

## Pengesanan Ralat Nahu Esei Bahasa Inggeris Melalui Pemprosesan Bahasa Tabii

NUR ASMA MOHD RAZALI, NAZLIA OMAR & SAADIYAH DARUS

### ABSTRAK

*Pelajar yang memasuki pusat pengajian tinggi dijangkakan mempunyai kemahiran bahasa Inggeris yang baik untuk membolehkan mereka menyediakan tugas kursus dan aktiviti akademik dalam bahasa Inggeris. Satu daripada tugas ialah menulis esei. Analisis kesalahan tatabahasa ke atas empat ratus (400) esei pelajar ESL (English as a Second Language) di Universiti Kebangsaan Malaysia (UKM) dijalankan dan hasil analisis menunjukkan ralat nahu (tenses) adalah ralat yang paling kerap dilakukan. Sebuah alat pengesanan ralat nahu dalam penulisan pelajar ESL diperlukan memandangkan alat yang sedia ada tidak dibangunkan khusus untuk pelajar ESL di Malaysia. Kertas ini bertujuan memperkenalkan teknik berasaskan heuristik dan peraturan untuk mengesan ralat nahu. Berasaskan heuristik dan peraturan ini, ATDEiT<sup>TM</sup> dibangunkan. ATDEiT<sup>TM</sup> diuji menggunakan 50 set data ujian (esei bahasa Inggeris). Keputusan yang dijana oleh ATDEiT<sup>TM</sup> dibandingkan dengan keputusan pakar bahasa. ATDEiT<sup>TM</sup> memperoleh 95% dapatan semula dan 76.2% kejituan. Ini menunjukkan bahawa pendekatan berasaskan heuristik dan peraturan membantu dalam proses mengesan ralat nahu dalam esei pelajar ESL secara automatik.*

*Kata kunci: Pengesanan ralat, pemeriksaan esei berbantuan komputer, heuristik, peraturan, kesalahan tatabahasa, esei ESL.*

### ABSTRACT

*Students entering tertiary level of education are expected to have the linguistic competence in English to be able to perform their academic activities. One of the tasks they need to perform is essay writing. Four hundred (400) essays of ESL (English as a Second Language) learners at Universiti Kebangsaan Malaysia (UKM) were analysed for grammatical errors and the most common error were tenses. An automated error detection tool is needed because currently available softwares are not developed for ESL learners in Malaysia. This paper aims to introduce a heuristics and rule based approach for detecting errors in tenses. To realize the utilities of these heuristics and rule, a tool called ATDEiT<sup>TM</sup> was implemented. ATDEiT<sup>TM</sup> was evaluated using 50 test data sets (English language essays). Results generated by ATDEiT<sup>TM</sup> were tested against human performance results. ATDEiT<sup>TM</sup> achieved 95% recall and 76.2% precision. The results revealed that heuristics and rule based approach can improve the process of detecting grammatical errors in tenses for automatic ESL writing.*

*Keywords: Error detection, computer-based essay marking system, heuristics, rule, grammatical errors, ESL essay.*

## PENGENALAN

Bahasa Inggeris merupakan bahasa yang digunakan secara meluas di Malaysia. Kedudukan bahasa Inggeris kini telah menjadi bahasa kedua dalam sistem pendidikan negara mahupun hubungan antarabangsa. Setiap rakyat Malaysia yang mengikuti sistem pendidikan negara diharapkan mempunyai kemahiran asas untuk berkomunikasi dalam bahasa Inggeris terutamanya dalam bidang ekonomi, perdagangan, pekerjaan (Richards 2002; Su Hie Ting 2002), sains dan teknologi (Foo & Richards 2004). Dalam sistem pendidikan negara ini, pembelajaran dan pengajaran bahasa Inggeris secara formal sebagai bahasa kedua (*English as a Second Language (ESL)*) dijalankan pada peringkat awal, daripada sekolah rendah sehingga ke institusi pengajian tinggi (IPT) (Foo & Richards 2004).

Pelbagai aktiviti yang bersesuaian dijalankan di dalam kelas bahasa Inggeris agar dapat menarik minat pelajar untuk mendalami dan menguasai bahasa tersebut. Empat kemahiran bahasa yang diserapkan ke dalam silabus pembelajaran ialah kemahiran bertutur, mendengar, membaca dan menulis. Menurut Ellis (1985), seseorang pelajar dapat menerima sesuatu peraturan (*rule*) dalam bahasa yang baru dipelajari, sekiranya berusaha dalam proses pembelajaran dan mendapatkan pengalaman yang berterusan. Tahap pemerolehan bahasa kedua (*second language acquisition*) bagi bahasa Inggeris berbeza bagi setiap individu. Untuk menguasai bahasa Inggeris bukan merupakan suatu perkara yang mudah yang boleh diperoleh dalam masa yang singkat. Proses pembelajaran bahasa Inggeris merupakan proses yang berterusan dan kompleks. Richards (2002: 2) menyenaraikan lapan persoalan penting dalam aspek pengajaran dan pembelajaran bahasa Inggeris seperti berikut:

- i. *Apakah matlamat pengajaran bahasa Inggeris?*,
- ii. *Apakah cara terbaik untuk belajar sesuatu bahasa?*,
- iii. *Apakah peranan tatabahasa dalam pengajaran bahasa?*
- iv. *Apakah proses yang terlibat dalam pembelajaran bahasa kedua?*
- v. *Apakah peranan sebagai seorang pelajar?*,
- vi. *Bagaimanakah cara untuk mengajar empat kemahiran?*
- vii. *Bagaimanakah menilai pembelajaran pelajar?*
- viii. *Bagaimanakah persediaan sebagai guru bahasa?*

Aspek ketiga adalah berkaitan tatabahasa (*grammar*), Richards (2002), melihat matlamat utama pengajaran dan pembelajaran tatabahasa sebagai memahami, membina dan mempelbagaikan ayat bahasa Inggeris yang mempunyai maksud berlainan, memahami peraturan dalam membina ayat seperti frasa dan klausa agar dapat meminimumkan kesalahan tatabahasa (*grammatical error*) atau ralat (*error*) dan melatih diri untuk menulis dan bertutur sebagai asas komunikasi yang penting. Penggunaan tatabahasa Inggeris yang betul hanya boleh diperoleh dengan latihan yang berterusan. Namun, pelajar sering kali melakukan ralat atau kesilapan sepanjang proses pembelajaran ini. Ellis (1985) menyatakan ralat atau kesilapan yang dilakukan merupakan norma semulajadi dan sebahagian daripada proses pembelajaran dan pemerolehan bahasa kedua, tidak perlu dianggap serius. Hudelson (1989:1) turut menyatakan:

*Don't worry about teaching writing until the children have mastered the oral language. Teach children to understand and speak English first. Language develops sequentially from listening to speaking to reading, and finally to writing. Writing is the last of the language processes to develop, and it is the last of the language processes to be taught. Elementary ESL educators should focus on developing children's speaking vocabularies and not concentrate on literacy*

Kenyataan di atas jelas menunjukkan bahawa menulis dalam bahasa Inggeris merupakan satu aspek yang paling sukar dikuasai berbanding dengan kemahiran mendengar, membaca dan bertutur. Silva dan Matsuda (2002) juga melihat aspek penulisan sebagai "...a complex, recursive and creative process that

is very similar in its general outlines for first and second language writer; learning to write requires the developmental of an efficient and effective composing process". Proses gabungan (*composing process*) yang dimaksudkan meliputi aspek seperti merancang, mendraf, menyemak, mengubah dan mendapatkan maklum balas (Cumming 2001; Myles 2002; Silva & Matsuda 2002).

