
ANALISIS KESALAHAN TERJEMAHAN *ARTIFICIAL INTELLIGENCE* PADA TEKS BISNIS JEPANG-INDONESIA MENGGUNAKAN PENDEKATAN *TRANSLATION QUALITY ASSESSMENT* BERBASIS *MULTIDIMENSIONAL QUALITY METRICS (MQM)*

Oleh

Diwana Fikri Aghniya¹, Muhammad Irsan²

^{1,2}Politeknik Takumi, Kab. Bekasi, Jawa Barat

Email: ¹diwana.dfa@takumi.ac.id, ²irsan.mir@takumi.ac.id

Article History:

Received: 22-05-2026

Revised: 19-06-2027

Accepted: 25-06-2026

Keywords:

Artificial Intelligence; Machine Translation; Japanese Business Language; Keigo; Multidimensional Quality Metrics

Abstract: *The advancement of Artificial Intelligence (AI) technology has increased the use of automatic translation systems in cross-language communication, including Japanese-Indonesian business translation. The characteristics of Japanese business language, which involve honorific expressions (keigo), indirect communication, professional terminology, and cultural conventions, remain challenging for AI-based translation systems. This study aims to identify the types of AI translation errors, classify them using a Translation Quality Assessment (TQA) approach based on Multidimensional Quality Metrics (MQM), and analyze the factors contributing to these errors. This research employed a descriptive qualitative method using translation error analysis. The corpus consisted of 50 Japanese business texts translated into Indonesian using three AI translation systems-Google Translate, DeepL, and ChatGPT-resulting in 150 translation units. The data were analyzed using six MQM categories: Accuracy, Terminology, Fluency, Grammar, Style, and Locale Convention. The findings revealed 90 translation errors among the 150 units analyzed. The dominant error category was Style/Keigo, with 45 cases (50%), followed by Accuracy with 23 cases (25.5%), Fluency with 9 cases (10%), Locale Convention with 8 cases (8.9%), and Terminology with 5 cases (5.6%), while the Grammar category showed no significant errors. These findings indicate that although AI translation systems demonstrate strong linguistic and grammatical capability, they continue to face challenges in interpreting pragmatic meaning, politeness levels, and the cultural values embedded in Japanese business communication.*

PENDAHULUAN

Perkembangan teknologi *Artificial Intelligence* (AI) telah membawa perubahan signifikan dalam bidang penerjemahan. Sistem penerjemahan berbasis AI seperti *Google Translate*, *DeepL*, dan *ChatGPT* semakin banyak digunakan karena mampu menghasilkan terjemahan secara cepat dengan kualitas linguistik yang terus meningkat. Perkembangan *Neural Machine Translation* (NMT) memungkinkan mesin memahami hubungan antarbahasa

melalui jaringan *neural* sehingga menghasilkan terjemahan yang lebih natural dibandingkan sistem penerjemahan berbasis aturan (*rule-based translation*) maupun statistik (Bahdanau et al., 2015; Koehn, 2020). Arsitektur *Transformer* yang diperkenalkan oleh Vaswani et al. (2017) menjadi fondasi teknis bagi hampir seluruh sistem NMT mutakhir, termasuk model-model bahasa berskala besar saat ini. Selain itu, kemunculan teknologi *Large Language Model* (LLM) memperluas kemampuan AI dalam memahami konteks dan menghasilkan teks berbasis bahasa alami (Brown et al., 2020), sehingga turut mendorong penggunaan model generatif seperti *ChatGPT* sebagai alat penerjemahan, di samping fungsi awalnya sebagai sistem percakapan (Jiao et al., 2023).

Meskipun kemampuan linguistik sistem AI terus mengalami peningkatan, kualitas penerjemahan masih menjadi persoalan penting terutama pada bahasa dengan karakteristik sosial-budaya yang kompleks, seperti bahasa Jepang. Bahasa Jepang bisnis (*bijinesu nihongo*) tidak semata-mata berfungsi menyampaikan informasi, tetapi juga menegosiasikan hubungan sosial antara pembicara dan lawan bicara melalui sistem *keigo*. Brown dan Levinson (1987) menjelaskan bahwa kesantunan berbahasa pada dasarnya merupakan strategi penjagaan muka (*face*) dalam interaksi sosial, dan strategi tersebut terealisasi secara khas pada setiap bahasa, termasuk melalui honorifik dalam bahasa Jepang (Fukada & Asato, 2004). Dengan demikian, *keigo* tidak hanya memiliki muatan gramatikal, tetapi juga mengandung nilai penghormatan dan hierarki sosial yang menjadi bagian inheren dari makna tuturan. Kesalahan penerjemahan pada aspek tersebut berpotensi menghilangkan fungsi interpersonal yang justru menjadi inti komunikasi bisnis Jepang.

