

# Acquiring Extended Units of Meaning: The Role of Learner Corpus in Teaching Indonesian as A Foreign Language

Totok Suhardijanto<sup>1</sup> and Deodatus Perdana Putra<sup>2</sup>

<sup>1</sup>Departemen Linguistik, Universitas Indonesia, Depok, Indonesia

<sup>2</sup>Program Studi Linguistik, Universitas Indonesia, Depok, Indonesia

<sup>1</sup>totok.suhardijanto@ui.ac.id, <sup>2</sup>deodatus.perdana@ui.ac.id

## ABSTRACT

One of the best ways to comprehend how people learn and acquire a foreign language is to look at what they produce in the target language. In this perspective, learner corpora have a special contribution to realize our goal. In this paper, we explore the relationship between learner corpus and teaching Indonesian as a foreign language. Data in this research is derived from Indonesian Learner Corpus (Korpus BIPA). At the moment, the Indonesian Learner Corpus consists of more than 2 million words. However, in this research, we only used learner sub-corpora which contains 815.470 words. This research employed method and techniques from corpus linguistics to depict how Indonesian learners acquire and master Indonesian. Learner corpus is a big thing, but in this research, we limit our discussion on how Indonesian learner acquire and master extended units of meaning in Indonesian which represent the concept of comfort, such as collocations with *baik* and *bagus*. This issue is related to a phenomenon called as an extended unit of meaning. The result shows that the student ability in differentiating the use of those words grew side by side with the increase of student competence levels.

**Keywords:** *teaching Indonesian as foreign language, learner corpus, collocation, phraseology, extended unit of meaning*

## 1. PENDAHULUAN

Dalam pendekatan tradisional, tiap kata dianggap memiliki satu atau lebih makna yang dapat dipilih, lalu disusun untuk membangun pesan atau pertanyaan dalam bentuk kalimat [2]. Dalam pengajaran bahasa, khususnya Pengajaran Bahasa Indonesia bagi Penutur Asing (BIPA), pandangan seperti itu masih banyak dianut sehingga dalam proses pembelajaran, pengajar masih sering mengajarkan kata sebagai batas satuan makna. Jadi, seolah-olah pada sebuah kata, maknanya dapat beragam sehingga disebut *polisemi* dan dapat saja makna tertentu justru dapat muncul pada dua atau lebih kata, yang kemudian menampilkan fenomena *sinonimi*. Jadi, selain kata tunggal (simpleks), gabungan kata merupakan hal yang juga penting dalam sebuah bahasa [3]. Pada bentuk seperti *kuda hitam* dan *argo kuda*, seorang penutur bahasa Indonesia yang mengetahui makna masing-masing untuk *kuda*, *hitam*, dan *argo* akan

membutuhkan status leksikal gabungan kata pada mental leksikon yang membentuk makna semantisnya tersendiri untuk memahami arti *kuda hitam* dan *argo kuda*.

Dengan mengetahui makna gabungan kata atau kata dalam konteksnya, proses pemahaman makna juga akan memberikan hasil yang akurat dan tepat. Perhatikan contoh berikut ini,

- (1) Saya bisa merekomendasikan tempat yang *bagus* di Indonesia.
- (2) Gagasan itu sangat *bagus* dan baru.
- (3) Lebih mahal daripada mie ayam, tetapi rasanya *bagus*.

Pada contoh (1), kata *bagus* muncul beriringan dengan kata *tempat*. Sementara itu, kata *bagus* pada contoh (2) muncul bersama dengan kata *gagasan* dan pada contoh (3) bersama dengan kata *rasa*. Bagi penutur jati bahasa Indonesia, kombinasi kata *bagus* dan *tempat* jelas merupakan kombinasi yang natural dan lazim. Begitu pula, kombinasi kata *bagus* dan *gagasan*. Pada kedua contoh tersebut, kata *bagus* dapat menjelaskan atau mendampingi kata yang bersifat konkret maupun abstrak. Namun, kombinasi kata *bagus* dan *rasa* jelas bukan merupakan pasangan yang natural. Ada kejanggalan dalam pemakaian kedua kata tersebut secara beriringan. Jadi, tampak di sini bahwa kata *bagus* merupakan kata yang berpolisemi, yakni memiliki gradasi, ketidakjelasan batasan, dan ketumpangtindihan.