Menulis esei dalam pelbagai bentuk genre seperti penceritaan (*narrative*), penghuraian (*descriptive*) dan penghujahan (*argumentative*) sudah tentu dapat mempelbagaikan dan meningkatkan kemahiran menulis pelajar *ESL*. Mutu penulisan esei pelajar *ESL* dapat dipertingkatkan dengan sentiasa menulis, mendapatkan maklum balas daripada pengajar, memperbaiki ralat yang dibuat dan proses pembelajaran ini berulang sebanyak yang mungkin. Menurut Myles (2002), banyak faktor yang perlu diambil kira semasa menulis esei seperti aspek perluasan idea (*idea development*), kejelasan (*clarity*), pertautan atau kohesi (*cohesion*) dan tatabahasa.

## TATABAHASA INGGERIS

Bahasa adalah sesuatu yang bersistem dan berperaturan. Setiap bahasa mempunyai tatabahasanya. Tatabahasa atau nahu ialah peraturan berasaskan teori tertentu yang menghuraikan ayat dalam suatu bahasa dengan kaedah yang paling mudah, tepat dan lengkap.

Menurut Nik Safiah Karim et al. (2008), tiga penanda tatabahasa yang penting dalam menentukan struktur sesuatu ayat bahasa Inggeris ialah penanda kasus, unsur keserasian dan kala. Contoh penanda kasus misalnya, *He kicked the ball* yang membawa maksud kasus nominatif untuk subjek *He*, tidak boleh diubah atau ditukar menjadi *\*Him kicked the ball*, yang frasa nama subjeknya *Him* membawa unsur kasus akusatif untuk objek langsung. Dalam bahasa Inggeris, urutan kata pada ayat aktif ialah *subjek + kata kerja + objek* (SKO atau SVO dalam bahasa Inggeris) seperti ayat *John kicked the ball*. Urutan kata dalam ayat pasif pula ialah *objek + kata kerja + subjek* seperti ayat *The ball was kicked by John*. Unsur keserasian dalam ayat aktif bahasa Inggeris boleh dilihat antara kata nama subjek dengan kata kerja seperti contoh ayat berikut:

- i. He helps the old man.
- ii. The boys help the old man.
- iii. \*He help the old man. (ayat ini tidak gramatis)
- iv. \*The boys helps the old man. (ayat ini tidak gramatis)

Kala atau lebih dikenali sebagai *tenses*, ialah penanda atau perubahan yang terjadi pada bentuk kata kerja sebagai menentukan kala atau waktu kejadian, sama ada *akan*, *sedang*, atau *sudah* berlaku (Nik Safiah Karim et al. 2008). Contoh ayat bahasa Inggeris yang menggunakan penanda kala adalah seperti berikut:

- i. I *kick* the ball today.
- ii. I *kicked* the ball yesterday.
- iii. I *will kick* the ball tomorrow.

Ayat *I kick the ball today* menggunakan *simple present tense* dan ini boleh dilihat pada kata kerja *kick* dan penggunaan perkataan *today*. Ayat *I kicked the ball yesterday* menunjukkan penggunaan *simple past tense* dan kata kerja *kick* tadi diubah menjadi *kicked* dan adanya perkataan *yesterday*. Ayat *I will kick the ball tomorrow* menunjukkan penggunaan *simple future tense* pada frasa kerja *will kick*. Noorizah dan Rosniah (2002) membincangkan dua belas jenis *tenses* Inggeris yang terdiri daripada *simple present tense*, *present progressive tense*, *present perfect tense*, *present perfect progressive tense*, *simple past tense*, *past progressive tense*, *past perfect tense*, *past perfect progressive tense*, *simple future tense*, *future progressive tense*, *future perfect tense*, dan *future perfect progressive tense*.

## LATAR BELAKANG PENGESANAN RALAT TATABAHASA

Para penyelidik mengadaptasikan pelbagai teknik dan algoritma dalam mengesan ralat dan memeriksa sesuatu esei. Menurut Valenti et al. (2003), terdapat lebih daripada 40 buah sistem pemeriksaan esei berbantuan komputer (*computer-based essay marking system (CBEM)*) di pasaran. Namun, kebanyakan sistem *CBEM* dan penyemak nahu (*grammar checker*) dibangunkan berdasarkan korpus data esei bahasa Inggeris oleh penutur bahasa pertama contohnya *Grammatik* dan *Critique*. Perisian yang dibangunkan khusus untuk memeriksa ralat tatabahasa atau ralat nahu berasaskan data esei bahasa Inggeris sebagai bahasa kedua iaitu L2 sebagai data dalam kajian di Malaysia masih belum wujud. Jadi, sebuah alatan pengesanan ralat tatabahasa atau ralat nahu bagi penulisan esei bahasa Inggeris di Malaysia perlu dibangunkan.

Menurut Brock (1993), Collins (1989) dan Vernon (2000), analisis teks yang dijalankan oleh alatan penyemak nahu seperti *Grammatik IV*, *Critique* dan *Microsoft Word 2000* adalah kurang tepat. Collins (1989) berpendapat bahawa antara sebab ketidaktepatan alatan ini menyemak ralat tatabahasa ialah kerana pelbagai sistem dibangunkan berasaskan *pattern matching*. Sistem ini membandingkan perkataan atau frasa dalam esei dengan satu senarai perkataan yang disimpan sebagai bermasalah (ralat) dan memberikan maklum balas sekiranya sepadan dengan memberi isyarat kepada pengguna. Namun, isyarat kesalahan tatabahasa yang diberi mengelirukan pelajar. Thiesmeyer (1989) mengatakan bahawa analisis yang tidak tepat dijana oleh sesuatu sistem disebabkan pengaturcara telah meletakkan satu senarai ralat untuk sesuatu perkataan dan bukan berdasarkan analisis ralat sebenar. Pennington (1992) turut mengkritik alatan penyemak nahu sebagai memberikan maklum balas yang luas dan tidak tepat; tidak dilatih untuk proses mengedit; kualiti penulisan yang kurang dan analisis yang dijana secara tidak tepat.

Haist (2000) membuat pengujian ke atas *Microsoft Word 97's Grammar Checker* dan mendapatinya berkeupayaan menanda (*flag*) dan memberi maklum balas berkaitan ralat *subject verb agreement*, *comma* dan *capitalisations*, tetapi tidak dapat mengesan kebanyakan ralat *pronoun* atau *modifier pronoun*. *Word 97's Grammar Checker* dapat menanda dan membetulkan ralat *subject-verb-agreement* dan mendapat keputusan sebanyak 70% daripada 200 ayat yang diuji (Haist 2000), namun tidak dapat mengesan atau menanda ralat kata kerja *tenses* dalam ayat, sebaliknya hanya menanda keseluruhan ayat tersebut sebagai bermasalah. Paparan maklum balas yang tidak tepat, seperti semak semula kata kerja (*consider revising (verb use)*) sudah tentu mengelirukan pelajar kerana tidak pasti kata kerja yang mana satu salah daripada segi tatabahasanya dalam ayat yang diuji. Contoh ayat berkaitan kata kerja *tenses* yang diuji ialah *Karen ran home and \*finds that she has forgotten to turn the water off* (Haist 2000).