Konsekuensinya, evaluasi kualitas penerjemahan AI tidak dapat dilakukan hanya dengan menilai benar atau salah secara gramatikal, melainkan memerlukan kerangka evaluasi yang sistematis dan multidimensi. Salah satu pendekatan yang relevan digunakan adalah *Translation Quality Assessment* (TQA) berbasis *Multidimensional Quality Metrics* (MQM). Kerangka MQM memungkinkan identifikasi kesalahan penerjemahan berdasarkan dimensi-dimensi spesifik, seperti akurasi makna, terminologi, kewajaran bahasa, tata bahasa, register/gaya, dan kesesuaian dengan konvensi budaya sasaran (Lommel et al., 2014). Pendekatan ini sejalan dengan model TQA yang dikembangkan House (2015) maupun model penilaian kualitas terjemahan yang diadaptasi dalam konteks Indonesia oleh Nababan dkk. (2012), serta selaras dengan prinsip kesepadanan dinamis yang dikemukakan Nida dan Taber (1982) bahwa keberhasilan penerjemahan idealnya diukur dari kesepadanan respons pembaca sasaran, bukan semata kesepadanan kata demi kata.

Studi terdahulu mengenai performa AI dalam penerjemahan menunjukkan hasil yang beragam. Castilho dkk. (2017) mendapati bahwa NMT secara umum unggul dibandingkan pendekatan statistik dalam hal kelancaran (*fluency*), tetapi masih menyisakan persoalan pada akurasi semantik untuk pasangan bahasa dengan struktur yang jauh berbeda. Studi yang lebih baru oleh Jiao dkk. (2023) terhadap *ChatGPT* sebagai mesin penerjemah menemukan bahwa model tersebut kompetitif untuk pasangan bahasa berdaya sumber tinggi (*high-resource*), tetapi performanya menurun pada bahasa dengan jarak tipologis yang lebih jauh-sebuah kondisi yang relevan dengan pasangan bahasa Jepang-Indonesia mengingat perbedaan struktur dan konvensi sosiopragmatik yang signifikan di antara keduanya. Berdasarkan pemetaan kondisi tersebut, masih terdapat kesenjangan penelitian (*research gap*) yang spesifik mengenai bagaimana ketiga sistem AI populer tersebut

menangani unsur *keigo* dan pragmatik bisnis Jepang secara komparatif, khususnya dalam konteks penerjemahan ke bahasa Indonesia.

Berdasarkan uraian tersebut, penelitian ini disusun untuk menjawab tiga pertanyaan penelitian: (1) Apa saja jenis kesalahan penerjemahan AI yang muncul pada teks bisnis Jepang-Indonesia? (2) Bagaimana klasifikasi kesalahan tersebut apabila dianalisis menggunakan kerangka MQM? (3) Faktor-faktor apa yang menyebabkan munculnya kesalahan penerjemahan AI tersebut? Penelitian ini diharapkan memberikan kontribusi baik secara teoretis, dengan memperkaya kajian translation quality assessment pada pasangan bahasa Jepang-Indonesia, maupun secara praktis, sebagai pertimbangan bagi pengguna dan penerjemah profesional dalam memanfaatkan AI translation secara lebih bijak pada konteks komunikasi bisnis formal.

TINJAUAN PUSTAKA

2.1 Artificial Intelligence Translation

Penerjemahan otomatis mengalami pergeseran paradigma dari *Statistical Machine Translation* (SMT) menuju *Neural Machine Translation* (NMT). NMT memanfaatkan jaringan neural untuk memprediksi kemungkinan terjemahan berdasarkan konteks keseluruhan kalimat, bukan sekadar kecocokan frasa secara statistik (Bahdanau et al., 2015). Pendekatan *encoder-decoder* dengan mekanisme atensi yang diusulkan Bahdanau dkk. (2015) kemudian disempurnakan melalui arsitektur Transformer (Vaswani et al., 2017), yang menjadi tulang punggung teknis hampir seluruh sistem penerjemahan neural kontemporer, termasuk *DeepL* dan *Google Neural Machine Translation*.

