection provenance remained unclear in five of the 16 configurations. For CLAOCS-TX, provenance of the teacher demonstrations, the exact adaptation pool and the French opinion-span benchmark version remained unresolved. For the Czech translated arm (C2), Fig. 4 of Šmíd et al. (2026a) illustrates the translation workflow with an ACOS prompt containing a labeled Czech output and describes five examples in the full prompt. The exact ASQP prompt and label provenance remain unverified, so H is unresolved for C2. For M-ABSA, reporting of the parallel translations and sampled human quality checks describes benchmark construction. It leaves which data were used for model selection open (Wu et al. 2025b, Sects. 5–6).

Two source comparisons illustrate why reported values are traced back to the original experiment. QPEN describes macro-averaged F1 and carries forward rows from Zhang et al. (2021), whose scoring section specifies micro-F1. For example, the Spanish XLM-R ACS-DISTILL-M value of 69.24 appears in both reports (Zhao et al. 2024, Table 2 and Comparison Results, p. 19675). The overlapping row belongs to the earlier experiment, and the description in QPEN leaves the scoring implementation for its new rows unresolved in this check. In CLAOCS-TX, the published TT-CSW comparator appears across benchmarks and languages that differ from the original Sheng et al. (2025) experiments. The table entry cannot be traced to a matched rerun by the row alone.

For the worked configurations, the evidence includes methods, tables, appendices and split descriptions in the primary reports. Online Resource 2 preserves the relevant locators, including five verification records covering score attribution, test-input access, unmatched comparators, target demonstration provenance and publication-version overlap. Where available, code locators support inspection of a scorer or data-loading procedure. Version correspondence, actual adaptation–test overlap and independent score reproduction remain separate verification steps.

5 Discussion

5.1 Interpretation of zero-shot terminology

A zero-shot claim in structured ABSA specifies a restriction on what is available to an experimental configuration. The worked cases show the value of stating that restriction explicitly. Direct transfer with donor-only training and selection satisfied both checks in the selected arms of Sheng et al. (2025) and Šmíd et al. (2025b, 2026a, b). The studies used different labels for these arms. LACA uses target-language text and teacher supervision, with demonstrations drawn from donor training data in the reported prompt. The Czech translated arm transfers donor annotations to Czech sentences and leaves the role and annotation provenance of any ASQP translation demonstrations unresolved. These configurations establish different target-resource conditions and support different claims about human annotation (Šmíd et al. 2025b, 2026a).

The number of demonstrations is a separate property. A prompt can include localized task instructions and zero demonstrations, as in Fehle et al. (2026, Sect. 3.4.3). In the French CLAOCS-TX configuration, the published teacher prompt shows two target-language input–output examples, and their annotation provenance is unspecified. The statement that benchmark gold is reserved for scoring therefore establishes the role of the benchmark labels while leaving the origin of the demonstrated labels unresolved (Dewangan and Maurya 2026b, Sect. 4.1 and Appendix A, Fig. 5). Recording the evaluator, teacher and student stages makes that distinction visible even when only the student is used at prediction time.

The meaning of access also depends on its unit. Translating or code-switching the current input introduces another prediction operation while leaving other test items outside that operation in the TT-CSW example (Sheng et al. 2025, Sect. 2.3). The published MSMO procedure describes student training with teacher predictions over target test inputs. Under that procedure, the score would be conditional on access to that collection (Wu et al. 2026a, Sect. 3.3 and Figs. 4–5). Study reports should specify whether a method can serve each item as it arrives or needs a batch before prediction. General multilingual pretraining belongs to checkpoint provenance. The two checks concern documented task-specific access and leave pretraining exposure to a particular benchmark open.

5.2 Reporting recommendations

General ABSA reviews emphasize the role of dataset choice and task definition, and the cross-lingual survey identifies translation and multilingual training as distinct transfer strategies (Ahmad et al. 2025; Hua et al. 2024; Šmíd and Král 2025). Table 6 extends that documentation to individual experimental arms. It follows the principle that describing the origin and scope of data makes generalization claims more specific (Bender and Friedman 2018). For each arm, the report should declare what the zero-shot label excludes and attach the corresponding resource and evaluation record to the score. A single dataset paragraph can describe shared settings. Document split, prompt and checkpoint changes against the relevant arm.

Two fields deserve particular attention. The model-selection record should identify the data used to choose checkpoints, thresholds and prompts. The baseline record should give the original source and scoring definition for copied results, or describe any rerun with its evaluation protocol. The QPEN and CLAOCS-TX examples in Sect. 4.5 show how copying a value without its original metric or rerun specification can obscure comparability. A report should distinguish confirmed absence of access from an applicable detail it could not document. A versioned configuration file with the same arm IDs as the results table can carry details such as prompt text, split identifiers and scorer options.

Table 6 Recommended reporting items for structured cross-lingual ABSA

Reporting item	Minimum information for each experimental arm
Claim and configuration	Give an arm identifier, the exact zero-shot claim, its restriction, the donor and target languages, the domain, the task and the predicted tuple.
Resource provenance and supervision	Give the original and rendered languages, source and split, quantity, preparation method and supervision for each material. Distinguish translation review, human label verification, projected labels and teacher predictions.
Model and prompt selection	Identify checkpoint and language or dataset split used to tune it. List task instruction and example provenance, example counts, and the data used to choose prompts, thresholds and settings.
Adaptation and prediction	Declare whether the system accesses independent target text, extra test items or only the current input. Describe any translation, retrieval, parameter updates or aggregation. State when the model is frozen and who can access the test gold.
Evaluation definition	Report test version, split and text provenance. Specify tuple fields, how implicit elements and spans are handled, the matching rule, F1 aggregation or continuous scoring definition, and scorer version.
Contrasts and uncertainty	Name the reference arm and document changes in data amount, training budget and inference calls. Report runs and variability when available. Give paired uncertainty for the difference where estimable.

Reporting item	Minimum information for each experimental arm
Outcome and artifact provenance	Identify which scores are newly computed, rerun or taken from earlier reports. Give the original result locator and link model, data, code or scorer revisions, prompts and configurations available for verification.

5.3 Implications and future research

A priority for low-resource evaluation is a baseline that can be run with donor labels and source validation, paired with arms that add one documented resource at a time. A benchmark can keep the target test, tuple schema and model fixed while varying access to donor translations, independently collected unlabeled target text or a small annotated target set. Matching the number of optimization steps or separately reporting the effect of extra training can help distinguish the resource change from the training budget. For input transformations, report performance together with the number of translation or model calls per item. The differing Czech ASQP responses across backbones and the inference change in TT-CSW motivate testing these effects within a fixed model and reporting where they vary (Šmíd et al. 2026a, Table 5; Sheng et al. 2025, Table 3).

Benchmark provenance should shape how such results are interpreted. A parallel test supports comparisons on aligned content, as in M-ABSA. An independently collected target test measures performance on reviews written directly for that setting, as illustrated by the Uzbek and Uyghur app reviews in LASQ (Wu et al. 2025b; Yusufu et al. 2026). A paired evaluation on both types can document whether a method retains its performance when the text is drawn from a target-language community. Such a design should harmonize category inventories and annotation policies, disclose any remaining domain differences, and keep parallel versions of each source item within one split. Report errors by span boundary, category and polarity. For languages where inflection or script variation affects span matching, evaluate these cases explicitly and publish the normalization policy alongside the scorer.

Future comparisons should also vary the donor. All worked configurations used English as the donor. Testing a non-English donor in the same domain alongside an English donor with matched amounts of labeled data would help separate source choice from data volume. If the aim is to measure adaptation on an unlabeled target test batch, report that protocol as a separate condition alongside the test in which each item is predicted with a fixed model. Report uncertainty on paired comparisons where runs or test predictions support it. This would allow readers to judge when a difference is reproducible under the stated resource conditions.

5.4 Review limitations

The structured search used English terms over ACL Anthology titles and available abstracts. For 45,815 records, only titles were searchable. Supplementary discovery widened the search, while advanced searches in Scopus and Web of Science were unavailable and ACM Digital Library and IEEE Xplore coverage relied on indexed pages, DOIs and citations. Research described using other terms or published outside these routes may be underrepresented. The eligibility rules also restrict inference to peer-reviewed, directed evaluations of linked ABSA outputs with a sentiment element.

