



PROSIDING
SEMINAR LEKSIKOGRAFI
INDONESIA



PROSIDING SEMINAR LEKSIKOGRAFI INDONESIA

TANTANGAN LEKSIKOGRAFIS BAHASA-BAHASA DAERAH DI INDONESIA

26—29 Juli 2016

Hotel Santika, Taman Mini Indonesia Indah, Jakarta

Pusat Pengembangan dan Pelindungan
Badan Pengembangan dan Pembinaan Bahasa
Kementerian Pendidikan dan Kebudayaan

PROSIDING INI BELUM DISUNTING

**PROSIDING
SEMINAR LEKSIKOGRAFI INDONESIA**

26—29 Juli 2016

Hotel Santika, Taman Mini Indonesia Indah, Jakarta

**TANTANGAN LEKSIKOGRAFIS
BAHASA-BAHASA DAERAH
DI INDONESIA**

**BADAN PENGEMBANGAN DAN PEMBINAAN BAHASA
KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN
JAKARTA
2016**

**MILIK NEGARA
TIDAK DIPERDAGANGKAN**

Hak Cipta 2016 milik Kementerian Pendidikan dan Kebudayaan
Dilindungi Undang-Undang

Desain Sampul dan Tata Letak
Yudi Nur Riyadi

KATALOG DALAM TERBITAN (KDT)

PB	
413.028.06	Prosiding Seminar Leksilografi: Tantangan
PRO	Leksikografis Bahasa-Bahasa Daerah di
P	Indonesia. Jakarta: Badan Pengembangan dan Pembinaan Bahasa, 2016.
	ISBN 978-979-069-211-4
	LEKSIKOGRAFI-TEMU ILMIAH

Pengarah

Dr. Sugiyono

Penanggung Jawab

Dr. Dora Amalia

Ketua

Azhari Dasman Darnis, M.Hum.

Anggota

Sutiman, Menuk Hardaniwati, Hari Sulastri, Adi Budiwiyanto, Meryna Afrila, Vita Luthfia Urfa, Dina Alfiyanti Fasa, Denda Rinjaya, Dira Hildayani, Winda Luthfiti, Ambiya Ikrami Adji, Dewi Khairiah, Kunkun Purwati

KATA PENGANTAR

Indonesia adalah gudang terbesar kedua bahasa-bahasa lokal. Terverifikasi sejumlah 617 bahasa daerah hidup di Indonesia dengan tingkat vitalitas yang tidak sama (Badan Bahasa, 2014). Belasan bahasa-bahasa dengan penutur lebih dari satu juta memiliki tingkat vitalitas tinggi. Sisanya, ratusan bahasa, berada di tubir jurang kepunahan karena penutur yang menjadi penopang hidupnya tidak memenuhi ‘kuota’ vitalitas. Menurut salah seorang peneliti, pada akhir abad ini hanya akan tersisa sepuluh persen bahasa. Artinya, di Indonesia hanya akan tinggal 60 bahasa saja.

Seminar Leksikografi Indonesia perdana ini mengangkat tema “Tantangan Leksikografis Bahasa-Bahasa Daerah di Indonesia”. Tujuan utama pemilihan tema ini adalah untuk mencari cara yang paling efektif dalam mendokumentasikan leksikon bahasa-bahasa daerah yang berjumlah ratusan dan terancam hilang itu. Melalui tema di atas, beberapa subtema diturunkan. Subtema tersebut adalah:

1. Leksikografi Lapangan dan Penyusunan Korpus Bahasa Daerah.
2. Pemanfaatan Teknologi Informasi dan Komunikasi (TIK) dalam Penyusunan Kamus.
3. Penanganan Dokumentasi Bahasa-Bahasa Daerah.
4. Aspek-Aspek dalam Penanganan Kata-Kata Budaya dalam Kamus.