Dalam kajian linguistik mutakhir, khususnya sejak perkembangan linguistik korpus modern, pendekatan terhadap makna telah banyak berubah. Makna tidak lagi dianggap inheren di dalam sebuah kata. Makna ditentukan bukan hanya oleh satu kata, melainkan juga oleh jejaring kata yang melingkunginya, meminjam istilah Firth, *the company it keeps* [4]. Dalam linguistik kognitif, makna kata memiliki jaringan semantik [5]. Sementara itu, Sinclair mengembangkan konsep Firth dengan mengeksplorasi gagasan satuan makna diperluas (*extended units of meaning*). Konsep ini berawal dari gagasan awal Sinclair tentang prinsip idiom, yakni gagasan bahwa pengguna bahasa secara rutin memanfaatkan khazanah yang disebutnya “semi-preconstructed phrases that constitute single choices” [6]. Prinsip yang sama disebut pula sebagai gugus leksikal (*lexical bundles*) [7].

Bagaimana pengaruh gugus leksikal tersebut dalam pembelajaran bahasa? Adakah para pemelajar menggunakan gugus leksikal dalam produksi bahasa yang mereka lakukan pada saat pembelajaran? Beberapa penelitian telah dilakukan di Indonesia, tetapi sasarannya pemelajar bahasa Inggris [8], [9], [10], [11]. Oleh sebab itu, dalam makalah ini, akan dieksplorasi lebih lanjut bagaimana pemelajar BIPA memperoleh kosakata dalam proses belajar bahasa Indonesia, khususnya kosakata yang berpolisemi, yakni *baik* dan *bagus* sebagaimana yang tampak dalam korpus pemelajar BIPA. Lebih lanjut, permasalahan tersebut dapat dipertajam dengan beberapa pertanyaan berikut. Pertama, bagaimana korpus pemelajar BIPA menunjukkan distribusi pemakaian satuan makna diperluas oleh pemelajar BIPA? Kedua, bagaimana pemelajar BIPA menggunakan kosakata berpolisemi, khususnya *baik* dan *bagus*?

## 2. METODE PENELITIAN

Penelitian ini menggunakan data korpus sebagai alat pembuktian terhadap permasalahan yang menjadi fokus penelitian. Korpus merupakan kumpulan teks yang disusun sedemikian rupa untuk tujuan merepresentasikan peristiwa kebahasaan pada sebuah dialek, variasi, atau ranah komunikasi tertentu [1]. Korpus bukan hanya kumpulan teks, melainkan ‘a collection of naturally occurring language texts, chosen to characterize a state or variety of a language’ [6]. Jadi, korpus merupakan koleksi teks yang disusun dengan desain tertentu [12]. Data dalam penelitian ini diperoleh dari korpus BIPA Universitas Indonesia yang disusun

dengan data tahun 2010-2017. Korpus tersebut berjumlah 2 juta token, namun penelitian ini hanya menggunakan subkorpus BIPA Universitas Indonesia yang memuat 815.470 kata.

Dalam analisis, diterapkan beberapa metode yang lazim digunakan dalam linguistik korpus. Dari metode-metode yang lazim dalam linguistik korpus [12], yang dimanfaatkan dalam kajian ini adalah frekuensi kata, konkordansi, dan kolokasi karena ketiganya berperan dalam menunjukkan bagaimana distribusi dan perilaku kata-kata berpolisemi, khususnya *baik* dan *bagus*. Pada analisis kolokasi, digunakan MI (*mutual information*) yang membuktikan apakah kekerapan kemunculan dua kata berpasangan terjadi secara kebetulan atau tidak [13]. Jika tidak kebetulan, kemunculan tersebut berarti signifikan. Dalam penelitian ini, ambang batas MI adalah sembilan dan dengan frekuensi sebanyak lima kali. Untuk lebar distribusi (*widexpan*) kolokasi, digunakan lebar 2 (dua) kata. Ini berarti bahwa kombinasi itu terdiri atas kosakata poros (*node*) dan dua kosakata pendamping masing-masing di sebelah kiri dan kanan.