The narrower criteria were applied in a targeted assessment of 38 of the 114 records. Nine decisions remained unresolved, and independent screening and reconciliation are pending. The 16 worked configurations were selected from nine reports to show different types of access. Several share a report or benchmark. The coding frequencies therefore describe these configurations and do not estimate prevalence across the literature.

The review codes access from documented procedures. Missing validation details, demonstration provenance or split membership can alter a judgment. The worked synthesis checks published results and locators. It leaves artifact version matching, independent score reproduction and unresolved adaptation–test overlap for further verification. In addition, incomplete pretraining corpus information limits checks for earlier benchmark exposure. The within-study contrasts quantify the published packages. Changes in data volume, training or inference can co-occur with changes in access, leaving the separate contribution of an added resource unidentified. Differences in benchmarks, output schemas and scoring rules precluded a pooled performance estimate.

6 Conclusion

What counts as zero-shot in structured ABSA depends on the target-language resources withheld and the stages covered by the claim. The worked cases distinguish donor-only training and selection from procedures that use translated examples, unlabeled target text or teacher supervision while withholding human-annotated target-task examples. A zero-demonstration claim specifies the prompt's example count. Adaptation on target test inputs changes the evidence available before prediction. The provenance of some prompts and selection data remains unresolved.

We propose reporting both restrictions for each experimental arm alongside the origin, supervision and role of each target-language resource. Performance comparisons should identify the shared test split, output structure and scorer, and document any accompanying changes in training or inference. The contrasts in this review quantify the configurations as reported. Linking each score to its data, configuration and original source lets readers assess whether a comparison answers the resource question they intend to ask.

Appendix A Bibliographic coverage and record-level assessment

Table 7 maps the 114 unique records to their role in the review. The dispositions follow Online Resources 1 and 2. The 38 targeted records comprise 20 with working support, nine contextual and nine unresolved. The other 76 retain their earlier flags. Source-specific notes indicate how each record informs eligibility, background, outcome interpretation or a boundary of the synthesis. The record ID links each row to the Screening worksheet. Independent screening and final reconciliation remain pending.

Table 7 Citation and disposition of each bibliographic record

Record	Source citation	Disposition	Evidence and role in the review
ABSA-001	Gazetas et al. 2026	Working support	English-trained dimensional quadruple extraction illustrates translate-test and target development scoring. Used for exposure mapping.
ABSA-002	Dewangan and Maurya 2026b	Working support	Teacher-supervised target adaptation with pseudo-triplets illustrates separation of target gold and teacher demonstrations.
ABSA-003	Wang et al. 2026	Targeted context	Target-supervised multilingual modeling is a boundary case. The extracted arms use target-language labels and validation.
ABSA-004	Šmíd et al. 2026a	Working support	Original and translated donor training are compared on the same Czech quadruple test. Used for a held-out contrast.
ABSA-005	Šmíd et al. 2026b	Working support	Source-only and code-switched training illustrate changes in target material under English validation. Used for a held-out contrast.
ABSA-006	Šmíd et al. 2026d	Working support	Extended ICAART report contributes donor directions and model configurations. Linked to ABSA-009 at study-family level.
ABSA-007	Wu et al. 2026a	Working support	Donor-derived bilingual training and teacher predictions on target test inputs illustrate split-specific adaptation.
ABSA-008	Fehle et al. 2026	Working support	German quadruple tests separate translated and code-switched training. Localized prompts show why demonstration count is a separate field.
ABSA-009	Šmíd et al. 2025a	Working support	ICAART conference report provides reported zero-shot English-to-Spanish pairs and constrained decoding. Linked to ABSA-006.
ABSA-010	Li and Zhang 2025	Working support	Donor-derived translations support directed aspect–polarity prediction. Validation language and score aggregation need confirmation.
ABSA-011	Wu et al. 2025a	Working support	Source-language fine-tuning supports directed target evaluation. Donor-unspecified no-demonstration prompts are a distinct setting.
ABSA-012	Šmíd et al. 2026c	Working support	A zero-target-example Czech condition supports comparison design. Few-shot target arms are linked controls.
ABSA-013	Šmíd et al. 2025b	Working support	Spanish aspect–polarity prediction contrasts direct donor transfer with teacher-supervised adaptation. Used for a held-out contrast.

Record	Source citation	Disposition	Evidence and role in the review
ABSA-014	Wu et al. 2025b	Working support	Translated multilingual benchmark separates pair and triplet scoring and illustrates benchmark provenance.
ABSA-015	Sheng et al. 2025	Working support	Code-switched donor training and current-input transformations illustrate item-level inference. Used for a held-out contrast.
ABSA-016	Liu et al. 2024	Working support	English-to-Chinese triplet extraction provides linked sentiment evidence. Aspect–opinion pairs without polarity are outside the outcome criterion.
ABSA-017	Zhao et al. 2024	Working support	Bilingual pair prediction illustrates primary-score attribution and a macro versus micro averaging discrepancy.
ABSA-018	Lin et al. 2023	Working support	Directed aspect–polarity prediction uses donor translations and code-switching. Target-text distillation constitutes a separate arm.
ABSA-019	Van Thin et al. 2023	Working support	Single and pooled donors predict linked category and sentiment. Separate category detection is background only.
ABSA-020	Szołomicka and Kocoń 2022	Working support	Polish-to-English pair prediction uses a translated target set. Exact and relaxed scores belong to one configuration.
ABSA-021	Zhang et al. 2021	Working support	Donor-derived variants and distillation on target training inputs illustrate two different training exposures. Used for a held-out contrast.
ABSA-022	Dewangan and Maurya 2026a	Targeted unresolved	Code-mixed datasets are documented. An identifiable donor setting and jointly scored target output require primary evidence.
ABSA-023	Affan et al. 2026	Targeted context	The cross-language ablation measures regression for given aspects. Linked dimensional results do not establish the same transfer condition.
ABSA-024	Cao et al. 2026	Targeted context	The linked dimensional results emphasize English tasks. The Chinese procedure does not establish a directed linked evaluation.
ABSA-025	Hellwig et al. 2026a	Targeted context	Zero-shot wording and linked outputs are documented. The evaluated benchmarks do not establish a distinct task-specific donor.
ABSA-026	Chen 2026	Targeted context	All-language training and language-specific adapters inform benchmark design. A qualifying directed linked arm could not be isolated.
ABSA-027	Satria et al. 2025	Targeted unresolved	Multilingual aspect extraction and classification are described. Donor direction and joint scoring require the experiment record.
ABSA-028	Zhao et al. 2025	Targeted unresolved	Structured target predictions and direction are documented. The explicit resource-selection details apply to a separate classification task.
ABSA-029	Zhang et al. 2025	Targeted unresolved	The publication and a secondary summary motivate a check of original linked scoring and stage-specific resources.