Berbeda dengan pendekatan NMT konvensional, teknologi berbasis LLM seperti ChatGPT menggunakan model bahasa berskala besar yang dilatih untuk memprediksi kelanjutan teks berdasarkan pola linguistik dan konteks yang luas (Brown et al., 2020). Kapabilitas ini memungkinkan ChatGPT menghasilkan terjemahan yang mempertimbangkan konteks wacana secara lebih fleksibel dibandingkan sistem NMT khusus (*dedicated NMT system*). Jiao dkk. (2023) menunjukkan bahwa performa ChatGPT sebagai penerjemah kompetitif dengan sistem komersial pada pasangan bahasa berdaya sumber tinggi, meskipun cenderung kurang stabil pada bahasa dengan jarak tipologis yang signifikan. Namun demikian, baik sistem NMT maupun LLM sama-sama menghadapi tantangan dalam menerjemahkan unsur pragmatik, sebab makna bahasa tidak semata bersumber dari kata dan struktur kalimat, melainkan juga dari hubungan sosial dan konteks budaya penuturnya (Castilho et al., 2017; House, 2015).

2.2 Bahasa Jepang Bisnis dan Keigo

Bahasa Jepang bisnis memiliki karakteristik utama berupa penggunaan *keigo*, yaitu sistem honorifik yang mengatur tingkat kesantunan tuturan berdasarkan relasi sosial antarpenerjemah. *Keigo* terdiri atas tiga kategori utama. *Sonkeigo*, yaitu bahasa penghormatan yang digunakan untuk meninggikan tindakan atau keadaan lawan bicara. *Kenjougo*, yaitu bahasa merendahkan diri yang digunakan penutur untuk menghormati lawan bicara secara tidak langsung. *Teineigo*, yaitu bentuk bahasa sopan baku yang menunjukkan kesantunan umum dalam interaksi formal.

Ide (1989) menjelaskan bahwa kesantunan dalam bahasa Jepang merupakan sistem sosial yang mencerminkan discernment, yakni kesadaran penutur terhadap posisi dirinya

dalam struktur hubungan interpersonal dan situasional, bukan semata pilihan strategis sebagaimana diasumsikan dalam teori kesantunan universal Brown dan Levinson (1987). Di sisi lain, Fukada dan Asato (2004) berargumen bahwa penggunaan honorifik Jepang tetap dapat dijelaskan secara memadai melalui kerangka *face-saving* Brown dan Levinson, sekalipun dengan penyesuaian kontekstual tertentu. Perdebatan teoretis ini menegaskan bahwa *keigo* merupakan fenomena linguistik yang kompleks dan multitafsir, sehingga kesalahan penerjemahan pada unsur *keigo* bukan sekadar kesalahan pemilihan register atau gaya bahasa, melainkan berpotensi menjadi kesalahan transfer budaya yang dapat mengubah persepsi hubungan sosial antarpihak dalam komunikasi bisnis.

2.3 Translation Quality Assessment dan Multidimensional Quality Metrics

Translation Quality Assessment (TQA) bertujuan mengevaluasi kualitas hasil penerjemahan berdasarkan hubungan fungsional antara teks sumber dan teks sasaran (House, 2015). Salah satu kerangka kerja TQA yang banyak diadopsi dalam riset maupun industri penerjemahan adalah *Multidimensional Quality Metrics* (MQM), yang dikembangkan oleh Lommel, Burchardt, dan Uszkoreit (2014). MQM mengembangkan evaluasi kualitas penerjemahan melalui klasifikasi kesalahan multidimensi yang mencakup enam kategori utama: *accuracy* (kesepadanan makna antara teks sumber dan sasaran), *terminology* (ketepatan penggunaan istilah khusus bidang), *fluency* (kewajaran dan kealamian bahasa sasaran), *grammar* (ketepatan struktur tata bahasa), *style* (kesesuaian register, formalitas, dan nada bahasa), dan *locale convention* (kesesuaian dengan konvensi budaya dan kebiasaan komunikasi pada locale sasaran).

Kerangka MQM dipilih dalam penelitian ini karena memungkinkan analisis kesalahan yang lebih granular dibandingkan metrik evaluasi otomatis berbasis kemiripan permukaan, seperti BLEU, yang cenderung kurang sensitif terhadap nuansa pragmatik dan sosiokultural (Castilho et al., 2017). Pendekatan ini juga selaras dengan model penilaian kualitas terjemahan yang dikembangkan Nababan dkk. (2012) dalam konteks penerjemahan Indonesia, yang menekankan pentingnya menilai keakuratan, keberterimaan, dan keterbacaan secara terpadu, serta konsisten dengan prinsip kesepadanan dinamis yang dirumuskan Nida dan Taber (1982).