Menurut Rosniah dan Noorizah (2008), penggunaan nahu *tenses* yang salah dan sering dilakukan dalam penulisan esei *ESL* adalah seperti berikut:

- i. Kesilapan menggunakan subjek dan kata kerja yang tidak bersesuaian dalam bentuk *simple present tense*. Tatabahasa Inggeris menggariskan panduan penggunaan subjek dan kata kerja yang bersesuaian dengan jenis *tenses* dalam ayat. Perkataan *goes* perlu ditukarkan kepada *go* bagi penggunaan *simple present tense* dalam ayat gramatis.

I \*goes to school by bus.

- ii. Kesilapan menggunakan bentuk *simple past tense* untuk sesuatu yang sentiasa benar. Contoh ayat di bawah memperkatakan tentang sesuatu yang sentiasa benar dari dahulu hingga sekarang iaitu “bumi ini bulat”. Peraturan *simple present tense* mesti digunakan dalam ayat tersebut walaupun merujuk kepada masa lampau seperti *yesterday*. Perkataan *was* mesti ditukarkan kepada *is*.

Yesterday I told him that the earth \*was round.

- iii. Kesilapan menggunakan berlainan jenis *tenses* untuk kata kerja dalam satu ayat. Penggunaan kata kerja utama dalam ayat di bawah iaitu *said* dalam bentuk *simple past tense*, maka kata kerja yang lain dalam ayat perlu menggunakan bentuk *simple past tense*. Perkataan *doesn't* mesti ditukarkan kepada *didn't*. Peraturan ini sentiasa benar kecuali jika ayat memperkatakan sesuatu yang sentiasa benar seperti ayat *Zatul told Fizah the earth is round*.

Gee said she \*doesn't want to go.

Kajian untuk mengesan ralat nahu menggunakan sistem *CBEM* secara automatik di Malaysia amat sukar ditemui. Dapatan kajian oleh Lim Ho Peng (1976), Nur Asma et al. (2007) dan Saadiyah Darus (2005) menunjukkan ralat nahu *tenses* adalah ralat yang paling banyak dalam penulisan esei *ESL*. Satu sistem *CBEM* atau sistem pengesanan ralat yang dapat mengesan ralat nahu terutamanya ralat *tenses* sudah tentu dapat membantu pelajar *ESL* dalam mengenalpasti ralat tatabahasa Inggeris yang dilakukan dan seterusnya dapat meningkatkan kemahiran menulis esei.

## KAJIAN LEPAS

Penyelidikan lepas menghasilkan banyak sistem *CBEM* yang berupaya memeriksa dan memarkahkan esei yang panjang iaitu sistem automatik dan sistem separa automatik. Menurut Saadiyah (1999, 2005), sistem separa automatik tidak berupaya memeriksa esei secara automatik sepenuhnya dan masih memerlukan pengawasan daripada pengajar terutamanya dalam membaca dan memberikan markah esei pelajar. Sistem automatik mampu memeriksa esei secara automatik. Manusia ataupun pengajar tidak perlu membaca kesemua set esei untuk memeriksa dan memberi markah, sebaliknya komputer menanda markah berdasarkan kriteria yang telah diberi. Sistem *CBEM* yang dibincangkan ialah E-rater.

Sistem pengesanan ralat esei berbantuan komputer adalah sistem yang berupaya mengesan ralat atau kesalahan tatabahasa dalam esei secara automatik menggunakan teknik dan kaedah tertentu. Sesuai dengan matlamat penyelidikan ini untuk membangunkan alatan pengesanan ralat nahu *tenses* secara automatik bagi penulisan esei *ESL*, maka kajian terhadap sistem pengesanan ralat yang sedia ada dijalankan. Antara sistem pengesanan ralat esei yang dibincangkan ialah Sistem Grammar Diagnostic Expert, English Learning, Rule Based Style and Grammar Checker, Sentence Analyzer and Viewer dan English Grammar Checking.

Burstein et al. (1998) membangunkan E-rater berasaskan ciri penulisan yang diperlukan bagi *Graduate Management Admission Test* (GMAT). E-rater menggunakan kaedah pengkomputeran linguistik dan teknik analisis berstatistik (Burstein et al. 1998). Bagi memarkahkan esei pelajar, analisis struktur hujahan (*argument partitioning*), analisis struktur sintaksis dan penggunaan perbendaharaan kata diambil kira. E-rater kini ditukarkan ke dalam Criterion (Burstein 2003), sistem versi masa nyata berasaskan web yang dibangunkan oleh ETS Technologies. E-rater dilatih menggunakan sampel esei yang telah diperiksa dan diberi markah oleh pemeriksa. Attali dan Burstein (2005) mendapati E-rater "...was trained on a sample of essays written on the same topic that had been scored by human readers. It measured more than 50 features in all and then computed a stepwise linear regression to select those features that made a significant contribution to the prediction of essay scores".

Selain membuat peramalan menggunakan Algoritma *stepwise linear progression*, heuristik APA digunakan dalam memeriksa dan mengenalpasti struktur sintaksis sesuatu esei (Burstein et al. 2002). *Argument Partitioning and Annotation Program* (APA) menggunakan terma kata kunci dan struktur bukan leksikal. Jadual 1 menunjukkan esei yang dianalisis menggunakan heuristik APA. Jadual 2 menunjukkan sebahagian daripada peraturan yang digunakan oleh E-rater.

Sistem *Grammar Diagnostic Expert* (Grades) dibangunkan oleh Fox dan Bowden (2002) yang bermatlamat untuk mengesan kesalahan tatabahasa dalam ayat bahasa Inggeris yang ditulis oleh pengguna

JADUAL 1. Contoh esei dan heuristik APA

<i>Essay text</i>	<i>Argument Partitioned / Annotated Essay Text</i>
<p>...Another problem with the argument is found in the evidence used to support it. Big boards takes responsibility for increasing River City's awareness of the marathon runner over the three month period by putting her name on billboards, but they also stated that there was...</p>	<p><u>Another</u> problem with the argument is found in the evidence used to support it.            &gt; Start argument at Parallel Word: <u>Another</u>            Big boards takes responsibility for increasing River City's awareness of the marathon runner over the three month period by putting her name on billboards, <u>but</u> they also stated <u>that</u> there was...            &gt; Argument Developmental at Contrast Word: <u>but</u>            &gt; Argument Developmental at Complement Clause : <u>that</u></p>

Sumber: Burstein et al. 2002

JADUAL 2. Sebahagian daripada peraturan yang digunakan oleh E-rater

<b>Peraturan</b>	<b>Penerangan</b>
a). "AFTER" RULE	Extracts "after", "after", and "afterwards" if they occur sentence initially as conjunction.
b). "BUT" RULE	Extracts "but" if it is labeled as a conjunction.
c). "FURTHER" RULE	Extracts "further" "overall" or "altogether" if they occurred sentence-intially and do not modify another constituent.
d). "OR" RULE	Extracts the conjunctions "or" and "either" when they occur in the second column of the parse tree, and the node immediately following the conjunction is a verb phrase. If the very first sentence of a paragraph begins with a subordinate clause, extract the noun or pronoun from the main clause of NP and consider it to be the beginning of a new argument.
e). "SUBORDINATE CLAUSE" RULE	