Prosiding seminar ini merupakan kumpulan makalah-makalah yang dipaparkan selama Seminar Leksikografi Indonesia. Terdapat dua alasan utama dalam menyusun prosiding seminar ini. Pertama, untuk memperkaya perbendaharaan ilmu tentang perkamusan atau leksikografi, khususnya leksikografi bahasa-bahasa lokal di Indonesia. Kedua, yaitu untuk merintis terwujudnya sarana ilmiah yang menjadi wadah para leksikograf, pengkritik kamus, dan pencinta kamus di Indonesia. Selain makalah-makalah yang dipaparkan selama seminar, prosiding ini juga memuat rekomendasi yang dihasilkan dari Séminar Leksikografi Indonesia.

**Kepala Pusat Pengembangan dan Pelindungan
Badan Pengembangan dan Pembinaan Bahasa**

Dr. Sugiyono

DAFTAR ISI

Kata Pengantar	iv
Daftar Isi	v
Kebijakan Pengembangan Kamus Bahasa Daerah di Indonesia <i>Prof. Dr. Dadang Sunendar, M.Hum.</i>	1
Tantangan Leksikografis Bahasa-Bahasa Daerah di Indonesia <i>Dr. Sugiyono</i>	7
Pemerian Makna <i>Dr. Hasan Alwi</i>	13
Kamus sebagai Kitab Undang-Undang <i>Deny Arnos Kwary, Ph.D.</i>	21
Pemanfaatan Korpus dalam Analisis Makna Kata Bersinonim <i>mau, ingin, hendak, dan akan</i> <i>Dewi Puspita, M.A.</i>	31
Menuju Pemanfaatan Leksikografis Korpus Bahasa Daerah: Keanekaragaman Tantangan dan Peluang <i>Yanti, Ph.D. dan Nany Setyono Kurnia, Ph.D.</i>	41
Survei Program Pengolah Korpus untuk Data Bahasa Indonesia dan Bahasa Daerah di Indonesia <i>Prihantoro</i>	59
Komputasi Linguistik untuk Penyusunan Kamus Bahasa Daerah: Studi Kasus Bahasa Sunda <i>Dadan Sutisna</i>	73
Penelitian Korpus Leksikografi: Aplikasi Fonetik Akustik Berkomputer <i>Dayang Hajah Norati binti Bakar</i>	87
Perancangan Aplikasi Kamus <i>Mobile</i> Bahasa Madura sebagai Upaya Pelestarian Bahasa Daerah <i>Salim Anshori, S.S. dan Iqbal Nurul Azhar, M.Hum.</i>	110

Pengembangan WordNet Bahasa-Bahasa Daerah di Indonesia: Kasus WordNet Bahasa Jawa <i>Totok Suhardijanto, Ph.D.</i>	121
Beberapa Kendala Pengayaan Kosakata Bahasa Indonesia dari Bahasa Daerah <i>Asep Rahmat Hidayat, M.Hum.</i>	129
Evaluasi Kamus Bilingual Kamus Bahasa Daerah <i>Using-Indonesia</i> sebagai Upaya Pemertahanan Eksistensi Bahasa Using di Kabupaten Banyuwangi <i>Fajar Erikha, S.Psi. dan Satwiko Budiono, S.Hum.</i>	135
Sejarah Perkembangan Leksikografi Tujuh Dialek di Brunei Darussalam <i>Hajah Rosmariah binti Haji Ali</i>	155
Pembentukan Kamus Pewayangan Bahasa Jawa–Indonesia Berbasis Korpus <i>Ika Nurfarida, S.S. dan Nita Suryawati, S.Hum.</i>	165
Redesain Kamus Bahasa Jawa Berdasarkan Tingkatan Bahasa Jawa (Bahasa Jawa Ngoko, Krama Madya, Krama Inggil) <i>Afwin Sulistiawati, S.Pd.</i>	181
Lema Budaya: Ketersebaran dan Keterwakilan Semantis <i>Ahmad Fadly, M.Hum.</i>	193
Mekanisme Pengolahan dan Penyajian Kosakata Budaya dalam Kamus Pemelajar BIPA <i>Atin Fitriana, M.Hum., Dien Rovita, M.Hum., dan Totok Suhardijanto, Ph.D.</i>	203
Aspek-Aspek dalam Penanganan Kata-Kata Budaya dalam Kamus Besar Bahasa Indonesia: Nama-Nama Makanan Indonesia <i>Defina, S.S., M.Si. dan Renny Soelistiyowati</i>	227
Lampiran	
• Rumusan Seminar Leksikografi Indonesia	236

Kebijakan Pengembangan Kamus Bahasa Daerah di Indonesia

Prof. Dr. Dadang Sunendar, M.Hum.