METODE PENELITIAN

Penelitian ini menggunakan metode kualitatif deskriptif dengan pendekatan *translation error analysis*. Pendekatan ini dipilih karena bertujuan mendeskripsikan dan mengklasifikasikan jenis-jenis kesalahan penerjemahan secara mendalam, bukan untuk menguji hubungan kausal antarvariabel secara statistik.

Sumber data penelitian berupa korpus yang terdiri atas 50 kalimat teks bisnis berbahasa Jepang otentik dari buku Matsumoto, S., Sakuma, Y., Nagatomo, E., Nanba, F., Matsukura, A., & Hamahata, Y. (2018). *Writing business emails in Japanese: The basics and practical examples*. The Japan Times yang mencakup ragam jenis teks seperti surel bisnis (*bijinesu meeru*), komunikasi internal perusahaan, surat permintaan (*irai*), ungkapan terima kasih (*kansha*), ungkapan permintaan maaf (*owabi*), dan komunikasi formal lain yang lazim dijumpai dalam korespondensi bisnis Jepang. Setiap kalimat sumber kemudian diterjemahkan ke dalam bahasa Indonesia menggunakan tiga sistem penerjemahan berbasis AI, yaitu *Google Translate*, *DeepL*, dan *ChatGPT*, sehingga diperoleh 50 kalimat hasil

terjemahan dari masing-masing sistem, atau total 150 unit data analisis terjemahan.

Instrumen penelitian menggunakan kategori-kategori dalam kerangka MQM (Lommel et al., 2014), yaitu *Accuracy*, *Terminology*, *Fluency*, *Grammar*, *Style*, dan *Locale Convention*, yang masing-masing dilengkapi dengan indikator tingkat keparahan kesalahan (*severity level*): *minor*, *major*, dan *critical*, mengikuti prinsip evaluasi yang juga diterapkan dalam model *TQA House* (2015).

Prosedur penelitian dilakukan melalui beberapa tahapan sistematis. Tahap pertama adalah pengumpulan teks sumber berbahasa Jepang dari berbagai konteks komunikasi bisnis. Tahap kedua adalah proses penerjemahan menggunakan ketiga sistem AI dengan kondisi input yang dikendalikan agar sebanding. Tahap ketiga adalah membandingkan teks sumber dengan hasil terjemahan untuk mengidentifikasi penyimpangan makna, bentuk, maupun fungsi pragmatik. Tahap keempat adalah mengklasifikasikan setiap kesalahan yang ditemukan ke dalam kategori MQM yang sesuai. Tahap kelima adalah melakukan interpretasi linguistik dan budaya terhadap pola kesalahan yang muncul, termasuk mengaitkannya dengan teori kesantunan dan karakteristik bahasa Jepang sebagai bahasa berkonteks tinggi (House, 2015; Ide, 1989).

HASIL DAN PEMBAHASAN

4.1 Distribusi Kesalahan Terjemahan AI Berdasarkan Kategori MQM

Berdasarkan analisis terhadap 150 unit data hasil penerjemahan teks bisnis Jepang-Indonesia menggunakan *Google Translate*, *DeepL*, dan *ChatGPT*, ditemukan bahwa ketiga sistem AI secara umum menunjukkan kemampuan yang baik dalam menghasilkan bahasa sasaran yang dapat dipahami pembaca. Meskipun demikian, sejumlah kesalahan tetap muncul, khususnya pada aspek pragmatik dan budaya komunikasi. Klasifikasi kesalahan berdasarkan kerangka MQM disajikan pada Tabel 1.

Tabel 1. Distribusi Kesalahan Terjemahan AI Berdasarkan Kategori MQM

Kategori MQM	Jumlah Kesalahan	Persentase
<i>Style / Keigo</i>	45	50%
<i>Accuracy</i>	23	25,5%
<i>Fluency</i>	9	10%
<i>Locale Convention</i>	8	8,9%
<i>Terminology</i>	5	5,6%
<i>Grammar</i>	0	0%
<i>Total</i>	90	100%