Sumber: Burstein et al. 2002

bahasa pertamanya (L1) iaitu bahasa Jepun. Grades terdiri daripada koleksi pangkalan data perkataan yang senang diakses. Satu proses klasifikasi untuk mendiagnos kesalahan tatabahasa berdasarkan jenis ralat dibuat berasaskan pepadanan peraturan. Sistem ini berupaya mendiagnos dua jenis kesalahan tatabahasa iaitu ralat berkaitan kata kerja (*verb-related errors*) dan ralat berkaitan kata nama (*noun-related errors*). Grades mengesan ralat berkaitan kata kerja dahulu diikuti dengan ralat berkaitan kata nama dalam sesuatu esei. Analisis berkenaan frasa kata kerja dijalankan dengan mengeluarkan setiap kata kerja dalam ayat dan diletakkan dalam senarai elemen kata kerja yang terdiri daripada empat bahagian iaitu "have", "be", "modals", "do", dan "non auxiliaries verb". Setelah Grades mengenali elemen kata kerja, satu corak susunan (*sequence pattern*) dalam ayat akan diperiksa dan ralat berkaitan kata kerja akan dikesan. Pada peringkat pengujian sintaktik ini, Grades menjalankan empat tugas utama iaitu menguji sama ada subjek dan kata kerja adalah serasi dengan bilangan orang, kewujudan perkataan "do" dalam ayat negatif, kewujudan ayat pasif yang gramatis dan kewujudan kata kerja utama dalam bentuk sedang (*present*) jika wujud perkataan "do". Pengesanan ralat berkaitan kata nama dijalankan dengan menjejaki frasa kata nama dalam ayat dan melihat *determiner* atau kata sifat. Grades mengesan jujukan kata nama dan diikuti dengan *modifier* sama ada *determiner* atau kata sifat.

Ying dan Chih (2003) membangunkan aplikasi pintar untuk pembelajaran bahasa Inggeris yang dinamakan sistem English Learning. Aplikasi ini dibangunkan menggunakan pangkalan data multimedia dan teknologi Internet iaitu korpus multimedia Inggeris. Aplikasi pintar ini mengandungi artikel Inggeris, dialog dan video. Pertanyaan semantik dan penghurai *Link Grammar* digunakan dalam membangunkan korpus multimedia Inggeris ini. Menurut Ying dan Chih (2003), pembelajaran bahasa Inggeris merupakan proses yang sukar dikuasai bagi pelajar Cina. Aplikasi yang dibangunkan sesuai digunakan untuk pembelajaran bahasa Inggeris dan mempunyai dua sub sistem iaitu sistem English Sentence Error Detection dan sistem English Pattern Query. Sistem Sentence Error Detection berfungsi untuk mengenalpasti ralat tatabahasa dalam ayat dengan menggunakan penghurai *Link Grammar* yang telah diubahsuai berserta korpus pelajar.

Kajian ke atas lima jenis ayat bahasa Inggeris iaitu *simple sentence pattern*, *negative sentence pattern*, *interrogative sentence pattern*, *WH question sentence pattern* dan *imperative sentence pattern* dijalankan sebelum membangunkan sistem English Pattern Query. Lima jenis ayat ini mewakili jenis nahu *tenses* yang pelbagai iaitu *simple present tense*, *simple past tense*, *simple future tense*, *present perfective tense*, *past perfective tense*, *future perfective tense*, *present continuous tense*, *past continuous tense*, *future continuous tense*, *present perfective continuous tense* dan *past perfective continuous tense*. Penghurai *Link Grammar* akan menghurai ayat untuk mendapatkan kolerasi label berasaskan corak ayat seperti Ss atau Sp bagi *simple present tense* ayat jenis aktif. Ss dan Sp membawa maksud penyambungan antara kata nama subjek dengan kata kerja. Pengguna diberi pilihan untuk memilih setiap jenis ayat yang mempunyai corak (*pattern*) tertentu menerusi menu pilihan dan sistem akan memaparkan dalam bentuk data multimedia (Yin Hon Wang 2001).

Rule-Based Style and Grammar Checker (Naber 2003) merupakan satu sistem pengesanan tatabahasa bahasa Inggeris. Sistem ini menggunakan teknik berasaskan peraturan semasa menjalankan proses pengesanan ralat tatabahasa. Dalam pendekatan ini, sistem menjana senarai teks dan memberi maklumbalas sekiranya terdapat ralat dalam teks tersebut. Setiap perkataan dalam teks ditanda dengan setiap label kelas kata (*part-of-speech*) dan jenis pecahan (*chunk*) seperti kata nama atau kata kerja. Pemandan peraturan dibuat semasa proses penghurian. Sekiranya ralat yang wujud sepadan dengan peraturan yang dibuat, maka sistem akan memaparkan ralat dan memberikan penjelasan kepada pengguna. Jenis ralat yang diambil kira oleh sistem ialah ralat *agreements*, *missing words*, *extra words*, *wrong words*, *confusion of similar words*, *wrong word order*, *comma errors* dan *whitespace errors*. Sebanyak 56 peraturan ralat tatabahasa dan 6 peraturan gaya ayat dibangunkan untuk mengesan ralat tatabahasa. Dalam ayat di bawah, jujukan label NN1 sebanyak dua kali akan dianggap sebagai *singular noun phrase* oleh sistem. Kata kerja *are* yang salah dalam ayat dilabelkan sebagai VBB (*simple present tense*) akan menghasilkan corak peraturan *\_NP VBB*. Rajah 1 menunjukkan pepadanan peraturan mengenai ralat *subject verb agreement* yang dimasukkan dalam fail XML.

Contoh ayat : The baseball team are established.

Ayat yang di labelkan: The baseball team are established.

AT0 NN1 NN1 VBB VVN

```
<rule id="_NP_INF" name="Subject/Verb agreement">
<pattern
lang="en">_NP(VDB|VDI|VHB|VHI|VBI|VBB|VVI|VVB)</pattern>
<message>Use a singular verb for singular
nouns.</message>
<error_rate warnings="248" sentences="75900" />
<example type="correct">A baseball team <em>is</em>
established.</example>
<example type="incorrect">A baseball team <em>are</em>
established.</example>
</rule>
```

RAJAH 1. Pepadanan peraturan dalam XML

Sentence Analyzer and Viewer (Amoah et al. 2006) adalah satu sistem yang dibina untuk mengesan ralat tatabahasa dalam ayat. Matlamat utama sistem ini dibangunkan untuk pelajar *ESL* yang ingin mengukuhkan kemahiran menulis bahasa Inggeris dengan betul. Fokus utama sistem adalah mengesan ralat *agreement* klausa dan ralat *agreement* frasa nama sesuatu ayat. Kedudukan ralat *agreement* klausa dalam ayat adalah antara subjek dan kata nama. Kedudukan ralat *agreement* frasa nama pula adalah antara *articles* dan *modified nouns*. Contoh ayat yang mengandungi ralat pada peringkat klausa ialah *Rose kiss John* dan pada peringkat frasa ialah *an apples*. Sistem ini digabungkan dengan satu penghurai iaitu *Definite Clause Grammar* (DCG) menggunakan bahasa Prolog. Setiap ayat input yang dimasukkan ke dalam sistem akan dibandingkan dengan peraturan penghurai DCG dan peraturan *agreement*. Sistem akan memberikan maklumbalas sekiranya terdapat ralat *agreement* dalam ayat input.