Badan Pengembangan dan Pembinaan Bahasa

I. LATAR BELAKANG

Indonesia adalah laboratorium bahasa terbesar kedua di dunia setelah Papua Nugini (Spolsky, 2004: 174). Badan Pengembangan dan Pembinaan Bahasa menetapkan jumlah bahasa daerah yang sudah diverifikasi sebesar 617 bahasa dalam hasil penelitian *Bahasa dan Peta Bahasa* yang akan diterbitkan pada tahun 2016 ini. Artinya, saat ini sekurang-kurangnya Indonesia memiliki 617 bahasa daerah yang masih hidup. Dari jumlah tersebut, tidak semua bahasa daerah memiliki tingkat ketahanan atau vitalitas yang sama. Saat ini Badan Bahasa sedang melakukan kajian vitalitas setiap bahasa daerah di Indonesia. Setiap bahasa memiliki ketahanan yang berbeda-beda bergantung pada berbagai faktor, seperti jumlah penutur, letak geografis, dan sikap bahasa penutur terhadap bahasanya. Salah satu aspek yang paling berpengaruh terhadap ketahanan suatu bahasa adalah jumlah penutur Bahasa. Ada penelitian yang memeringkat ketahanan bahasa berdasarkan jumlah penuturnya. Makin besar jumlah penutur bahasa tertentu makin jauh bahasa tersebut dari kepunahan. Sebaliknya, makin sedikit jumlah penuturnya makin bahasa tersebut mendekati jurang kepunahan. Penyusutan jumlah penutur ini biasanya disebabkan antara lain oleh perang, bencana alam yang sangat besar, kawin campur antarsuku, dan sikap bahasa penuturnya.

Letak geografis ikut menentukan ketahanan bahasa daerah. Bahasa yang terletak di tempat yang menjadi lalu lintas perekonomian antaretnik, seperti di daerah pesisir dan perkotaan, akan cenderung berubah atau bahkan hilang, meskipun jumlah penuturnya cukup besar. Kontak bahasa yang terjadi akibat komunikasi antaretnik tersebut sedikit banyak akan memengaruhi bahasa setempat. Bahasa yang lebih dominan akan memengaruhi bahasa yang lain.

Ketahanan sebuah bahasa daerah tidak selalu berkaitan dengan jumlah penutur yang ada. Sikap penutur terhadap lebih banyak memengaruhi ketahanan sebuah bahasa. Sikap bahasa tersebut sangat ditentukan oleh pandangan penutur apakah bahasa yang mereka gunakan tersebut berprestise atau membawa keuntungan, baik secara ekonomis maupun sosial, terhadap kehidupannya. Kecenderungan penutur menggunakan bahasa asing, terutama bahasa Inggris, merupakan cerminan sikap penutur terhadap pilihan bahasanya. Bahasa yang dianggap tidak ekonomis atau memiliki nilai prestise yang rendah, seperti bahasa daerah, rentan untuk ditinggalkan penuturnya.

II. KEBIJAKAN PENANGANAN BAHASA DAERAH

Sejak tahun 1970-an penanganan bahasa-bahasa di Indonesia didasarkan pada –Politik Bahasa Nasional dan Keputusan Kongres Bahasa Indonesia. Sejak tahun 2004, Politik Bahasa Nasional dan keputusan-keputusan kongres itu dilebur menjadi draf RUU Kebahasaan yang akhirnya lahir dalam bentuk Undang-Undang Nomor 24 tahun 2009 pada tanggal 9 Juli 2009. Selanjutnya, sejak tahun 2009 itu, penanganan bahasa-bahasa di Indonesia, baik bahasa negara, bahasa daerah, maupun bahasa asing, didasarkan pada undang-undang itu.