Data pada Tabel 1 menunjukkan bahwa kesalahan terbesar terdapat pada kategori *Style*, khususnya yang berkaitan dengan *keigo*, yang mencapai 50% dari keseluruhan kesalahan yang teridentifikasi. Temuan ini membuktikan bahwa tantangan utama AI dalam menerjemahkan bahasa Jepang bisnis bukan terutama berkaitan dengan pemahaman

kosakata atau struktur kalimat, melainkan dengan kemampuan mempertahankan nilai sosial yang terkandung dalam ekspresi kesantunan bahasa Jepang (Fukada & Asato, 2004; Ide, 1989). Sebaliknya, kategori *Grammar* tidak menunjukkan kesalahan yang signifikan, yang mengindikasikan bahwa teknologi penerjemahan neural mutakhir telah cukup mampu menghasilkan struktur bahasa Indonesia yang gramatikal dan dapat diterima (Bahdanau et al., 2015; Vaswani et al., 2017). Akan tetapi, kemampuan gramatikal yang baik tersebut tidak selalu berbanding lurus dengan kemampuan memahami konteks sosial dan budaya, sejalan dengan temuan Castilho dkk. (2017) bahwa keunggulan NMT pada aspek *fluency* tidak serta-merta menjamin keakuratan pada level makna pragmatik.

4.2 Analisis Kesalahan Style/Keigo

Kesalahan terbesar ditemukan pada penerjemahan bentuk *keigo*, khususnya *kenjougo* atau bahasa merendahkan diri. Tabel 2 merangkum sejumlah bentuk *kenjougo* yang paling sering mengalami reduksi atau kehilangan makna dalam proses penerjemahan AI.

Tabel 2. Bentuk Kenjougo yang Sering Mengalami Kehilangan Makna

Ekspresi Jepang	Jenis	Fungsi
申し上げます	<i>Kenjougo</i>	Menghormati lawan bicara dengan merendahkan tindakan diri
いたします	<i>Kenjougo</i>	Bentuk sopan dari “ <i>suru</i> ” (melakukan)
いただく	<i>Kenjougo</i>	Menunjukkan penerimaan dengan hormat
伺う	<i>Kenjougo</i>	Bentuk rendah hati dari “pergi/bertanya”
存じます	<i>Kenjougo</i>	Bentuk hormat dari “berpikir”

Contoh Data:

Teks sumber : 明日の打ち合わせは約束どおり13時にお邪魔いたします。

(*Ashita no uchiawase wa yakusoku doori juusan-ji ni*

ojamaitashimasu)

Hasil Google Translate : “Saya akan hadir di pertemuan besok pukul 13.00 seperti yang telah dijanjikan.”

Secara makna dasar, hasil terjemahan tersebut dapat diterima dan dipahami pembaca sasaran. Namun, terdapat kehilangan fungsi pragmatik pada ekspresi *ojama itashimasu*. Secara harfiah, ungkapan ini berarti “mengganggu”, tetapi dalam komunikasi bisnis Jepang digunakan sebagai bentuk kerendahan hati ketika seseorang akan mengunjungi atau mendatangi tempat pihak lain. Dengan demikian, makna ungkapan tersebut bukan sekadar “datang”, melainkan lebih mendekati “saya datang dengan menghormati ruang dan waktu Anda”. Kesalahan ini dikategorikan sebagai *MQM Style Error* dengan tingkat keparahan minor, karena informasi utama (waktu dan kesediaan hadir) masih tersampaikan, tetapi nilai kesantunan dalam tuturan mengalami reduksi. Temuan ini menunjukkan bahwa sistem AI mampu membaca makna semantik permukaan, tetapi belum sepenuhnya memahami makna interpersonal yang melekat dalam komunikasi Jepang (Ide, 1989).

4.3 Analisis Kesalahan Accuracy

Kategori kesalahan terbanyak kedua adalah *Accuracy Error*. Kesalahan pada kategori ini umumnya muncul dalam bentuk perubahan makna, penghilangan informasi, penambahan informasi yang tidak terdapat pada teks sumber, atau kesalahan interpretasi konteks situasional.

Contoh Data:

Teks sumber : 突然メールを差し上げる失礼をお許してください。

(*Totsuzen meeru o sashiageru shitsurei o oyurushi kudasai*)

Hasil Google Translate : "Mohon maaf atas keterlambatan saya mengirimkan email ini."

Kesalahan utama terdapat pada penerjemahan kata *totsuzen* yang bermakna "tiba-tiba", namun diterjemahkan menjadi "keterlambatan". Kekeliruan ini menyebabkan pergeseran konteks komunikasi secara signifikan. Makna asli teks sumber adalah permintaan maaf penutur karena menghubungi pihak lain secara tiba-tiba atau tanpa pemberitahuan sebelumnya, sedangkan hasil terjemahan AI justru menyampaikan permintaan maaf atas keterlambatan pengiriman pesan-dua situasi komunikasi yang secara pragmatik berbeda. Dalam konteks komunikasi bisnis formal, pergeseran makna sekecil ini berpotensi menimbulkan kesalahpahaman mengenai situasi yang sesungguhnya terjadi. Kesalahan ini dikategorikan sebagai *MQM Accuracy-Mistranslation* dengan tingkat keparahan major.