English Grammar Checking (Kumar & Nair 2007) merupakan satu sistem yang menggunakan pendekatan berasaskan sistem imun manusia (*Artificial Immune System* (AIS)). Pengesanan ralat tatabahasa yang diadaptasikan daripada konsep AIS yang boleh mengesan sebarang entiti bahaya di dalam badan manusia. Sebarang entiti di luar korpus dianggap sebagai ralat. Korpus yang digunakan ialah Reuters-21578 yang telah dilabelkan kelas katanya menggunakan penghurai Montylingua. Dapatan kajian menunjukkan hanya lapan jenis ralat tatabahasa yang dapat dikenalpasti oleh sistem iaitu *subject verb agreement*, *choice of tenses*, *article misuse*, *wrong pronouns*, *wrong number*, *wrong adverbs*, *wrong adjectives* dan *misusing verb*. *Self* dianggap sebagai korpus yang telah dilabelkan kelas katanya. Secara analoginya, korpus dianggap seperti badan manusia dan sama seperti sel badan atau *self cell*. Struktur *Self* terdiri daripada jenis ayat yang telah dilabelkan kelas katanya dan terdiri daripada tiga peringkat iaitu *bigrams*, *trigrams* dan *tetagrams* seperti ayat di bawah:

Official/NNS were/VBD not/RB immediately/RB available/JJ for/IN coment/NN. /.

Bigrams: <NNS,VBD>, <VBD,RB>...

Trigrams: <NNS,VBD,RB>, <VBD,RB,RB>...

Tetagrams:<NNS,VBD, RB, RB>, <VBD,RB,RB, JJ>...

Ketiga peringkat ayat di atas akan diuji ke atas antibodi yang dibangunkan. Teknik algoritma pemilihan negatif (*Negative Selection*) dijalankan dan dianggap sebagai antibodi (*antibody*) badan manusia yang dapat mengenalpasti *antigen* (dianggap ralat tatabahasa) dalam sesuatu ayat.

Kesemua sistem pengesanan ralat esei menggunakan teknik pemprosesan bahasa tabii semasa mengesan ralat tatabahasa kecuali sistem English Grammar Checking yang berkonsepkan AIS. Pengesanan ralat antara kata nama dengan kata kerja dijalankan oleh sistem Grades, Rule Based Style and Grammar Checker dan Sentence Analyzer and Viewer. Didapati hanya sistem Rule Based Style and Grammar Checker yang dapat mengesan ralat daripada aspek tatabahasa dan gaya. Pemanduan peraturan berasaskan sintaktik bahasa Inggeris semasa mengesan ralat digunakan secara meluas oleh sistem Grades dan Rule Based Style and Grammar Checker.

Sistem Rule Based Style and Grammar Checker menunjukkan satu set peraturan pengesanan ralat tatabahasa. Namun bagi kesemua sistem yang dibincangkan, tidak terdapat sebarang penilaian secara formal (*formal evaluation*) atau pengujian yang dilakukan untuk menunjukkan keberkesanan sistem terutamanya penggunaannya dalam aplikasi sebenar (*real world application*). Sistem E-rater menggunakan heuristik APA dan peraturan dalam menjalankan analisis struktur hujahan dan analisis sintaktik manakala sistem Grades dan sistem Rule Based Style and Grammar Checker menggunakan sepenuhnya peraturan semasa mengesan ralat tatabahasa. Penggunaan teknik sama ada berasaskan heuristik ataupun peraturan sintaktik bahasa Inggeris dilihat dapat membantu dalam proses mengesan ralat tatabahasa esei pelajar.



Walaupun pelbagai alatan dibangunkan bagi memudahkan pengesanan ralat dan pemeriksaan esei pelajar, namun alatan tersebut tidak dibangunkan khusus untuk pelajar *ESL* Malaysia. Satu alat pengesanan ralat untuk mengesan ralat tatabahasa dalam penulisan *ESL* amat diperlukan. Menyedari kepentingan pelajar untuk mengenalpasti ralat tatabahasa dalam esei yang ditulis, satu prototaip perisian ATDEiT™ dibangunkan. Kajian ini adalah merupakan kesinambungan kajian lepas yang telah dicadangkan oleh Saadiyah Darus (2005). Analisis empat ratus (400) esei pelajar *ESL* di Universiti Kebangsaan Malaysia (UKM) menunjukkan kesalahan tatabahasa yang agak ketara ialah nahu (Saadiyah Darus et al. 2007). Berdasarkan hasil analisis ini, teknik berasaskan heuristik dan peraturan berserta algoritma untuk mengesan ralat tatabahasa khususnya ralat nahu dalam esei pelajar dibangunkan.

## HEURISTIK UNTUK MENGESAN RALAT NAHU

Menurut Kamus Dewan (2005), heuristik membantu seseorang mempelajari, mengetahui sesuatu perkara melalui pengalaman sendiri. Heuristik mewakili satu andaian yang tidak tentu (Tjoa & Berger 1993), sering dipandu oleh akal, untuk menyediakan penyelesaian yang baik tetapi tidak semestinya satu penyelesaian yang optimum bagi sesuatu masalah yang sukar (Zanakis & Evans 1981). Penyelidikan yang menggunakan heuristik untuk mengesan ralat tatabahasa atau memeriksa esei pelajar adalah sukar didapati. Penyelidikan lepas yang mencadangkan penggunaan heuristik dan peraturan dalam pembangunan sistem pemeriksaan esei berbantuan komputer adalah heuristik APA dan peraturan (Burstein et al. 2002) untuk Sistem E-rater. Burstein et al. (2002) mencadangkan penggunaan esei diperlukan untuk menggunakan heuristik *APA dan peraturan* dalam mengenalpasti struktur sintaksis sesuatu esei.

Dalam penyelidikan ini, heuristik yang dicadangkan merupakan kaedah baru dan dibina berdasarkan pemerhatian terhadap penulisan esei bahasa Inggeris oleh pelajar *ESL* di Universiti Kebangsaan Malaysia. Kajian Nur Asma et al. (2007) menunjukkan kesemua set heuristik yang telah dibangunkan. Dalam kajian Nur Asma et al. (2008), beberapa contoh aplikasi heuristik dipaparkan. Heuristik yang dicadangkan adalah berasaskan sintaksis dan konteks tertentu.

### **Heuristik untuk mengesan ralat *simple present tense 1*:**

Jika kata nama atau kata ganti nama dalam set *present tense 1* wujud, periksa kata kerja. Jika kata kerja dilabelkan VB, maka ayat tersebut tidak mempunyai kesalahan tatabahasa.

Contoh ayat : *Cats like milk.*

Ayat yang dilabelkan : *Cats/NNP like/IN milk/NN ./.*

### **Heuristik untuk mengesan ralat *simple present tense 2*:**

Jika kata nama atau kata ganti nama dalam set *present tense 2* wujud, periksa kata kerja. Jika kata kerja dilabelkan VBZ, maka ayat tersebut tidak mempunyai kesalahan tatabahasa.

Contoh ayat : *Aziz calls Izah everyday.*

Ayat yang dilabelkan : *Aziz/NN calls/VBZ Izah/NNP everyday/JJ ./.*

### **Heuristik untuk mengesan ralat *simple present tense 3*:**

Jika kata nama atau kata ganti nama dalam set *present tense 3* wujud diikuti perkataan *do*, *do not* dan *don't*, periksa kata kerja. Jika kata kerja dilabelkan VB, maka ayat tersebut tidak mempunyai kesalahan tatabahasa.