Record	Source citation	Disposition	Evidence and role in the review
ABSA-030	Zeng et al. 2025	Targeted unresolved	Code-mixed demonstrations and language tags are documented. Direction, linked output and demonstration provenance need experiment-level evidence.
ABSA-031	Nour et al. 2025	Targeted unresolved	Arabic zero-, one- and few-shot triplet extraction motivates checking whether a separate task-specific donor was used.
ABSA-032	Ahmed et al. 2025	Targeted unresolved	Urdu quadruple extraction is documented. Architecture reuse leaves task-specific donor information to be verified.
ABSA-033	Lin et al. 2024	Targeted unresolved	Directional linked results are described. The selected run's target-label exposure and per-language scores require the underlying table.
ABSA-034	Nam 2024	Targeted context	Preprint and separate category and polarity scoring provide a publication and outcome boundary example.
ABSA-035	Cotfas 2023	Targeted context	Master's thesis provides contextual directional end-to-end results and an example of the publication-type boundary.
ABSA-036	Pfister et al. 2022	Targeted context	General opinion graph prediction provides a comparison with aspect-specific joint output and scoring.
ABSA-037	Li et al. 2020	Targeted context	Preprint describes directed linked output and illustrates the need to verify a qualifying publication.
ABSA-038	Lin et al. 2014	Targeted unresolved	Directional aspect-specific analysis is reported. A quantitative joint aspect–sentiment outcome requires the original experiment table.
ABSA-039	Bapanov et al. 2026	Initial context	Hotel-review triplet extraction across three languages illustrates multilingual coverage without an isolated donor–target design.
ABSA-040	Haralovic et al. 2026	Initial context	Structured crisis sentiment informs task coverage. Directional cross-language transfer is identified as future work.
ABSA-041	Lee et al. 2026a	Initial context	Dimensional datasets and continuous scoring provide benchmark context. A separate donor condition was not identified.
ABSA-042	Yusufu et al. 2026	Initial context	Native Uzbek and Uyghur app reviews illustrate target benchmark provenance. No directed transfer was found in the checked evidence.
ABSA-043	Gheorghe et al. 2026	Initial context	Romanian linked-task benchmark informs language coverage. The isolated English-transfer result concerns category classification.
ABSA-044	Yu et al. 2026	Initial context	The official shared-task report defines dimensional output units and continuous scoring. Team reports supply the directed configurations.
ABSA-045	Santin and Rezende 2026	Initial context	Cross-domain review informs the distinction between domain transfer and directed language transfer.
ABSA-046	Strothe 2026	Initial context	Combined multilingual training illustrates why exposure to the evaluation language in training needs explicit reporting.

Record	Source citation	Disposition	Evidence and role in the review
ABSA-047	Fatemian et al. 2026	Initial context	Transfer-learning review supplies background and citation trails to original transfer reports.
ABSA-048	He and Zhou 2026	Initial context	English linked dimensional outputs and multilingual given-aspect regression illustrate distinct task units.
ABSA-049	Galhom et al. 2025	Initial context	Arabic restaurant triplet benchmark informs target resources. The documented results do not isolate a distinct task-specific donor.
ABSA-050	Hammad et al. 2025	Initial context	Arabic game reviews illustrate domain diversity. Reported component scores leave joint output and direction unverified.
ABSA-051	Ahmad et al. 2025	Initial context	Supplied comprehensive ABSA review informs taxonomies of tasks and resources and the identification of primary sources.
ABSA-052	Dewangan et al. 2025	Initial context	Odia annotation and multilingual-model fine-tuning inform target benchmark lineage. Opinion-span provenance in later use remains to be verified.
ABSA-053	Šmíd and Král 2025	Initial context	Cross-lingual ABSA survey supplies task and dataset taxonomy and citation trails to primary transfer work.
ABSA-054	Chiril et al. 2025	Initial context	English and French media corpus provides context for multilingual ABSA resources and domain breadth.
ABSA-055	Hua et al. 2024	Initial context	Supplied systematic review informs domain, method and benchmark context. The preprint and journal version share one registry entry.
ABSA-056	Rink et al. 2024	Initial context	Dutch open-ended survey analysis illustrates a domain application in which a distinct zero-shot donor was not established.
ABSA-057	Šmíd et al. 2024a	Initial context	Czech structured-task benchmark documents annotation compatibility and provides a resource for later transfer experiments.
ABSA-058	Dejaeghere et al. 2024	Initial context	Multilingual literary corpus illustrates non-review domains. The checked evaluation emphasizes the English part.
ABSA-059	Zhang et al. 2024	Initial context	Review of aspect quadruples informs implicit-element policies and quadruple task definitions.
ABSA-060	Nawawi et al. 2024	Initial context	Tourism application illustrates zero-shot terminology without verified structured transfer from a separate task-specific donor.
ABSA-061	Luo et al. 2024	Initial context	Multilingual multimodal sextuples illustrate a broader task boundary. Directed transfer is future work in the cited report.
ABSA-062	Lango et al. 2024	Initial context	Polish triplet dataset informs resource coverage and structured task definitions.
ABSA-063	Matlatipov et al. 2024	Initial context	Uzbek restaurant-review dataset informs native target benchmark availability.
ABSA-064	Šmíd 2023	Initial context	Student proceedings report directional category–term and term–category–sentiment results. Publication status and overlap remain open.

Record	Source citation	Disposition	Evidence and role in the review
ABSA-065	Šmíd and Přibáň 2023	Initial context	Czech prompting provides structured benchmark context. The checked zero-shot use does not identify a separate ABSA donor.
ABSA-066	Chen et al. 2022a	Initial context	General cross-language opinion graphs provide context for holder, target, expression and polarity outputs.
ABSA-067	Barnes et al. 2022	Initial context	Shared-task specification establishes the broader structured sentiment graph used in comparison.
ABSA-068	De Geyndt et al. 2022	Initial context	The multilingual sentiment platform provides pipeline context. A jointly scored directed ABSA outcome was not identified.
ABSA-069	Yang et al. 2021	Initial context	Chinese-oriented multi-task component results illustrate why multilingual datasets and separate component scores require a directional linked test.
ABSA-070	Gimadi et al. 2021	Initial context	Kazakh app reviews and naturally code-switched text provide target-language and domain context.
ABSA-071	Regatte et al. 2020	Initial context	Telugu annotation and component baselines document resources for a low-resource target language.
ABSA-072	Ruder et al. 2016	Initial context	Multilingual shared-task entry reports category and polarity outputs separately and provides a component-scoring example.
ABSA-073	Pontiki et al. 2016	Initial context	Original SemEval benchmark provides domain, language, split and annotation provenance for downstream comparisons.
ABSA-074	Klinger and Cimiano 2015	Initial context	English–German aspect and subjective expression spans are evaluated separately. The linked polarity criterion remains unestablished.
ABSA-075	Hu 2026	Initial outside scope	Dimensional system report specifies English-only evaluation and illustrates a transfer-direction mismatch.
ABSA-076	Suryabrata et al. 2026	Initial outside scope	Pooled Indonesian and English hotel reviews illustrate multilingual training without an isolated directed transfer test.
ABSA-077	Dharpure and Rusnachenko 2026	Initial outside scope	Multilingual training is evaluated through valence and arousal regression for supplied aspects.
ABSA-078	Li and Yang 2026	Initial outside scope	Translation assists Nigerian Pidgin and Swahili regression for supplied aspects. The output lacks a jointly extracted structure.
ABSA-079	Hellwig et al. 2026b	Initial outside scope	Structured dimensional outputs are reported across languages after training within task–language–domain splits.
ABSA-080	Wu et al. 2026b	Initial outside scope	Language- and domain-specific models estimate valence and arousal for supplied aspects.
ABSA-081	S JM and S AR 2026	Initial outside scope	Regression conditions on a supplied aspect and does not jointly extract linked elements.
ABSA-082	Tu et al. 2026	Initial outside scope	Vietnamese-to-Swedish transfer is evaluated as paragraph sentiment toward a specified topic. The unit is not a linked ABSA output.