4.4 Analisis Kesalahan Terminology

Kesalahan pada kategori terminologi relatif sedikit ditemukan dalam penelitian ini, yang mengindikasikan bahwa ketiga sistem AI cukup andal dalam mengenali istilah-istilah bisnis umum. Meskipun demikian, beberapa istilah khusus masih mengalami generalisasi makna yang kurang tepat.

Contoh Data:

Teks sumber : 請求書を本日郵送いたしました。

(*seikyuusho o honjitsu yuusou itashimashita*)

Hasil DeepL : "Tagihan telah kami kirimkan melalui pos hari ini."

Istilah *seikyuusho* sesungguhnya lebih tepat diterjemahkan sebagai "faktur" atau "invoice", bukan sekadar "tagihan", karena dalam konteks administrasi perusahaan Jepang, *seikyuusho* merujuk pada dokumen transaksi resmi dengan fungsi akuntansi dan legal yang spesifik, berbeda dengan istilah "tagihan" yang konotasinya lebih umum dan informal. Kesalahan ini dikategorikan sebagai MQM: *Terminology Error*.

4.5 Analisis Kesalahan Fluency

Kesalahan fluency terjadi ketika hasil terjemahan secara semantik dapat dikatakan benar, tetapi terasa kurang natural atau tidak lazim digunakan dalam register bahasa Indonesia yang setara.

Contoh Data:

Teks sumber : 心よりお祝い申し上げます。

(*Kokoro yori oiwai moushiagemasu*)

Hasil DeepL : "Saya mengucapkan selamat dari lubuk hati yang paling dalam."

Terjemahan tersebut secara literal dapat dikatakan benar, karena frasa *kokoro yori* memang bermakna "dari hati". Namun, dalam konteks komunikasi bisnis Jepang, ungkapan tersebut merupakan formula kesantunan formal yang lazim digunakan dalam korespondensi

resmi perusahaan, bukan ungkapan emosional personal yang intens sebagaimana kesan yang ditimbulkan terjemahan literal tersebut dalam bahasa Indonesia. Padanan yang lebih sesuai dengan register bisnis formal adalah “Kami menyampaikan ucapan selamat dengan tulus.” Kesalahan ini dikategorikan sebagai MQM: *Fluency Error*.

4.6 Analisis Kesalahan Locale Convention

Kesalahan *locale convention* muncul ketika sistem AI gagal mempertahankan norma dan konvensi komunikasi yang berlaku dalam budaya Jepang ketika dialihkan ke bahasa Indonesia.

Contoh Data

Teks sumber

: 今後とも何卒よろしくお願ひ申し上げます。

(*kongo tomo nanitozo yoroshiku onegaiitashimasu*)

Hasil Google Translate : “Kami menantikan dukungan Anda yang berkelanjutan.”

Hasil DeepL : “Mohon dukungan Anda di masa mendatang.”

Hasil ChatGPT: “Kami berharap dapat terus menjalin kerja sama yang baik ke depannya.”

Ungkapan *yoroshiku onegaiitashimasu* merupakan salah satu formula komunikasi bisnis Jepang yang paling khas dan tidak memiliki padanan langsung dalam bahasa Indonesia. Ungkapan ini mengandung makna budaya yang kompleks, yaitu sekaligus berfungsi menjaga keberlangsungan hubungan bisnis, menunjukkan sikap penghormatan kepada mitra, dan membangun fondasi kerja sama jangka panjang. Dari ketiga hasil terjemahan di atas, terlihat bahwa *Google Translate* dan *DeepL* cenderung menerjemahkan ungkapan tersebut secara lebih literal, sehingga kesan yang muncul adalah relasi satu arah (permintaan dukungan semata), sedangkan *ChatGPT* menghasilkan terjemahan yang secara pragmatik lebih mendekati makna asli karena berhasil mempertahankan nuansa resiprokal dalam hubungan kerja sama. Pola ini sejalan dengan temuan Jiao dkk. (2023) bahwa model berbasis LLM cenderung lebih adaptif dalam mempertimbangkan konteks wacana yang lebih luas dibandingkan sistem NMT konvensional.

4.7 Perbandingan Karakteristik Tiga Sistem AI

Berdasarkan hasil analisis korpus secara keseluruhan, ditemukan bahwa masing-masing sistem AI memiliki karakteristik dan kecenderungan kesalahan yang berbeda, sebagaimana dirangkum pada Tabel 3.