Contoh ayat : *You do not speak English.*

Ayat yang dilabelkan : *You/PRP do/VBP not/RB speak/VB English/NNP /./*

### **Heuristik untuk mengesan ralat *simple future tense 1*:**

Jika kata nama atau kata ganti nama dalam set *future tense 1* diikuti dengan set *modal* wujud, periksa kata kerja. Jika kata kerja dilabelkan VB, maka ayat tersebut tidak mempunyai kesalahan tatabahasa.

Contoh ayat : *The teacher will dance well.*

Ayat yang dilabelkan : *The/DT teacher/NN will/MD dance/VB well/RB ./*

### **Heuristik untuk mengesan ralat *present progressive tense 2*:**

Jika kata nama atau kata ganti nama dalam set *present progressive tense* diikuti dengan *is* wujud, periksa kata kerja. Jika kata kerja dilabelkan VBG, maka ayat tersebut tidak mempunyai kesalahan tatabahasa.

Contoh ayat : *Mary is reading the newspaper.*

Ayat yang dilabelkan : *Mary/NNP is/VBZ reading/VBG the/DT newspaper/NN /./*

### **Heuristik untuk mengesan ralat *present perfect progressive tense 2*:**

Jika kata nama wujud dalam set *present perfect progressive tense* diikuti *have been*, periksa kata kerja. Jika kata kerja dilabelkan VBG, maka ayat tersebut tidak mempunyai kesalahan tatabahasa.

Contoh ayat : *Your brothers have been waiting since 3.00 p.m.*

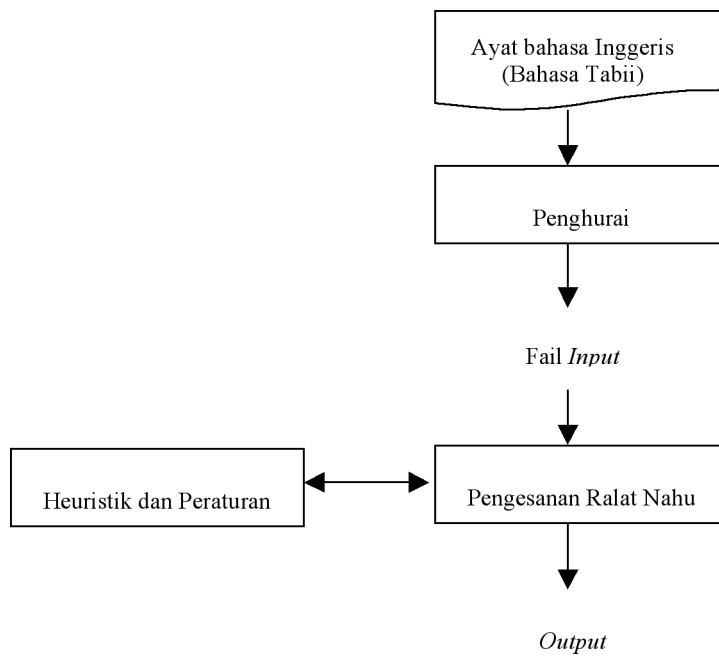
Ayat yang dilabelkan : *Your/PRP brothers/NNS have/VBP been/VBN waiting/VBG since/IN 3.00/CD p.m/NN /./*

## IMPLEMENTASI ATDEiT™

Proses yang terlibat dalam pengesanan ralat esei pelajar *ESL* oleh ATDEiT™ dapat dilihat dalam Rajah 2. Proses dimulakan dengan membaca teks input yang terdiri daripada ayat bahasa Inggeris. Untuk tujuan ini, satu penghurai diperlukan untuk menghuraikan ayat bahasa Inggeris ini kepada label kelas perkataan atau golongan kelas kata (*Part-of-speech – POS tagger*). Penghurai yang digunakan ialah CST Part Of Speech Tagger (2009) di [http://cst.dk/online/pos\\_tagger/uk/index.html](http://cst.dk/online/pos_tagger/uk/index.html). Ayat dilabelkan ini seterusnya akan dimasukkan ke dalam ATDEiT™. ATDEiT™ akan mengesan ralat nahu yang wujud di dalam ayat berdasarkan heuristik dan peraturan yang telah dibangunkan. *Output* akhir dipaparkan dalam bentuk yang mesra pengguna dan pelajar dapat melihat ralat nahu yang dikesan dan ditanda oleh ATDEiT™. Maklum balas daripada ATDEiT™ dapat meningkatkan kefahaman pelajar tentang kesalahan tatabahasa nahu yang dilakukan.

## ALGORITMA

Sebahagian daripada algoritma untuk mengesan ralat nahu *simple present tense* ditunjukkan dalam Rajah 3. Rajah 4 menunjukkan sebahagian daripada algoritma untuk mengesan ralat nahu *simple past tense*. Dalam algoritma ini, 'check' membawa maksud keadaan pemadanan antara ciri yang ada dalam heuristik dengan ayat input. Setiap perkataan dalam ayat input akan diperiksa oleh ATDEiT™.



RAJAH 2. Aliran proses pengesanan ralat nahu esei oleh ATDEiT™

1. FOR each sentence in parsed and tagged text file, DO;
2. FOR each sentence, DO;
  - 2.1 IF word is of type 'noun' DO;
    - IF noun is in the set of heuristic **Simple Present 1**,
    - IF noun is tagged as NNS, check the immediate verb after the noun.
    - IF verb is tagged as VBP or VB, print the following sentence: The sentence has no grammatical error.
    - IF noun is tagged as PRPplural, check the immediate verb after the noun.
    - IF verb is tagged as VBP or VB, print the following sentence: The sentence has no grammatical error.
    - IF noun is in the set of heuristic **Simple Present 2**,
    - IF noun is tagged as NN, check the immediate verb after the noun.
    - IF verb is tagged as VBZ, print the following sentence: The sentence has no grammatical error.

RAJAH 3. Sebahagian algoritma untuk mengesan ralat *simple present tense*

3. FOR each sentence, DO,
  - 3.1 FOR each noun, DO,
    - IF noun is in the set of heuristic **Past Tense 1 with or without Set T2**,
    - IF noun is tagged as NN, check the immediate verb.
    - IF verb is tagged as VBD, print the following sentence: The sentence has no grammatical error.
    - IF noun is tagged as NNP, check the immediate verb.
    - IF verb is tagged as VBD, print the following sentence: The sentence has no grammatical error.