Record	Source citation	Disposition	Evidence and role in the review
ABSA-083	Lu et al. 2026	Initial outside scope	Transductive zero-shot quadruple prediction uses English benchmark datasets. The transfer axis differs from directed language transfer.
ABSA-084	Lee et al. 2026b	Initial outside scope	The reported learning-transfer comparison measures valence and arousal regression with root mean squared error.
ABSA-085	Frolov and Rykov 2026	Initial outside scope	Dimensional scores are reported for supplied aspects without a jointly extracted structured outcome.
ABSA-086	Dmonte et al. 2026	Initial outside scope	Selective unmasking includes an aspect span task. A linked sentiment output was not established.
ABSA-087	Zhou et al. 2026	Initial outside scope	Multilingual training leads to dimensional regression for provided aspects rather than joint prediction.
ABSA-088	Yang and Yang 2026	Initial outside scope	Valence and arousal are predicted for given aspects. The stance task is also outside the linked ABSA outcome.
ABSA-089	Fan et al. 2025	Initial outside scope	Prompted structured analysis uses zero-shot reasoning. A separate cross-language donor test was not established.
ABSA-090	Xu et al. 2025	Initial outside scope	Few-shot learning and label refinement address label scarcity. Distinct donor and target languages were not established.
ABSA-091	Shahana et al. 2025	Initial outside scope	English and Bengali evaluation classes are predicted for supplied gold aspects rather than jointly recovered structures.
ABSA-092	Rahman et al. 2025	Initial outside scope	Review-level and aspect-level classification use multilingual review data and back-translation. Joint directional transfer is unverified.
ABSA-093	Shen et al. 2025	Initial outside scope	The zero-shot condition changes domains. Distinct donor and target languages were not demonstrated.
ABSA-094	Vacareanu et al. 2024	Initial outside scope	Weak supervision and few-shot extraction provide a resource-scarcity comparison without documented directional transfer.
ABSA-095	Chen and Pan 2024	Initial outside scope	English-to-Chinese transfer predicts polarity for a supplied aspect. There is no jointly recovered output.
ABSA-096	Wang et al. 2024	Initial outside scope	The shift is between domains in triplet extraction. A controlled donor-to-target language comparison was not identified.
ABSA-097	Šmíd et al. 2024b	Initial outside scope	Structured evaluations cover multiple English datasets rather than an additional target language.
ABSA-098	Garg and Caragea 2024	Initial outside scope	Stance detection is the main task. The ABSA evaluation does not establish a directed linked extraction condition.
ABSA-099	Gou et al. 2023	Initial outside scope	Prompt ordering improves tuple prediction under resource scarcity. A distinct donor and target language were not identified.
ABSA-100	Hosseini-Asl et al. 2022	Initial outside scope	Few-shot generation covers multiple ABSA outputs. Directional language transfer was not established.

Record	Source citation	Disposition	Evidence and role in the review
ABSA-101	Mamta and Ekbal 2022	Initial outside scope	Adversarial perturbations are evaluated for aspect-conditioned polarity classification.
ABSA-102	Chen et al. 2022b	Initial outside scope	Polarity is predicted for an already supplied aspect. Opinion-tree induction is latent rather than a jointly scored output.
ABSA-103	Hu et al. 2022	Initial outside scope	Template order varies in quadruple prediction. The experiment does not isolate distinct donor and target languages.
ABSA-104	Sattar et al. 2021	Initial outside scope	Cross-language polarity classification conditions on a supplied aspect and informs the distinction between input and prediction fields.
ABSA-105	Bhattacharya et al. 2021	Initial outside scope	English translations support Hindi aspect extraction. Sentiment polarity is outside the output.
ABSA-106	Phan et al. 2021	Initial outside scope	Directed multilingual transfer is evaluated separately for aspect extraction and category detection.
ABSA-107	Wu et al. 2021	Initial outside scope	Cross-language distillation predicts polarity for a supplied aspect category.
ABSA-108	Jacqmin et al. 2021	Initial outside scope	Annotation projection supports opinion-target span extraction. The reported target outcome is a single extraction field.
ABSA-109	Bensoltane and Zaki 2022	Initial outside scope	Arabic aspect-term extraction and category detection are reported as separate component outcomes.
ABSA-110	Akhtar et al. 2019b	Initial outside scope	Multilingual aspect extraction and classification are evaluated separately. A joint sentiment-bearing score was not identified.
ABSA-111	Jebbara and Cimiano 2019	Initial outside scope	Directed cross-language opinion-target extraction recovers target spans without jointly scored polarity.
ABSA-112	Akhtar et al. 2019a	Initial outside scope	Cross-language embedding experiments report aspect extraction and polarity classification separately. Classification uses an aspect-centered context.
ABSA-113	Akhtar et al. 2018	Initial outside scope	Aspect extraction and polarity classification are separate experiments in multilingual data-sparsity analysis.
ABSA-114	Wang and Pan 2018	Initial outside scope	Directed transition-based aspect extraction predicts aspect terms without a linked sentiment output.

Publication-type notes for the search register

These three records are cited in Table 7 for publication-type screening. The register identifies two preprints and one master's thesis.

Cotfas RA (2023) Analysis of Cross-Lingual Models for Aspect-Based Sentiment Analysis: A Case Study in Dutch as a Low-Resource Language. Master's thesis, Tilburg University. https://arno.uvt.nl/show.cgi?fid=172664

Li X, Bing L, Zhang W, Li Z, Lam W (2020) Unsupervised Cross-lingual Adaptation for Sequence Tagging and Beyond. arXiv:2010.12405 [preprint]. https://doi.org/10.48550/arXiv.2010.12405

Nam K (2024) Korean Aspect-Based Sentiment Analysis via Implicit-Feature Alignment with Corpus Filtering. arXiv:2407.00342v1 [preprint]. https://doi.org/10.48550/arXiv.2407.00342 First posted version: https://arxiv.org/abs/2407.00342v1

Supplementary Information

Online Resource 1 Executed search strategies and dates, protocol decisions, deduplication, documented provisional review flow, and the 114-record screening register (ESM_1.pdf)

Online Resource 2 Machine-readable search and screening logs, extraction codebook, configuration-level evidence, outcome values, and comparison and verification records (ESM_2.xlsx)

Statements and Declarations

Funding

This work received no specific funding from any agency in the public, commercial, or not-for-profit sectors.

Competing interests

The authors declare no competing interests.

Author contributions

Botir Elov conceived the review, formulated the research questions, coordinated the work, and wrote the initial draft. Ruhillo Alaev contributed to the review methodology, helped develop the coding scheme for target-language resource use, and assessed the comparability of the reported performance scores. Zilola Xusainova curated the supplementary evidence tables and checked their consistency with the figures and the main text. All authors critically revised the manuscript and approved the final version.

Data availability

Online Resource 1 and Online Resource 2 accompany the article and document the search, working decisions and extracted results. Online Resource 2 supplies stable bibliographic, configuration and outcome identifiers, with links to the cited source material. Appendix A maps all search records to their working dispositions and bibliographic citations. A public persistent repository identifier for the compiled review files remains pending.

Code availability

The scripts used to prepare the review records and figures are retained in the review working files.

Ethics approval

Not applicable. This review examined published reports and associated materials and involved no new collection of participant data.

Consent to participate

Not applicable. No participants were recruited for this review.

Consent for publication

Not applicable. The review does not report identifiable participant information.

References

Affan M, Shahzad MH, Imam M, Zulfiqar M, Kumar S, Samad A (2026) Habib University at SemEval-2026 Task 3: A Pipeline Approach for Dimensional Aspect-Based Sentiment Analysis. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 3449–3459. https://doi.org/10.18653/v1/2026.semeval-1.428

Ahmad W, Khan HU, Alarfaj FK, Alreshoodi M (2025) Aspect-based sentiment analysis: a comprehensive review and open research challenges. IEEE Access 13:65138–65182. https://doi.org/10.1109/access.2025.3555744

Ahmed N, Aziz K, Tariq U, Hadi HJ, Alshara MA, Harris S (2025) Urdu Aspect-Category-Opinion-Sentiment (UACOS) Quadruple Extraction: A Transfer Learning Approach. In: 2025 4th International Conference on Communication, Computing and Digital Systems (C-CODE). IEEE, pp 1–6. https://doi.org/10.1109/c-code67372.2025.11204128

Akhtar MS, Sawant P, Sen S, Ekbal A, Bhattacharyya P (2018) Solving Data Sparsity for Aspect Based Sentiment Analysis Using Cross-Linguality and Multi-Linguality. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). Association for Computational Linguistics, pp 572–582. https://doi.org/10.18653/v1/N18-1053

Akhtar MS, Sawant P, Sen S, Ekbal A, Bhattacharyya P (2019a) Improving Word Embedding Coverage in Less-Resourced Languages Through Multi-Linguality and Cross-Linguality: A Case Study with Aspect-Based Sentiment Analysis. ACM Trans Asian Low-Resour Lang Inf Process 18(2):15. https://doi.org/10.1145/3273931

Akhtar MS, Kumar A, Ekbal A, Biemann C, Bhattacharyya P (2019b) Language-Agnostic Model for Aspect-Based Sentiment Analysis. In: Proceedings of the 13th International Conference on Computational Semantics – Long Papers. Association for Computational Linguistics, pp 154–164. https://doi.org/10.18653/v1/W19-0413