Selain menggunakan kolokasi, dalam linguistik korpus juga dapat digunakan analisis ungkapan multikata (*multiword expression*) untuk menggali perilaku semantis kata. Untuk menjaring ungkapan multikata dari korpus, digunakan teknik n-gram, yakni rangkaian berurutan unsur-unsur berjumlah *n* dari sebuah sampel teks atau ujaran [14]. Penjaringan ungkapan multikata dilakukan dengan menggunakan perhitungan *Dice*, *Jaccard*, dan *Pointwise Mutual Information (PMI)*. Formula *Dice* mempunyai kalkulasi, yaitu probabilitas merupakan kemungkinan hasil yang diharapkan dibagi kemungkinan hasil yang mungkin muncul [13]. Sementara itu, indeks *Jaccard* dihitung sebagai kesamaan di antara set sampel terbatas dan didefinisikan sebagai ukuran irisan dibagi dengan ukuran gabungan set sampel yang ada [15]. Kemudian, *Pointwise Mutual Information* adalah penghitungan asosiasi yang diperoleh melalui kalkulasi peristiwa kemunculan pasangan dua kata *x* dan *y* sebagai kebetulan dan/atau distribusi bersama [13]. Karena dalam bahasa Indonesia pada umumnya makna pada ungkapan multikata terwujud dalam rangkaian dua hingga tiga kata, pada penelitian akan digunakan trigram, atau gram dalam bentuk tiga unsur kata, seperti *defisit transaksi berjalan* atau *oleh karena itu*.

### 3. HASIL DAN PEMBAHASAN

Berdasarkan analisis kolokasi kata *bagus* pada data Korpus BIPA dengan perhitungan MI di atas 9 dan lebar distribusi 2, diperoleh hasil yang dapat dilihat pada Tabel 1 berikut ini. Yang ditampilkan pada tabel berikut hanya kata penuh dengan kelas kata nomina, pronominal, verba, dan adjektiva. Kelas kata lainnya tidak ditampilkan dengan alasan bahwa keempat kelas kata tersebut merupakan kolokat yang dapat langsung memperlihatkan preferensi semantis dari sebuah kata [16].

Pada Tabel 1, terdapat informasi mengenai posisi kemunculan kata kolokat. Jika tertulis L, hal itu berarti bahwa kolokat tersebut muncul sebelum atau di sebelah kiri kata yang menjadi porosnya. Misalnya, pada gabungan kata *jam tangan emas*, kata *jam* terletak sebelum atau di sebelah kiri kata *emas* yang menjadi poros. Jika tertulis R, itu berarti bahwa kolokat muncul setelah atau di sebelah kanan kata yang menjadi poros. Kemudian, jika tertulis M, hal itu berarti bahwa kolokat tersebut muncul baik di sebelah kanan atau kiri kata yang menjadi porosnya.

**Tabel 1 Kolokat Signifikansi Kata *Bagus* pada Korpus BIPA (MI<sub>3</sub>>9, frek>5, dan span L/R= 2)**

Indeks	Position	Collocate	Stat	Freq Coll	Freq Corp
1	L	kualitasnya	15.2871	10	25
2	L	BIPA	15.13383	59	5710
3	L	ide	13.67104	11	102
4	L	fasilitasnya	12.91271	6	28
5	L	pantai	12.69255	10	151
6	L	nilai	12.53864	10	168
7	R	belajar	11.83257	26	4817
8	M	hotel	11.56429	8	169