Tabel 3. Perbandingan Karakteristik Tiga Sistem AI dalam Menerjemahkan Teks Bisnis Jepang

Sistem	Keunggulan	Kelemahan
Google Translate	Cepat dan baik pada struktur kalimat sederhana	Lemah pada konteks budaya dan <i>keigo</i>
DeepL	Naturalitas bahasa tinggi	Masih cenderung literal pada ekspresi budaya
ChatGPT	Lebih baik dalam memahami konteks pragmatik	Berpotensi melakukan interpretasi yang terlalu bebas

4.8 Pembahasan Terpadu

Hasil penelitian secara keseluruhan menunjukkan bahwa perkembangan AI *translation*

telah berhasil mengatasi sebagian besar persoalan linguistik permukaan, seperti ketepatan tata bahasa dan pemilihan kosakata umum-hal ini tampak jelas dari tidak ditemukannya kesalahan signifikan pada kategori Grammar. Akan tetapi, penerjemahan bahasa Jepang bisnis menuntut lebih dari sekadar kompetensi linguistik; ia membutuhkan pemahaman pragmatik yang mendalam terhadap fungsi sosial bahasa.

Bahasa Jepang dikenal sebagai bahasa dengan konteks komunikasi tinggi (*high-context language*), yang berarti sebuah ekspresi tidak hanya menyampaikan informasi proposisional, tetapi juga secara simultan menegosiasikan dan menegaskan hubungan sosial antara pembicara dan lawan bicara (Ide, 1989). Sebagai ilustrasi, ungkapan *osoreimasuga* tidak semata-mata bermakna “maaf”, tetapi berfungsi sebagai strategi komunikasi untuk melembutkan permintaan, menunjukkan sikap penghormatan, sekaligus menjaga keharmonisan hubungan sosial antarpihak-sebuah fungsi yang sejalan dengan konsep *negative politeness strategy* dalam kerangka Brown dan Levinson (1987), yaitu strategi kesantunan yang berorientasi pada penghormatan terhadap ruang dan kebebasan bertindak lawan bicara.

Pola kesalahan yang ditemukan dalam penelitian ini mengindikasikan bahwa sistem AI cenderung lebih mudah memproses makna referensial (apa yang secara harfiah dirujuk oleh sebuah kata atau frasa) dibandingkan makna pragmatik (mengapa dan dalam kondisi sosial seperti apa kata tersebut dipilih dan digunakan). Kecenderungan ini sejalan dengan karakteristik arsitektur NMT dan LLM yang dioptimalkan untuk memprediksi kesepadanan linguistik berdasarkan data pelatihan berskala besar, namun belum sepenuhnya mampu memodelkan pengetahuan sosiokultural implisit yang melekat pada komunitas tutur tertentu (Brown et al., 2020; House, 2015). Temuan ini juga memperkuat hasil studi Castilho dkk. (2017) yang menyimpulkan bahwa kemajuan NMT secara *fluency* tidak otomatis menyelesaikan persoalan akurasi pragmatik, terutama untuk pasangan bahasa dengan jarak budaya dan struktur yang signifikan.

Dengan demikian, sekalipun sistem AI khususnya model berbasis LLM seperti *ChatGPT* menunjukkan kemajuan dalam mempertimbangkan konteks wacana yang lebih luas (Jiao et al., 2023), peran penerjemah manusia tetap krusial sebagai evaluator dan penyunting akhir melalui proses *Machine Translation Post-Editing* (MTPE), khususnya dalam ranah komunikasi profesional Jepang-Indonesia yang menuntut ketepatan pragmatik dan kepekaan budaya tinggi.

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisis terhadap 150 unit data penerjemahan AI Jepang-Indonesia menggunakan pendekatan *Translation Quality Assessment* berbasis *Multidimensional Quality Metrics*, dapat disimpulkan bahwa sistem AI translation telah memiliki kemampuan tinggi dalam menghasilkan terjemahan yang secara umum dapat dipahami pembaca sasaran, tetapi masih menunjukkan keterbatasan signifikan pada aspek budaya dan pragmatik.

Jenis kesalahan yang ditemukan dalam penelitian ini meliputi kesalahan gaya bahasa (*Style Error*), kesalahan makna (*Accuracy Error*), kesalahan kewajaran bahasa (*Fluency Error*), kesalahan budaya (*Locale Convention Error*), dan kesalahan terminologi (*Terminology Error*). Kesalahan terbesar terdapat pada kategori *Style/Keigo*, yaitu sebanyak 45 dari 90

kasus kesalahan (50%), yang menunjukkan bahwa sistem AI masih mengalami kesulitan dalam memahami fungsi sosial bahasa Jepang bisnis, terutama yang berkaitan dengan bentuk penghormatan dan kerendahan hati dalam keigo.