Bapanov A, Murzakhmetov A, Beisov N, Satymbekov M, Murotboyev M (2026) A Generative Aspect Sentiment Triplet Extraction Approach Using Fine-Tuned LLMs. Informatics 13(10):164. https://doi.org/10.3390/informatics13100164

Barnes J, Oberlaender L, Troiano E, Kutuzov A, Buchmann J, Agerri R, Øvrelid L, Velldal E (2022) SemEval 2022 Task 10: Structured Sentiment Analysis. In: Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). Association for Computational Linguistics, pp 1280–1295. https://doi.org/10.18653/v1/2022.semeval-1.180

Bender EM, Friedman B (2018) Data statements for natural language processing: toward mitigating system bias and enabling better science. Trans Assoc Comput Linguist 6:587–604. https://doi.org/10.1162/tacl_a_00041

Bensoltane R, Zaki T (2022) Towards Arabic aspect-based sentiment analysis: a transfer learning-based approach. Soc Netw Anal Min 12:7. https://doi.org/10.1007/s13278-021-00794-4

Bhattacharya A, Debnath A, Shrivastava M (2021) Enhancing Aspect Extraction for Hindi. In: Proceedings of the 4th Workshop on e-Commerce and NLP. Association for Computational Linguistics, pp 140–149. https://doi.org/10.18653/v1/2021.ecnlp-1.17

Cai H, Xia R, Yu J (2021) Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, pp 340–350. https://doi.org/10.18653/v1/2021.acl-long.29

Cao A, Hoang L, Toan LN, Linh H (2026) HUS@NLP-VNU at SemEval-2026 Task 3: Dual-Stream Syntax-Aware Modeling and Direct Preference Optimization for Dimensional ABSA. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 2200–2208. https://doi.org/10.18653/v1/2026.semeval-1.278

Chen C (2026) PALI at SemEval-2026 Task 3: LoRA Fine-Tuning with Validation for DimABSA. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 641–649. https://doi.org/10.18653/v1/2026.semeval-1.93

Chen J, Pan L (2024) An AI based cross-language aspect-level sentiment analysis model using English corpus. Eng Rep 6(12):e12969. https://doi.org/10.1002/eng2.12969

Chen C, Chen J, Liu C, Yang F, Wan G, Xia J (2022a) MT-Speech at SemEval-2022 Task 10: Incorporating Data Augmentation and Auxiliary Task with Cross-Lingual Pretrained Language Model for Structured Sentiment Analysis. In: Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). Association for Computational Linguistics, pp 1329–1335. https://doi.org/10.18653/v1/2022.semeval-1.185

Chen C, Teng Z, Wang Z, Zhang Y (2022b) Discrete Opinion Tree Induction for Aspect-based Sentiment Analysis. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, pp 2051–2064. https://doi.org/10.18653/v1/2022.acl-long.145

Chiril P, Spreadbury T, Rock JS, Dowd-Uribe B, Uminsky D (2025) Seeds of Discourse: A Multilingual Corpus of Direct Quotations from African Media on Agricultural Biotechnologies. In: Findings of the Association for Computational Linguistics: NAACL 2025. Association for Computational Linguistics, pp 8509–8515. https://doi.org/10.18653/v1/2025.findings-naacl.473

De Geyndt E, De Clercq O, Van Hee C, Lefever E, Singh P, Parent O, Hoste V (2022) SentEMO: A Multilingual Adaptive Platform for Aspect-based Sentiment and Emotion Analysis. In: Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis. Association for Computational Linguistics, pp 51–61. https://doi.org/10.18653/v1/2022.wassa-1.5

Dejaeghere T, Singh P, Lefever E, Birkholz J (2024) Exploring Aspect-Based Sentiment Analysis Methodologies for Literary-Historical Research Purposes. In: Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024. ELRA and ICCL, pp 129–143. https://doi.org/10.63317/3tzzsuwra4gn

Dewangan L, Maurya CK (2026a) CMF_Hit: Enhancing code-Mixed aspect-Based sentiment analysis via language-Aware gradient-Based tokenization and feature fusion. Expert Syst Appl 316:131639. https://doi.org/10.1016/j.eswa.2026.131639

Dewangan L, Maurya CK (2026b) CLAOCS-TX: cross-lingual triplet extraction with aspect-opinion-aware code-switched prompting and LLM-guided contrastive distillation. In: Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, pp 44561–44577. https://doi.org/10.18653/v1/2026.acl-long.2063

Dewangan L, Sayeed ZA, Maurya CK (2025) Benchmark Creation for Aspect-Based Sentiment Analysis in Low-Resource Odia Language and Evaluation through Fine-Tuning of Multilingual Models. In: Proceedings of the 31st International Conference on Computational Linguistics. Association for Computational Linguistics, pp 5863–5869. https://aclanthology.org/2025.coling-main.391/

Dharpure H, Rusnachenko N (2026) hdharpure at SemEval-2026 Task 3: BERT-Based Modeling and Prediction Behavior Analysis for Multilingual Valence–Arousal Scoring. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 2228–2232. https://doi.org/10.18653/v1/2026.semeval-1.281

Dmonte A, Ranasinghe T, Zampieri M (2026) SUNDER: Selective Unmasking for Text Understanding. In: Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp 19171–19188. https://doi.org/10.18653/2026.emnlp-main.825

Fan R, Li S, He T, Liu Y (2025) Aspect-Based Sentiment Analysis with Syntax-Opinion-Sentiment Reasoning Chain. In: Proceedings of the 31st International Conference on Computational Linguistics. Association for Computational Linguistics, pp 3123–3137. https://aclanthology.org/2025.coling-main.210/

Fatemian SH, Farhadi A, Zamanifar A (2026) Transfer learning for aspect based sentiment analysis. Discov Artif Intell 6:639. https://doi.org/10.1007/s44163-026-01153-y

Fehle J, Hellwig NC, Kruschwitz U, Wolff C (2026) Zero-shot to full-resource: cross-lingual transfer strategies for aspect-based sentiment analysis. In: Proceedings of the Fifteenth Language Resources and Evaluation Conference. ELRA Language Resource Association, pp 7999–8013. https://doi.org/10.63317/3fpqpgsdobd6

Frolov A, Rykov E (2026) ssurface3 at SemEval-2026 Task 3: Efficient Methods for Multilingual Dimensional Aspect-Based Sentiment Analysis. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 2910–2919. https://doi.org/10.18653/v1/2026.semeval-1.365

Galhom A, Abobakr A, Noseer M, Abdalgwad S, Nour R, Fares A (2025) AASTE: Arabic Aspect Sentiment Triplet Extraction. In: Proceedings of the 2025 7th Novel Intelligent and Leading Emerging Sciences Conference (NILES). IEEE, pp 453–456. https://doi.org/10.1109/niles68063.2025.11232333

Garg K, Caragea C (2024) Stanceformer: Target-Aware Transformer for Stance Detection. In: Findings of the Association for Computational Linguistics: EMNLP 2024. Association for Computational Linguistics, pp 4969–4984. https://doi.org/10.18653/v1/2024.findings-emnlp.286

Gazetas S, Filandrianos G, Lymperaiou M, Tzouveli P, Voulodimos A, Stamou G (2026) AILS-NTUA at SemEval-2026 Task 3: efficient dimensional aspect-based sentiment analysis. In: Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026). Association for Computational Linguistics, pp 1212–1227. https://doi.org/10.18653/v1/2026.semeval-1.163

Gheorghe AA, Andrei C, Ionescu E, Ruseti S, Dascalu M (2026) RO-ABSA: A Romanian Dataset and Baselines for Aspect-Based Sentiment Analysis. In: Proceedings of the Fifteenth Language Resources and Evaluation Conference. ELRA Language Resource Association, pp 6373–6382. https://doi.org/10.63317/3y4i7u4c78n5

Gimadi D, Evans R, Simov K (2021) Web-sentiment analysis of public comments (public reviews) for languages with limited resources such as the Kazakh language. In: Proceedings of the Student Research Workshop Associated with RANLP 2021. INCOMA Ltd, pp 65–68. https://doi.org/10.26615/issn.2603-2821.2021_010

Gou Z, Guo Q, Yang Y (2023) MvP: Multi-view Prompting Improves Aspect Sentiment Tuple Prediction. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, pp 4380–4397. https://doi.org/10.18653/v1/2023.acl-long.240