Dari data di atas diperoleh gambaran bagaimana kata *bagus* digunakan oleh pemelajar BIPA. Secara garis besar, kata *bagus* dikaitkan dengan kata-kata berkelas nomina seperti *kualitas, ide, BIPA, fasilitas, pantai, nilai, hotel, akting, restoran, film, tempat*, dan sebagainya. Muncul satu kata yang sangat terikat konteks, yaitu *BIPA* karena korpus ini sebagian besar disusun dari teks yang dihasilkan pemelajar BIPA di lingkungan Universitas Indonesia. Sementara itu, untuk verba atau kata kerja, kata *bagus* secara signifikan muncul bersama dengan kata *belajar*. Bentuk yang muncul pada umumnya berpola *bagus untuk belajar*. Beberapa contoh kemunculan kata *bagus* dan kedua verba tersebut di antaranya adalah

... BIPA kepada Anda sebagai tempat yang	<b>bagus</b>	untuk belajar bahasa Indonesia. Karena ...
Dalam pelajaran dan tugas menulis ulasan	<b>bagus</b>	untuk <i>belajar</i> bahasa Indonesia untuk ...

Sementara itu, kata sifat atau adjektiva yang secara signifikan muncul bersama dengan kata *bagus* adalah *efektif, menyenangkan*, dan *indah*. Beberapa contoh kemunculan ketiga adjektiva tersebut dengan kata *bagus* dapat dilihat pada contoh berikut ini.

Saya pikir Bank Sampah itu ide yang	<b>bagus</b>	dan <i>efektif</i> untuk mengatasi masalah ...
... saya berharap Dongdeamun yang sangat	<b>bagus</b>	dan <i>menyenangkan</i> ini makin terkenal di ...
... lihat pemandangan Jepang yang paling	<b>bagus</b>	atau <i>indah</i> daripada yang lain. Tapi ini

Dari contoh pemakaian di atas, tampak bahwa adjektiva *efektif, menyenangkan*, dan *indah* yang bermakna positif digunakan untuk mengamplifikasi atau menggandakan kesan positif yang disampaikan oleh kata *bagus*—apalagi hal tersebut diperkuat dengan penggunaan kata hubung *dan* yang bermakna penambahan. Dalam data, tidak ditemukan penggunaan kata *bagus* yang bersanding atau muncul bersama dengan adjektiva bermakna negatif.

Lalu bagaimana dengan ungkapan multikata (*multiword expression*) yang mengandung kata *bagus* yang digunakan para pemelajar BIPA? Dari ungkapan multikata tersebut, dapat diketahui bagaimana dan dalam konteks apa kata *bagus* tersebut digunakan.

**Tabel 2 Ungkapan Multikata dengan Kata *Bagus* dalam Korpus BIPA**

No	Kata 1	Kata 2	Kata 3	Dice	Jaccard	PMI	Total
1	piano	lebih	bagus	0.222	0.143	5.642	6.007
2	lebih	bagus	daripada	0.333	0.25	4.836	5.419
3	bermain	piano	bagus	0.211	0.133	5.527	5.871
4	adalah	hari	bagus	0.167	0.111	2.639	2.917

5	membeli	barang	bagus	0.133	0.08	5.604	5.817
6	restoran	sandwich	bagus	0.133	0.077	5.868	6.078

Dari Tabel 2, ungkapan multikata dengan kata *bagus* pada umumnya menjelaskan aspek kualitas dari sebuah entitas, misalnya seperti *film, fisik, cuaca, restoran, barang, suara*, dan sebagainya. Selain entitas, tindakan yang juga memiliki aspek yang dapat diberi nilai positif atau negatif dapat muncul dengan penanda kualitas seperti kata *bagus*, misalnya *bermain piano, berenang, menulis*, dan sebagainya. Meskipun pada umumnya penggunaan kolokasi kata *bagus* oleh pemelajar BIPA berkesan natural dan lazim, ada beberapa kombinasi tak tepat (*inappropriate collocation*) yang dihasilkan pemelajar dengan level pemula. Beberapa kombinasi tak tepat tersebut adalah *merasa makanan bagus* dan *akan hidup bagus*.