Berdasarkan UU No. 24/2009 tersebut, penangan terhadap bahasa dan sastra daerah diklasifikasikan ke dalam tiga hal, yaitu pengembangan, pembinaan, dan perlindungan bahasa dan sastra daerah. Dalam pengembangan bahasa dilakukan upaya memodernkan bahasa melalui pemerayaan kosakata, pemantapan dan pembakuan sistem bahasa, dan pengembangan laras bahasa. Dalam pembinaan bahasa dilakukan upaya meningkatkan mutu penggunaan bahasa melalui pembelajaran bahasa serta pemasyarakatan bahasa ke berbagai lapisan masyarakat. Selain itu, pembinaan bahasa juga dimaksudkan untuk meningkatkan kedisiplinan, keteladanan, dan sikap positif masyarakat terhadap bahasa itu. Sementara itu, dalam upaya perlindungan dilakukan upaya menjaga dan memelihara kelestarian bahasa melalui penelitian, pengembangan, pembinaan, dan pengajarannya.

Di sisi lain, upaya mengarahkan perkembangan bahasa menuju ke keadaan yang dirancang oleh Pusat Pengembangan dan Pelindungan didasarkan pada kondisi kebahasaan di Indonesia yang sangat kaya dengan bahasa daerah. Kekayaan itu di satu sisi merupakan kebanggaan, di sisi yang lain menjadi tugas yang tidak ringan terutama apabila memikirkan bagaimana cara melindungi, menggali manfaat, dan mempertahankan keragamannya. Menurut UNESCO seperti yang tertuang dalam *Atlas of the World's Language in Danger of Disappearing*, di Indonesia terdapat kurang lebih 154 bahasa yang harus diperhatikan, yaitu sekitar 139 bahasa terancam punah dan 15 bahasa yang benar-benar telah mati. Bahasa-bahasa yang terancam terdapat di Kalimantan (1 bahasa), Maluku (22 bahasa), Papua Barat dan Kepulauan Halmahera (67 bahasa), Sulawesi (36 bahasa), Sumatera (2 bahasa), dan Timor-Flores dan Bima-Sumbawa (11 bahasa). Sementara bahasa yang telah punah berada di Maluku (11 bahasa), Papua Barat dan Kepulauan Halmahera, Sulawesi, dan Sumatera (masing-masing 1 bahasa).

Di antara bahasa-bahasa yang berstatus terancam punah itu pun dari tahun ke tahun mengalami penurunan status, bahkan terdapat 75 bahasa tergolong sekarat (*dying*). Sementara itu, dari ratusan bahasa daerah itu, hanya sedikit bahasa yang memiliki tata aksara dan tradisi sastra, tetapi sebagian besar hanya berbentuk ujaran lisan. (Sumber: Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.). 2014. *Ethnologue: Languages of the World, Seventeenth edition*. Dallas, Texas: SIL International).

Catatan di atas dengan jelas menunjukkan bahwa status bahasa daerah ada yang tetap bertahan, tetapi ada yang terancam kepunahan. Dalam keadaan itu, dapat dipastikan

bahasa Indonesia dapat hidup dan berkembang secara lebih baik. Tuntutan komunikasi di daerah-daerah urban serta komunikasi di bidang politik, sosial, ekonomi, dan ipteks di Indonesia memberi peluang hidup yang lebih baik bagi bahasa Indonesia.

Penelitian menunjukkan bahwa masyarakat Indonesia yang tinggal di perkotaan mempunyai sikap yang cukup positif terhadap bahasa daerah (Sugiyono dan Sasangka, 2011). Artinya, orang Indonesia masih menganggap bahasa daerah perlu dilestarikan dan digunakan dalam komunikasi di ranah domestik. Hanya saja penelitian ini dilakukan dengan sampel masyarakat perkotaan, terutama kota besar di Indonesia, sehingga belum menggambarkan secara utuh sikap bahasa masyarakat Indonesia secara keseluruhan.