Hammad M, AbuEnnab N, Al-Refai M (2025) An aspect-based sentiment analysis model for Arabic game reviews based on hybrid transformers models. Neural Comput Appl 37:10309–10331. https://doi.org/10.1007/s00521-025-11032-9

Haralovic M, Akca OD, Yücetürk SE, Li M, Theune M (2026) CrisisKD: Five-Stage Knowledge Distillation for Aspect-Level Sentiment and Emotion Analysis in Crisis Discourse. In: Proceedings of the 11th Workshop on Natural User-generated Text. Association for Computational Linguistics, pp 49–64. https://doi.org/10.18653/2026.wnut-1.4

He Q, Zhou X (2026) YNU-ABSA at SemEval-2026 Task 3: A Unified Pipeline for Continuous and Structured Dimensional ABSA. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 1100–1105. https://doi.org/10.18653/v1/2026.semeval-1.150

Hellwig NC, Fehle J, Kruschwitz U, Wolff C (2026a) LLM-as-an-Annotator: Training Lightweight Models with LLM-Annotated Examples for Aspect Sentiment Tuple Prediction. In: Proceedings of the Fifteenth Language Resources and Evaluation Conference. European Language Resources Association, pp 7955–7972. https://doi.org/10.63317/43srcdyc52cd

Hellwig NC, Fehle J, Kruschwitz U, Wolff C (2026b) nchellwig at SemEval-2026 Task 3: Self-Consistent Structured Generation (SCSG) for Dimensional Aspect-Based Sentiment Analysis using Large Language Models. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 37–47. https://doi.org/10.18653/v1/2026.semeval-1.6

Hosseini-Asl E, Liu W, Xiong C (2022) A Generative Language Model for Few-shot Aspect-Based Sentiment Analysis. In: Findings of the Association for Computational Linguistics: NAACL 2022. Association for Computational Linguistics, pp 770–787. https://doi.org/10.18653/v1/2022.findings-naacl.58

Hu S (2026) kirito at SemEval-2026 Task 3: Dimensional Aspect-Based Sentiment Analysis via Sentence Structure Parsing Preprocessing and Prompt-Enhanced Instruction Tuning. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 913–918. https://doi.org/10.18653/v1/2026.semeval-1.125

Hu M, Wu Y, Gao H, Bai Y, Zhao S (2022) Improving Aspect Sentiment Quad Prediction via Template-Order Data Augmentation. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp 7889–7900. https://doi.org/10.18653/v1/2022.emnlp-main.538

Hua YC, Denny P, Wicker J, Taskova K (2024) A systematic review of aspect-based sentiment analysis: domains, methods, and trends. Artif Intell Rev 57:296. https://doi.org/10.1007/s10462-024-10906-z

Jacqmin L, Marzinotto G, Gromada J, Szczekocka E, Kołodyński R, Chauvière A, Damnati G (2021) SpanAlign: Efficient Sequence Tagging Annotation Projection into Translated Data applied to Cross-Lingual Opinion Mining. In: Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021). Association for Computational Linguistics, pp 238–248. https://doi.org/10.18653/v1/2021.wnut-1.27

Jebbara S, Cimiano P (2019) Zero-Shot Cross-Lingual Opinion Target Extraction. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, pp 2486–2495. https://doi.org/10.18653/v1/N19-1257

Klinger R, Cimiano P (2015) Instance Selection Improves Cross-Lingual Model Training for Fine-Grained Sentiment Analysis. In: Proceedings of the Nineteenth Conference on Computational Natural Language Learning. Association for Computational Linguistics, pp 153–163. https://doi.org/10.18653/v1/K15-1016

Lango M, Naglik B, Lango M, Naglik I (2024) Polish-ASTE: Aspect-Sentiment Triplet Extraction Datasets for Polish. In: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). ELRA and ICCL, pp 12821–12828. https://doi.org/10.63317/32nv97efm2tr

Lee LH, Yu LC, Loukachevitch NV, Alimova I, Panchenko A, Lin TM, Xu ZY, Zhou JY, Zheng G, Wang J, Awasthi S, Becker J, Wahle JP, Ruas T, Muhammad SH, Mohammad SM (2026a) DimABSA: Building Multilingual and Multidomain Datasets for Dimensional Aspect-Based Sentiment Analysis. In: Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, pp 40492–40518. https://doi.org/10.18653/v1/2026.acl-long.1881

Lee CY, Pleva M, Hladek D, Su MH (2026b) SCU_Mesclab at SemEval-2026 Task 3: An Adaptive Dual-Track Framework for Dimensional Aspect-Based Sentiment Analysis. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 1482–1488. https://doi.org/10.18653/v1/2026.semeval-1.192

Li J, Yang Y (2026) hllwan at SemEval-2026 Task 3: Dimensional Aspect-Based Sentiment Analysis via LLM Feature Fusion and Test-Time Adaptation. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 1147–1152. https://doi.org/10.18653/v1/2026.semeval-1.157

Li X, Zhang K (2025) Contrastive Learning Pre-Training and Quantum Theory for Cross-Lingual Aspect-Based Sentiment Analysis. Entropy 27(7):713. https://doi.org/10.3390/e27070713

Lin Z, Jin X, Xu X, Wang W, Cheng X, Wang Y (2014) A Cross-Lingual Joint Aspect/Sentiment Model for Sentiment Analysis. In: Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM). ACM, pp 1089–1098. https://doi.org/10.1145/2661829.2662019

Lin N, Fu Y, Lin X, Zhou D, Yang A, Jiang S (2023) CL-XABSA: Contrastive Learning for Cross-Lingual Aspect-Based Sentiment Analysis. IEEE/ACM Trans Audio Speech Lang Process 31:2935–2946. https://doi.org/10.1109/taslp.2023.3297964

Lin N, Zeng M, Liao X, Liu W, Yang A, Zhou D (2024) Addressing class-imbalance challenges in cross-lingual aspect-based sentiment analysis: Dynamic weighted loss and anti-decoupling. Expert Syst Appl 257:125059. https://doi.org/10.1016/j.eswa.2024.125059

Liu Y, Li F, Ji D (2024) Improving cross-lingual aspect-based sentiment analysis with sememe bridge. ACM Trans Asian Low-Resour Lang Inf Process 23(12):166. https://doi.org/10.1145/3691342

Lu HY, Jiang YF, Liu TC, Fang W, Wang C (2026) RePD: Reliable Pseudo-label Construction and Dynamic Inference for Transductive Zero-shot Aspect-based Sentiment Quadruple Prediction. IEEE Trans Affect Comput 1–15. https://doi.org/10.1109/TAFFC.2026.3733678

Luo M, Fei H, Li B, Wu S, Liu Q, Poria S, Cambria E, Lee ML, Hsu W (2024) PanoSent: A Panoptic Sextuple Extraction Benchmark for Multimodal Conversational Aspect-based Sentiment Analysis. In: Proceedings of the 32nd ACM International Conference on Multimedia. Association for Computing Machinery, pp 7667–7676. https://doi.org/10.1145/3664647.3680705

Mamta, Ekbal A (2022) Adversarial Sample Generation for Aspect based Sentiment Classification. In: Findings of the Association for Computational Linguistics: AACL-IJCNLP 2022. Association for Computational Linguistics, pp 478–492. https://doi.org/10.18653/v1/2022.findings-aacl.44

Matlatipov SG, Rajabov J, Kuriyozov E, Aripov M (2024) UzABSA: Aspect-Based Sentiment Analysis for the Uzbek Language. In: Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages @ LREC-COLING 2024. ELRA and ICCL, pp 394–403. https://doi.org/10.63317/2sriwt3ca358

Nawawi I, Ilmawan KF, Maarif MR, Syafrudin M (2024) Exploring Tourist Experience through Online Reviews Using Aspect-Based Sentiment Analysis with Zero-Shot Learning for Hospitality Service Enhancement. Information 15(8):499. https://doi.org/10.3390/info15080499