Bagaimana selanjutnya dengan kolokasi kata *baik*? Berdasarkan analisis kolokasi, kata *baik* secara signifikan muncul bersamaan dengan kata-kata berikut sebagaimana yang tampak pada Tabel 3 di bawah ini.

**Tabel 3 Kolokat Signifikan Kata *Baik* pada Korpus BIPA (MI3>9, frek>5, dan span L/R= 2)**

Index	Position	Collocate	Stat	Freq Coll	Freq Corp
1	R	hati	21.44819	137	465
2	R	ramah	17.08784	41	256
3	L	berkomunikasi	15.99282	32	260
4	L	bersifat	15.38478	21	112
5	L	cukup	15.24288	40	854
6	L	merawatnya	15.01518	8	8
7	L	bersikap	13.86318	12	60
8	L	berkelakuan	13.39601	5	6

Berbeda dengan kata *bagus*, dari tabel di atas tampak bahwa kolokat signifikan kata *baik* cukup banyak yang berupa verba, misalnya *senyum, berkomunikasi, bersifat, merawatnya, bersikap, berkelakuan, menjadi*, dan sebagainya. Sementara itu, nomina yang sering muncul dengan kata *baik* adalah *hati, kesan, BIPA, orang, teman*, dan sebagainya. Selanjutnya, adjektiva yang sering muncul dengan kata *baik* adalah *ramah, rajin, benar, sopan*, dan sebagainya.

Beberapa penggunaan verba seperti *menjadi, berkelakuan, bersikap, merawat*, dan *berkomunikasi* yang muncul bersamaan dengan kata *baik* dapat dilihat pada contoh berikut.

... supaya BIPA ini <i>menjadi</i> lebih	<b>Baik</b>	untuk junior-junior saya. Menurut saya, kualitas ...
Tambahan pula, saya <i>berkelakuan</i>	<b>Baik</b>	dan berbadan sehat. Demikianlah surat lamaran ...
... pula, mereka selalu <i>bersikap</i>	<b>baik</b>	kepada peserta. Para dosen yang mudah senyum
... berjanji akan <i>merawatnya</i> dengan	<b>baik.</b>	Mengerti, kan? Tanya kakek. Tentu saja aku ...
... pasti bisa <i>berkomunikasi</i> dengan	<b>baik.</b>	Namun, ada beberapa masalah yang harus ...

Sementara itu, ungkapan multikata dengan kata *baik* dapat dilihat pada Tabel 4. Dari daftar tersebut, menarik untuk diperhatikan bahwa kata *baik* sangat sering muncul dalam bentuk kombinasi *baik hati, baik dan (ramah, rajin, ...), baik untuk (badan, kesehatan, ...), baik-baik (saja)* dan *lebih baik daripada*. Sementara itu, kata *baik* sering digunakan dengan verba yang bersinggungan dengan karakter, misalnya *bersifat baik, bersikap baik*, dan *berkata-kata baik*.

**Tabel 4 Ungkapan Multikata dengan Kata *Baik* dalam Korpus BIPA**

Index	Kata_1	Kata_2	Kata_3	Dice	Jaccard	PMI	Total
1	dia	baik	hati	0.018	0.009	3.753	3.78
2	bicara	yang	baik	0.017	0.009	4.846	4.872
3	kamu	selalu	baik-baik	0.017	0.009	4.486	4.512
4	ayahnya	baik	hati	0.015	0.008	2.791	2.814
5	aku	baik-baik	saja	0.014	0.007	4.768	4.789
6	yang	baik	hati	0.012	0.007	3.647	3.666

#### 4. SIMPULAN

Secara umum, dari analisis berbasis korpus, pemelajar BIPA telah dapat membedakan dan menggunakan kata *bagus* dan *baik*. Dari segi makna pemakaiannya, kata *bagus* lebih sering digunakan untuk menjelaskan kadar dari kualitas, fasilitas, pelayanan, dan hal maujud lainnya seperti restoran, pantai, organisasi, dan sebagainya. Meskipun demikian, kata *bagus* juga dapat digunakan untuk menandai kadar dari nilai, ujian, cuaca, dan performa yang lebih abstrak. Oleh sebab itu, kata *bagus* juga sering dikaitkan dengan adjektiva *efektif* dan *indah*.