RAJAH 4. Sebahagian algoritma untuk mengesan ralat *Simple Past Tense*

Untuk memahami dengan lebih mendalam proses yang berlaku semasa pepadanan antara kriteria heuristik, perhatikan senario ‘Teenagers’ di bawah:

*“Nowadays, we heard a lot of infants abandonment news being revealed in the newspapers in Malaysia. Teenagers who have busy parents will found that they actually have freedom to do anything be it legally or illegally, without their parents knowing what their teenage children do while they are busy at work. The saddest thing is pornographic reading materials and videos are sold illegally all around Malaysia which encouraging teenagers to purchase them. The other effect of illegal sex is that it can leads to another bigger social problem which is prostitution.”*

Pemprosesan dalam ATDEiT™ ini bermula dengan membaca ayat pada baris pertama. Daripada baris pertama, sistem mengenal pasti kata ganti nama *we* diikuti oleh kata kerja *heard*. ATDEiT™ dapat mengenalpasti kesalahan tatabahasa nahu dalam ayat tersebut berdasarkan pepadanan heuristik *Simple Present Tense* pada perkataan *heard*. ATDEiT™ juga dapat mengenalpasti perkataan *nowadays* yang termasuk dalam set heuristik *simple present tense*. Kemudian, dalam ayat kedua, ATDEiT™ mengenalpasti kesalahan tatabahasa nahu pada ayat *found* berdasarkan pepadanan bentuk *tenses* yang tidak sepadan antara ayat pertama dan ayat kedua. Keputusan untuk setiap ayat akan dijana setelah semua ayat dibaca oleh ATDEiT™. Rajah 5 menunjukkan keputusan daripada senario ‘Teenagers’.

```
Nowadays , we heard a lot of infants abandonment news
being revealed in the newspapers in
Malaysia

-->This sentence has grammatical error at word : heard
-->This sentence has grammatical error at word : being
-->This sentence has grammatical error at word : revealed
##
Teenagers who have busy parents will found that they
actually have freedom to do anything be it legally or
illegally , without their parents knowing what their
```

RAJAH 5. Keputusan ralat yang dipaparkan oleh ATDEiT™

Perbandingan keputusan yang diberikan oleh ATDEiT™ dan keputusan yang diperiksa oleh pemeriksa dapat dilihat dalam Jadual 3. Dalam konteks kajian ini, pemeriksa esei merupakan pakar yang mempunyai pengalaman mengajar bahasa Inggeris lebih sepuluh tahun di institusi pengajian tinggi. ATDEiT™ dapat memeriksa dengan betul kebanyakan ralat nahu dalam data ujian. Namun, hanya terdapat dua kesilapan semasa mengesan ralat nahu. Hasil pemerhatian menunjukkan ATDEiT™ menunjukkan kesilapan semasa memeriksa ayat pasif. Ini disebabkan ATDEiT™ hanya menjana perkataan yang sesuai dan sepadan dengan heuristik dan peraturan yang dibina untuk ayat aktif sahaja.

JADUAL 3. Perbandingan keputusan daripada ATDEiT™ dan pemeriksa

Ayat	Ralat dikesan ATDEiT™	Ralat dikesan pemeriksa
Pertama	<i>heard, being revealed</i>	<i>heard</i>
Kedua	<i>found</i>	<i>found</i>
Ketiga	<i>sold</i>	
	<i>encouraging</i>	
Keempat	<i>leads</i>	<i>leads</i>

### KEPUTUSAN EKSPERIMEN

Satu pengujian dijalankan untuk menguji keberkesanan heuristik yang dibina dan seterusnya menguji ketepatan ATDEiT™. Tiga output yang diambil kira ialah ralat yang dikesan betul, ralat yang dikesan salah dan ralat yang tidak dikesan. Teknik yang digunakan dalam kajian ini berdasarkan kaedah pengujian dalam *Information Extraction* iaitu dapatan semula (*recall*) dan kejitian (*precision*) (Jurafsky & Martin 2000). Dapatan semula adalah peratus ketepatan yang wujud dalam maklumat yang dihasilkan oleh sistem. Kejitian merupakan peratus ketepatan maklumat yang dihasilkan.

#### DAPATAN SEMULA

Dapatan semula adalah peratus ketepatan yang wujud dalam maklumat yang dihasilkan oleh sistem. Maklumat yang tepat tersebut kemudiannya dibandingkan dengan keputusan yang dihasilkan oleh pakar atau jawapan sebenar. Berikut merupakan formula piawai yang digunakan untuk mengira dapatan semula:

$$\text{Dapatan Semula} = \frac{N_{\text{Betul}}}{N_{\text{Jawapan_Sebenar}}}$$

Berdasarkan formula di atas, bilangan betul yang dihasilkan oleh ATDEiT™ diwakilkan dengan  $N_{\text{Betul}}$  dan  $N_{\text{Jawapan_Sebenar}}$  pula adalah merujuk kepada bilangan yang dihasilkan oleh pakar atau jawapan sebenar.

#### KEJITUAN

Kejitian adalah peratus ketepatan maklumat yang telah dihasilkan. Kajitian menggambarkan ketepatan sistem dalam mendapatkan keputusan yang tepat. Berikut merupakan formula piawai yang digunakan untuk mengira kejitian:

$$\text{Kejitian} = \frac{N_{\text{Betul}}}{N_{\text{Betul}} + N_{\text{Salah}}}$$

ATDEiT™ diuji menggunakan 50 set data ujian esei. ATDEiT™ mencapai keputusan yang memberansangkan dengan 95% dapatan semula dan 76.2% kejitian. Keputusan ini menunjukkan potensi penggunaan heuristik dan peraturan untuk pengesanan ralat ke atas esei bahasa Inggeris adalah mengujakan. Keputusan ini menandakan bahawa pembangunan sintaktik heuristik dan peraturan sesuai dijadikan asas pada peringkat permulaan dalam pembangunan sistem *CBEM*.

## KESIMPULAN DAN CADANGAN MASA DEPAN

Kajian ini telah mencadangkan pembangunan heuristik dan peraturan untuk pengesanan ralat nahu dan direalisasikan dengan sebuah alatan yang dinamakan ATDEIT™. Berdasarkan pemerhatian kepada struktur ayat dalam penulisan esei *ESL* dan tatabahasa Inggeris itu sendiri, kaedah heuristik dan peraturan ini boleh dikembangkan lagi. Heuristik dan peraturan yang dibangunkan hanya meliputi ayat aktif dan berjenis mudah (*simple sentence*). Maka pembangunan heuristik dan peraturan dapat dikembangkan dengan menambah lagi heuristik dan peraturan untuk ayat jenis pasif dan ayat jenis *compound sentence*, *complex sentence* dan *compound-complex sentences*. Selain jenis ralat nahu *tenses*, beberapa jenis ralat tatabahasa seperti *articles*, *prepositions* dan *subject verb agreement* boleh diambil kira pada masa hadapan.

## PENGHARGAAN

Penyelidikan ini dibiayai oleh Kementerian Sains, Teknologi dan Inovasi, Malaysia (MOSTI). Kod nombor penyelidikan: e-ScienceFund 01-01-02-SF0092.