Faktor utama yang menyebabkan munculnya kesalahan tersebut antara lain: (1) perbedaan mendasar pada sistem kesantunan antara bahasa Jepang dan bahasa Indonesia; (2) sifat komunikasi Jepang yang sangat kontekstual (*high-context*); dan (3) keterbatasan sistem AI dalam memodelkan hubungan sosial dan pengetahuan budaya implisit. Penelitian ini menegaskan bahwa *AI translation* dapat dimanfaatkan secara efektif sebagai alat bantu penerjemahan awal, tetapi hasilnya tetap memerlukan evaluasi dan penyuntingan oleh penerjemah manusia, terutama dalam konteks komunikasi bisnis profesional yang menuntut ketepatan pragmatik tinggi.

5.2 Keterbatasan dan Saran Penelitian Lanjutan

Penelitian ini memiliki sejumlah keterbatasan yang perlu dikemukakan secara transparan. Pertama, jumlah data sebanyak 50 teks sumber (150 unit terjemahan) bersifat purposif dan belum tentu merepresentasikan seluruh ragam genre teks bisnis Jepang secara proporsional. Kedua, proses klasifikasi kesalahan berdasarkan MQM dalam penelitian ini dilakukan secara kualitatif oleh peneliti, sehingga berpotensi mengandung unsur subjektivitas interpretatif, meskipun telah merujuk pada definisi kategori MQM yang baku (Lommel et al., 2014). Ketiga, penelitian ini belum melibatkan uji reliabilitas antarpenilai (*inter-rater reliability*) secara kuantitatif.

Penelitian selanjutnya disarankan untuk memperluas cakupan data dengan jumlah dan ragam genre teks yang lebih besar, melibatkan beberapa penilai (*rater*) ahli bahasa Jepang untuk meningkatkan keandalan klasifikasi MQM, serta mempertimbangkan perbandingan dengan versi model AI yang lebih mutakhir mengingat kecepatan perkembangan teknologi LLM. Penelitian lanjutan juga dapat mengeksplorasi efektivitas strategi *prompting* tertentu dalam memperbaiki akurasi penerjemahan *keigo* pada model berbasis LLM, serta mengkaji persepsi pengguna profesional Jepang-Indonesia terhadap keberterimaan hasil terjemahan AI dalam konteks bisnis nyata.

DAFTAR PUSTAKA

- [1] Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *International Conference on Learning Representations*.
- [2] Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.
- [3] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- [4] Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics*, 108, 109–120. <https://doi.org/10.1515/pralin-2017-0013>
- [5] Fukada, A., & Asato, N. (2004). Universal politeness theory: Application to the use of Japanese honorifics. *Journal of Pragmatics*, 36(11), 1991–2002.

- <https://doi.org/10.1016/j.pragma.2003.11.006>
- [6] House, J. (2015). *Translation quality assessment: Past and present*. Routledge.
- [7] Ide, S. (1989). Formal forms and discernment: Two neglected aspects of linguistic politeness. *Multilingua*, 8(2-3), 223-248. <https://doi.org/10.1515/mult.1989.8.2-3.223>
- [8] Jiao, W., Wang, W., Huang, J., Wang, X., Shi, S., & Tu, Z. (2023). *Is ChatGPT a good translator? A preliminary study*. arXiv. <https://doi.org/10.48550/arXiv.2301.08745>
- [9] Koehn, P. (2020). *Neural machine translation*. Cambridge University Press.
- [10] Lommel, A., Burchardt, A., & Uszkoreit, H. (2014). Multidimensional Quality Metrics: A framework for declaring translation quality metrics. *Tradumàtica*, 12, 455-463. <https://doi.org/10.5565/rev/tradumatica.77>
- [11] Matsumoto, S., Sakuma, Y., Nagatomo, E., Nanba, F., Matsukura, A., & Hamahata, Y. (2018). *Writing business emails in Japanese: The basics and practical examples*. The Japan Times.
- [12] Nababan, M., Nuraeni, A., & Sumardiono. (2012). Pengembangan model penilaian kualitas terjemahan. *Kajian Linguistik dan Sastra*, 24(1), 39-57.
- [13] Nida, E. A., & Taber, C. R. (1982). *The theory and practice of translation*. Brill.
- [14] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998-6008.

HALAMAN INI SENGAJA DIKOSONGKAN