Sementara itu, kata *baik* lebih sering digunakan untuk menandai kadar dari karakter, sifat, watak, dan sikap. Karena terkait dengan sifat, kata *bagus* sering muncul terkait dengan adjektiva lain, khususnya seperti *ramah*, *rajin*, *sopan*, dan *benar*. Dari segi struktur, pola *bagus untuk ...* dan *baik untuk ...* merupakan pola yang paling sering digunakan selain—tentu saja—pola perbandingan seperti *lebih baik/bagus/enak daripada ...* yang juga cukup sering muncul.

Dari kajian korpus, juga dapat diketahui bagaimana permasalahan penggunaan kata, khususnya terkait dengan pemahaman makna kontekstual sebuah kata. Dalam korpus pemelajar BIPA itu, ditemukan beberapa kali penggunaan kolokasi tak tepat sebagai akibat kurangnya pemahaman pemelajar terhadap makna penggunaan tiap-tiap kata, khususnya kata *bagus* dan *baik*. Terlihat bahwa kesalahan penggunaan—yang berakibat pada penghasilan kolokasi tak tepat—cenderung terjadi pada peserta BIPA pada tahap awal. Penguasaan mereka terhadap makna dan pemakaian kosakata membuat mereka sering melakukan kesalahan dengan menyusun kombinasi kolokasional yang tidak lazim.

#### REFERENCES

- [1] J. M. Sinclair and R. Carter, *Trust the text : language, corpus and discourse*. Routledge, 2004.
- [2] P. Hanks and M. Oakes, "Flexibility of multiword expressions and Corpus Pattern Analysis," in *Multiword expressions: Insights from a multi-lingual perspective*, 2018, pp. 93–119.
- [3] P. Schone and D. Jurafsky, "Is knowledge-free induction of multiword unit dictionary headwords a solved problem?," *Proc. 2001 Conf. Empir. Methods Nat. Lang. Process.*, 2001.
- [4] J. R. Firth, *Studies in Linguistic Analysis*. Oxford: Blackwell, 1957.
- [5] R. B. Dewell, "Over again: Image-schema transformations in semantic analysis," *Cogn. Linguist.*, 1994.

- [6] J. M. Sinclair, *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.
- [7] D. Biber, "A corpus-driven approach to formulaic language in English," *Int. J. Corpus Linguist.*, 2009.
- [8] R. R. D. Moehkardi, "Grammatical and lexical English collocations: Some possible problems to Indonesian learners of English," *Humaniora*, vol. XIV, no. 1, pp. 53–62, 2002.
- [9] V. Cortes, "Lexical bundles in published and student disciplinary writing: Examples from history and biology," *English Specif. Purp.*, 2004.
- [10] V. Cortes, "Teaching lexical bundles in the disciplines: An example from a writing intensive history class," *Linguist. Educ.*, 2006.
- [11] Y. H. Chen and P. Baker, "Lexical bundles in l1 and l2 academic writing," *Lang. Learn. Technol.*, 2010.
- [12] W. Cheng, *Exploring Corpus Linguistics: Language in Action*. London: Routledge, 2012.
- [13] C. D. Manning, H. Schütze, and G. Weikurn, "Foundations of Statistical Natural Language Processing," *SIGMOD Rec.*, 2002.
- [14] D. Jurafsky and J. H. Martin, "N-Grams," in *Speech and Language Processing*, 2007.
- [15] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu, "Using of jaccard coefficient for keywords similarity," in *Lecture Notes in Engineering and Computer Science*, 2013.
- [16] M. Bednarek, "Semantic preference and semantic prosody re-examined," *Corpus Linguist. Linguist. Theory*, 2008.