Nour R, Abdalgwad S, Noseer M, Abobakr A, Galhom A, Fares A (2025) Evaluating Large Language Models for Arabic Aspect Sentiment Triplet Extraction using Zero, One, and Few-Shot Prompting. In: 2025 13th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC). IEEE, pp 327–330. https://doi.org/10.1109/jac-ecc67970.2025.11417612

Page MJ, McKenzie JE, Bossuyt PM et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372:n71. https://doi.org/10.1136/bmj.n71

Peng H, Xu L, Bing L, Huang F, Lu W, Si L (2020) Knowing what, how and why: a near complete solution for aspect-based sentiment analysis. Proc AAAI Conf Artif Intell 34:8600–8607. https://doi.org/10.1609/aaai.v34i05.6383

Pfister J, Wankerl S, Hotho A (2022) SenPoi at SemEval-2022 Task 10: Point me to your Opinion, SenPoi. In: Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022). Association for Computational Linguistics, pp 1313–1323. https://doi.org/10.18653/v1/2022.semeval-1.183

Phan KTK, Ngoc Hao D, Thin DV, Luu-Thuy Nguyen N (2021) Exploring Zero-shot Cross-lingual Aspect-based Sentiment Analysis using Pre-trained Multilingual Language Models. In: 2021 International Conference on Multimedia Analysis and Pattern Recognition (MAPR). IEEE, pp 1–6. https://doi.org/10.1109/MAPR53640.2021.9585242

Pontiki M, Galanis D, Papageorgiou H et al (2016) SemEval-2016 Task 5: aspect based sentiment analysis. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). Association for Computational Linguistics, pp 19–30. https://doi.org/10.18653/v1/S16-1002

Rahman A, Khan MA, Bishnu KK, Rozario U, Ishraq A, Mridha MF, Aung Z (2025) Multilingual sentiment analysis in restaurant reviews using aspect focused learning. Sci Rep 15:28371. https://doi.org/10.1038/s41598-025-12464-y

Regatte YR, Gangula RRR, Mamidi R (2020) Dataset Creation and Evaluation of Aspect Based Sentiment Analysis in Telugu, a Low Resource Language. In: Proceedings of the Twelfth Language Resources and Evaluation Conference. European Language Resources Association, pp 5017–5024. https://doi.org/10.63317/5929697gfgk9

Rethlefsen ML, Kirtley S, Waffenschmidt S et al (2021) PRISMA-S: an extension to the PRISMA Statement for reporting literature searches in systematic reviews. Syst Rev 10:39. https://doi.org/10.1186/s13643-020-01542-z

Rink L, Meijdam J, Graus D (2024) Aspect-Based Sentiment Analysis for Open-Ended HR Survey Responses. In: Proceedings of the First Workshop on Natural Language Processing for Human Resources (NLP4HR 2024). Association for Computational Linguistics, pp 16–26. https://doi.org/10.18653/v1/2024.nlp4hr-1.2

Ruder S, Ghaffari P, Breslin JG (2016) INSIGHT-1 at SemEval-2016 Task 5: Deep Learning for Multilingual Aspect-based Sentiment Analysis. In: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). Association for Computational Linguistics, pp 330–336. https://doi.org/10.18653/v1/S16-1053

S JM, S AR (2026) Pixel Phantoms at SemEval-2026 Task 3: Language-Specific Transformer Regression for Dimensional Aspect-Based Sentiment Analysis. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 803–810. https://doi.org/10.18653/v1/2026.semeval-1.112

Santin RV, Rezende SO (2026) Systematic review on aspect-based sentiment analysis in cross-domain. Artif Intell Rev 59:37. https://doi.org/10.1007/s10462-025-11437-x

Satria GP, Gunawan KW, Rasjid ZE (2025) Aspect-Based Sentiment Analysis of Multilingual Customer Product Reviews: Leveraging Unsupervised Aspect Extraction with ABAE for XLM-RoBERTa-based Classification. Procedia Comput Sci 269:239–248. https://doi.org/10.1016/j.procs.2025.08.276

Sattar K, Umer Q, Vasbieva DG, Chung S, Latif Z, Lee C (2021) A Multi-Layer Network for Aspect-Based Cross-Lingual Sentiment Classification. IEEE Access 9:133961–133973. https://doi.org/10.1109/ACCESS.2021.3116053

Shahana SA, Afrin NN, Anwar MM, Jahan I (2025) Gen-mABSA-T5: A Multilingual Zero-Shot Generative Framework for Aspect-Based Sentiment Analysis. In: Proceedings of the Second Workshop on Bangla Language Processing (BLP-2025). Association for Computational Linguistics, pp 142–150. https://doi.org/10.18653/v1/2025.banglalp-1.12

Shen C, Wei W, Wang D, Wang ZH (2025) Zero-Shot Cross-Domain Aspect-Based Sentiment Analysis via Domain-Contextualized Chain-of-Thought Reasoning. In: Findings of the Association for Computational Linguistics: EMNLP 2025. Association for Computational Linguistics, pp 4558–4573. https://doi.org/10.18653/v1/2025.findings-emnlp.245

Sheng D, Han K, Li H, Zhang Y, Huang Y, Lang J, Liu W (2025) Test-time code-switching for cross-lingual aspect sentiment triplet extraction. In: Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers). Association for Computational Linguistics, pp 5041–5053. https://doi.org/10.18653/v1/2025.naacl-long.260

Šmíd J (2023) Cross-lingual Aspect-Based Sentiment Analysis. In: Student Scientific Conference 2023, pp 66–67. https://naos-be.zcu.cz/server/api/core/bitstreams/3311962d-f076-409b-8368-6fc4bf1e1d6f/content

Šmíd J, Král P (2025) Cross-lingual aspect-based sentiment analysis: a survey on tasks, approaches, and challenges. Inf Fusion 120:103073. https://doi.org/10.1016/j.inffus.2025.103073

Šmíd J, Přibáň P (2023) Prompt-Based Approach for Czech Sentiment Analysis. In: Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing. INCOMA Ltd., Shoumen, Bulgaria, pp 1110–1120. https://doi.org/10.26615/978-954-452-092-2_118

Šmíd J, Přibáň P, Pražák O, Král P (2024a) Czech Dataset for Complex Aspect-Based Sentiment Analysis Tasks. In: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). ELRA and ICCL, pp 4299–4310. https://doi.org/10.63317/2atzhbwhyw3z

Šmíd J, Přibáň P, Král P (2024b) LLaMA-Based Models for Aspect-Based Sentiment Analysis. In: Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis. Association for Computational Linguistics, pp 63–70. https://doi.org/10.18653/v1/2024.wassa-1.6

Šmíd J, Přibáň P, Král P (2025a) Advancing cross-lingual aspect-based sentiment analysis with LLMs and constrained decoding for sequence-to-sequence models. In: Proceedings of the 17th International Conference on Agents and Artificial Intelligence, vol 2. SciTePress, pp 757–766. https://doi.org/10.5220/0013349400003890

Šmíd J, Přibáň P, Král P (2025b) LACA: improving cross-lingual aspect-based sentiment analysis with LLM data augmentation. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, pp 839–853. https://doi.org/10.18653/v1/2025.acl-long.41

Šmíd J, Přibáň P, Král P (2026a) Extending Czech aspect-based sentiment analysis with opinion terms: dataset and LLM benchmarks. In: Proceedings of the Fifteenth Language Resources and Evaluation Conference. European Language Resources Association, pp 7973–7984. https://doi.org/10.63317/4hkzdnwfztkz

Šmíd J, Přibáň P, Král P (2026b) Generative models enhanced by sequence labelling and aspect-code switching improve cross-lingual aspect-based sentiment analysis. In: Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp 34966–34981. https://doi.org/10.18653/2026.emnlp-main.1531

Šmíd J, Přibáň P, Král P (2026c) Few-Shot Cross-Lingual Aspect-Based Sentiment Analysis with Sequence-to-Sequence Models. In: Text, Speech, and Dialogue. TSD 2025. Lecture Notes in Artificial Intelligence, vol 16030. Springer, pp 27–38. https://doi.org/10.1007/978-3-032-02551-7_4