## REFERENCES

- Amoah, P, Lupiana, R, Ghemri, L. 2006. A sentence analyzer and viewer for Detecting grammatical errors. *Texas Southern University Research Week 2006*.
- Attali, Y. and Burstein, J. 2005. Automated essay scoring with E-rater® V.2.0. Kertas Kerja *Conference of the International Association for Educational Assessment*. Philadelphia, PA.
- Burstein, J. 2003. The E-rater® scoring engine: Automated essay scoring with natural language processing. Dlm. Shermis, M. D & Burstein, J.C. (pnyt.). *Automated essay scoring. A cross-disciplinary perspectives*, hlm. 113-121. London: Lawrence Erlbaum Associates, Inc.
- Burstein, J.C., Braden-Harder, L., S. Chodrow, M., A. Kaplan, B., Kukich, K., Lu, C., A. Rock, D and Wolff, S. 2002. System and method for computer based automatic essay scoring, *United State Patent*.
- Burstein, J., Kukich, K., Wolff, S., Lu, C. & Chodorow, M. 1998. Enriching automated scoring using discourse marking. *Workshop on Discourse Relations and Discourse Marking. Annual Meeting of the Association of Computational Linguistics*, Montreal, Canada.
- Brock, M. N. 1993. Should we do what we can or can we do what we should? Three disk-based text analyzers and the ESL writer. *Journal of Second Language Writing*, 8(5): 51-60.
- Collins, J. L. 1989. Computerized text analysis and the teaching of writing. In *Critical Perspectives on Computers and Composition Instruction*, eds. Hawisher, G.E. & Selfe, C. L, 30-34. New York: Teachers College Press.
- CST's Part Of Speech tagger. (atas talian). [http://cst.dk/online/pos\\_tagger/uk/index](http://cst.dk/online/pos_tagger/uk/index). (1 August 2009).
- Cumming, A. 2001. Learning to write in a second language: Two decades of research. *International Journal of English Studies* 1(2): 1-23.
- Ellis, R. 1985. *Understanding second language acquisition*. Oxford University Press.
- Foo, B & Richard, C. 2004. English in Malaysia. *RELC Journal* 35 (2): 229-240.
- Fox, R. and Bowden, M. 2002. Automated diagnosis of non-native English speaker's natural language. *Proceeding of the 14<sup>th</sup> IEEE International Conference on Tools with Artificial Intelligent (ICTA'02)*, 301-306.
- Haist, C. 2000. An evaluation of Microsoft Word 97's Grammar Checker. *ERIC 438538*, CS 216991, hlm. 1 - 42.
- Hudelson, S. 1989. *Write on: Children writing in ESL*. Englewood Cliffs, NJ: Prentice Hall Regents and Center for Applied Linguistics.
- Jurafsky, D., & Martin, J. H. 2000. *Speech and language processing; An introduction to natural language processing, computational linguistics and speech recognition*. Upper Saddle River, New Jersey: Prentice Hall, Inc.
- Kamus Dewan*. 2005. Ed. Keempat. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Kumar, A. and Nair, S., 2007. An artificial immune system based approach for English grammar checking, *ICARIS 2007, Lecture Notes in Computer Science*, 4628, 348-357.
- Lim Ho Peng. 1976. Errors and error analysis in TESL: The Malaysian Experience, *RELC Journal*, 7 (2): 23-29.

- Myles, J. 2002. Second language writing and research: The writing process and error analysis in student texts. *TESL-EJ Journal*, 6(2): A-1. ISSN 1072-4303.
- Naber, D., 2003. A rule based style and grammar checker, Diplomarbeit Thesis, Universitat Bielefeld.
- Nik Safiah Karim., Farid M. Onn., Hashim Haji Musa & Abdul Hamid Mahmood. 2008. *Tatabahasa dewan*. Edisi ke-3. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Noorizah Mohd Nor & Rosniah Mustaffa. 2002. *Tatabahasa Inggeris*. Edisi ke-5. Kuala Lumpur: Utusan Publications & Distributors Sdn Bhd.
- Nur Asma Mohd Razali, Nazlia Omar, Saadiyah Darus, Tg Nor Rizan Tg Mohd Maasum, Mohd Juzaidin Ab. Aziz. 2007. Pendekatan heuristik dan peraturan untuk pemeriksaan nahu secara automatik bagi penulisan esei bahasa Inggeris sebagai bahasa kedua. *Prosiding Persidangan Kebangsaan Sains Pengaturcaraan Atur '07*, in Malay. Pan Pacific KLIA, 77-88.
- Nur Asma Mohd Razali, Nazlia Omar and Saadiyah Darus. 2008. Heuristics and rule-based approach for automated marking tool for ESL writing. *Proceedings of the International Symposium on Information Technology 2008 (ITSIM 2008)*. Kuala Lumpur Convention Centre, Malaysia, 26-29 August 2008, IEEE Cat. No. CFP0833E-PRT, ISBN 978-1-4244-2327-9, 144-149.
- Pennington, M. C. 1992. Beyond off - The shelf computer remedies for student writers: Alternatives to canned feedback. *System*, 20(4): 423-437.
- Rosniah Mustaffa & Noorizah Mohd Nor. 2008. *Kesilapan tatabahasa Inggeris*. Edisi ke-2. Kuala Lumpur: Utusan Publications & Distributors Sdn Bhd.
- Richards, J. C. 2002. 30 years of Tefl / Tesl: A personal reflection. *RELC Journal*. 33: 1-35.
- Saadiyah Darus. 1999. A prospect of automatic essay marking. International Seminar on Language in the Global Context: Implications for the Language Classroom, RELC, Singapore, 19-21 April.
- Saadiyah Darus. 2005. A framework of a CBEM for ESL writing at institutions of higher learning in Malaysia. Phd Tesis. Universiti Kebangsaan Malaysia.
- Saadiyah Darus., Tg Nor Rizan Tg Mohd Maasum, Siti Hamin Stapa, Nazlia Omar, Mohd Juzaidin Ab Aziz. 2007. Developing an error analysis marking tool for ESL learners. *7th WSEAS International Conference on Applied Computer Science*, 356-259.
- Silva, T & Matsuda, P. K. 2002. Writing. Dlm. N. Schmit. (pnyt.) *An Introduction to Applied Linguistics*, hlm. 251-266. London: Oxford University Press.
- Su Hie Ting. 2002. Is English needed in a Malay workplace? *RELC Journal*. 33:137-153.
- Thiesmeyer, J. 1989. Should we do what we can?. In *Critical Perspectives on Computers and Composition Instruction*. G.E. Hashiwer & C.L.Selfe, eds. 30-43, New York: Teachers College.
- Tjoa, A. M and Berger, L. 1993. Transformation of requirements specifications expressed in natural language into an EER model, *Proceeding of the 12<sup>th</sup> International Conference on Approach*, Airlington, Texas, USA, 206-217.
- Valenti, S., Neri, F., & Chuchiareli, A. 2003. An overview of current research on automated essay grading. *Journal of Information Technology Education*. DIIGA- Universita' Politecnica delle Marce, Ancona, Italy, Vol 2:319-320.
- Vernon, A. 2000. Computerized Grammar Checkers 2000: Capabilities, limitations, and pedagogical possibilities. *Computers and Composition*, 17 (3): 329-349.
- Ying Hong Wang. 2001. E-learning for English based on multimedia database and internet. *Tamkang Journal of Science and Engineering* 4 (1): 19-36.
- Ying, H.W., Chih, H.L., 2003. Two intelligent applications supports English distance learning. *Proceedings of the 23<sup>rd</sup> International Conference on Distributed Computing Systems Workshop*.
- Zanakis, S.H. and Evans, J. R. 1981. Heuristics 'Optimization': Why, when and how to use it, *Interfaces*, 11(5): 84-91.

Nur Asma Mohd Razali @ Mohd Rozly  
 Jabatan Sains Komputer  
 Fakulti Teknologi dan Sains Maklumat  
 43600 UKM Bangi, Selangor.  
 nur\_asma84@yahoo.com

Nazlia Omar  
Jabatan Sains Komputer  
Fakulti Teknologi dan Sains Maklumat  
43600 UKM Bangi, Selangor.  
no@ftsm.ukm.my

Saadiyah Darus  
Pusat Pengajian Bahasa dan Linguistik  
Fakulti Sains Sosial dan Kemanusiaan  
43600 UKM Bangi, Selangor.  
adi@ukm.my