Šmíd J, Přibáň P, Král P (2026d) Improving Generative Cross-Lingual Aspect-Based Sentiment Analysis with Constrained Decoding. In: Agents and Artificial Intelligence. ICAART 2025 revised selected papers. Lecture Notes in Artificial Intelligence, vol 16518. Springer, pp 419–446. https://doi.org/10.1007/978-3-032-25035-3_20

Strothe L (2026) TeamLasse at SemEval-2026 Task 3: A Hybrid Generative-Discriminative Framework for Dimensional Aspect-Based Sentiment Analysis. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 2155–2162. https://doi.org/10.18653/v1/2026.semeval-1.273

Suryabrata AW, Warnars HLHS, Muyeba MK (2026) Aspect-Based Sentiment Analysis of Multilingual Hotel Reviews in Jakarta Using Multilingual Transformer Models: A Comparative Study of Multilingual BERT and Cross-Lingual RoBERTa. Ingénierie des Systèmes d'Information 31(1):123–135. https://doi.org/10.18280/isi.310112

Szołomicka J, Kocoń J (2022) MultiAspectEmo: multilingual and language-agnostic aspect-based sentiment analysis. In: 2022 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, pp 443–450. https://doi.org/10.1109/ICDMW58026.2022.00065

Tu LH, Nguyen Q, Nguyen V, Björklund J, Vu XS (2026) Reason-to-Learn (R2L): Multi-Agent Knowledge Distillation for Lightweight LLMs in Sentiment Analysis. In: Proceedings of the Fifteenth Language Resources and Evaluation Conference. ELRA Language Resource Association, pp 10301–10312. https://doi.org/10.63317/3aygmawej3my

Vacareanu R, Varia S, Halder K, Wang S, Paolini G, Anna John N, Ballesteros M, Muresan S (2024) A Weak Supervision Approach for Few-Shot Aspect Based Sentiment Analysis. In: Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, pp 2734–2752. https://doi.org/10.18653/v1/2024.eacl-long.167

Van Thin D, Quoc Ngo H, Ngoc Hao D, Luu-Thuy Nguyen N (2023) Exploring zero-shot and joint training cross-lingual strategies for aspect-based sentiment analysis based on contextualized multilingual language models. J Inf Telecommun 7(2):121–143. https://doi.org/10.1080/24751839.2023.2173843

Wan H, Yang Y, Du J, Liu Y, Qi K, Pan JZ (2020) Target-aspect-sentiment joint detection for aspect-based sentiment analysis. Proc AAAI Conf Artif Intell 34:9122–9129. https://doi.org/10.1609/aaai.v34i05.6447

Wang W, Pan SJ (2018) Transition-Based Adversarial Network for Cross-Lingual Aspect Extraction. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. IJCAI, pp 4475–4481. https://doi.org/10.24963/ijcai.2018/622

Wang G, Wang B, Xu F, Wang R, Zhu Z, Liu P (2024) Domain-consistent syntactic representation for cross-domain aspect sentiment triplet extraction. Expert Syst Appl 256:124854. https://doi.org/10.1016/j.eswa.2024.124854

Wang Y, Jiang R, Wang Z, Zhou G (2026) Exploring multilingual pre-trained language model for aspect-based sentiment analysis. In: Findings of the Association for Computational Linguistics: ACL 2026. Association for Computational Linguistics, pp 6011–6023. https://doi.org/10.18653/v1/2026.findings-acl.298

Wu H, Wang Z, Qing F, Li S (2021) Reinforced Transformer with Cross-Lingual Distillation for Cross-Lingual Aspect Sentiment Classification. Electronics 10(3):270. https://doi.org/10.3390/electronics10030270

Wu C, Ma B, Zhang Z, Deng N, He Y, Xue Y (2025a) Evaluating zero-shot multilingual aspect-based sentiment analysis with large language models. Int J Mach Learn Cybern 16:8079–8101. https://doi.org/10.1007/s13042-025-02711-z

Wu C, Ma B, Liu Y, Zhang Z, Deng N, Li Y, Chen B, Zhang Y, Xue Y, Plank B (2025b) M-ABSA: a multilingual dataset for aspect-based sentiment analysis. In: Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp 2530–2557. https://doi.org/10.18653/v1/2025.emnlp-main.128

Wu C, Ma B, Deng N, He Y, Xue Y, Liu X (2026a) MSMO-ABSA: multi-scale and multi-objective optimization for cross-lingual aspect-based sentiment analysis. In: Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, pp 18723–18735. https://doi.org/10.18653/v1/2026.acl-long.851

Wu T, Rusnachenko N, Liang H (2026b) NCL-BU at SemEval-2026 Task 3: Fine-tuning XLM-RoBERTa for Multilingual Dimensional Sentiment Regression. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 1911–1918. https://doi.org/10.18653/v1/2026.semeval-1.240

Xu H, Zhang Y, Wang Q, Xu R (2025) DS²-ABSA: Dual-Stream Data Synthesis with Label Refinement for Few-Shot Aspect-Based Sentiment Analysis. In: Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, pp 15460–15478. https://doi.org/10.18653/v1/2025.acl-long.752

Yang TH, Yang SF (2026) YangSteam at SemEval-2026 Task 3: Transformer-Based Aspect-Aware Regression for Dimensional Sentiment and Stance Analysis. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 1533–1538. https://doi.org/10.18653/v1/2026.semeval-1.199

Yang H, Zeng B, Yang J, Song Y, Xu R (2021) A multi-task learning model for Chinese-oriented aspect polarity classification and aspect term extraction. Neurocomputing 419:344–356. https://doi.org/10.1016/j.neucom.2020.08.001

Yu LC, Becker J, Muhammad SH et al (2026) SemEval-2026 Task 3: dimensional aspect-based sentiment analysis (DimABSA). In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 3753–3778. https://doi.org/10.18653/v1/2026.semeval-1.452

Yusufu A, Liu J, Aziz K et al (2026) LASQ: a low-resource aspect-based sentiment quadruple extraction dataset. J King Saud Univ Comput Inf Sci 38:337. https://doi.org/10.1007/s44443-026-00701-x

Zeng M, Liao X, Zhou Y, Huang W, Lin N, Yang A (2025) Enhancing Cross-Lingual Aspect-Based Sentiment Analysis with Code-Mixed In-Context Demonstrations and Language-Specific Tags. In: 2025 IEEE 28th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, pp 716–721. https://doi.org/10.1109/cscwd64889.2025.11033355

Zhang W, He R, Peng H, Bing L, Lam W (2021) Cross-lingual aspect-based sentiment analysis with aspect term code-switching. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp 9220–9230. https://doi.org/10.18653/v1/2021.emnlp-main.727

Zhang H, Cheah YN, Alyasiri OM, An J (2024) Exploring aspect-based sentiment quadruple extraction with implicit aspects, opinions, and ChatGPT: a comprehensive survey. Artif Intell Rev 57:17. https://doi.org/10.1007/s10462-023-10633-x

Zhang B, Zhao Y, Jin G, Cui R (2025) Cross-Lingual Aspect-Based Sentiment Analysis Based on Semi-supervised Knowledge Distillation. In: 2025 IEEE 5th International Conference on Electronic Technology, Communication and Information (ICETCI). IEEE, pp 57–61. https://doi.org/10.1109/icetci64844.2025.11084138

Zhao X, Wan H, Qi K (2024) QPEN: quantum projection and quantum entanglement enhanced network for cross-lingual aspect-based sentiment analysis. Proc AAAI Conf Artif Intell 38(17):19670–19678. https://doi.org/10.1609/aaai.v38i17.29940

Zhao W, Yang Z, Yu S, Zhu S, Li L (2025) Contrastive pre-training and instruction tuning for cross-lingual aspect-based sentiment analysis. Appl Intell 55:358. https://doi.org/10.1007/s10489-025-06251-5

Zhou Y, Wang W, Wang S, Bao M, Fang R, Song S, Li Y, Li X (2026) TeleAI at SemEval-2026 Task 3: Large Language Models for Dimensional Aspect-Based Sentiment Analysis. In: Proceedings of the 20th International Workshop on Semantic Evaluation (2026). Association for Computational Linguistics, pp 1846–1852. https://doi.org/10.18653/v1/2026.semeval-1.233

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM2.xlsx
- ESM1.pdf