III. PENYUSUNAN KAMUS BAHASA DAERAH

Kamus bahasa daerah yang ada dalam koleksi perpustakaan Badan Bahasa diterbitkan sejak tahun 1975, yaitu Kamus Bahasa Indonesia-Bali (A-K). Sejak tahun itu sampai dengan tahun 2014 setidaknya ada 123 kamus bahasa daerah yang diterbitkan oleh Badan Bahasa dan UPT-nya. Pada umumnya, tim penyusun kamus tersebut adalah karyawan Badan Bahasa dan hanya sedikit sekali melibatkan orang di luar Badan Bahasa. Oleh karena itu, selama lebih dari empat puluh tahun tersebut kamus yang berhasil diterbitkan hanya sekitar 17% dari 716 bahasa daerah hasil verifikasi Badan Bahasa.

Lambatnya laju penambahan jumlah kamus ini menjadi hal yang penting untuk ditangani. Untuk mempercepat laju pertambahan tersebut, mulai tahun 2016 ini, penyusunan kamus bahasa daerah diserahkan kepada masyarakat. Untuk tahun anggaran 2016 ini, Pusat Pengembangan dan Pelindungan mengalokasikan anggaran untuk lebih dari sepuluh kamus bahasa daerah yang saat ini sedang dalam proses penyusunan. Setidaknya pada tahun 2016 ini ada dua belas kamus yang disusun dan pengerjaannya diserahkan kepada masyarakat dan Badan Bahasa memberi fasilitasi berupa konsultasi dan pendampingan. Untuk tahun-tahun mendatang, jumlah penyusunan kamus bahasa daerah ini akan lebih ditingkatkan dengan melibatkan lebih banyak lagi peran serta masyarakat.

IV. PERAN BAHASA DAERAH DALAM PENGEMBANGAN KOSAKATA BAHASA INDONESIA

Bahasa daerah sangat berperan penting dalam pengembangan bahasa Indonesia. Peran tersebut terutama dalam hal kontribusi kosakata bahasa daerah dalam memperkaya kosakata bahasa Indonesia dan daya ungkap bahasa Indonesia. Kosakata budaya merupakan kekayaan budaya yang harus ikut mewarnai kosakata bahasa Indonesia. Selain itu, konsep baru dari bahasa asing yang masuk ke dalam bahasa Indonesia dan belum ada padanannya dapat dipadankan dengan kata bahasa daerah yang mungkin memiliki kesamaan konsep dengan bahasa asing.

Kebijakan pengembangan kosakata bahasa Indonesia dengan memanfaatkan sumber bahasa daerah tersebut merupakan kebijakan yang dicanangkan oleh Menteri Pendidikan dan Kebudayaan. Oleh karena itu, dalam menyikapi hal ini, Badan Bahasa telah menetapkan sumber terbesar sumbangan kosakata bahasa Indonesia berasal dari bahasa daerah dengan target minimal 13.000 kata per tahun. Target tersebut akan dapat tercapai jika ada peran aktif dari setiap Balai/Kantor Bahasa yang merupakan UPT Badan Bahasa di setiap provinsi. Keterlibatan masyarakat dalam hal ini telah difasilitasi oleh Badan Bahasa dengan menyediakan kemudahan akses menyumbangkan kosakata daerah melalui aplikasi daring.

Inventarisasi kosakata bahasa daerah untuk disumbangkan ke dalam bahasa Indonesia sekaligus dapat dilakukan bersamaan dengan penyusunan kamus bahasa daerah. Oleh sebab itu, kedua kegiatan ini dapat dilakukan secara paralel karena saling mendukung satu sama lain. Melalui seminar ini pula, peserta yang khususnya berasal dari daerah, dapat menyumbangkan kosakata bahasa daerahnya dan sekaligus dapat menyebarluaskan informasi ini kepada masyarakat yang lebih luas.

V. PEMBENTUKAN ASOSIASI PEKAMUS INDONESIA

Penyusun kamus atau pekamus merupakan profesi yang masih sangat baru di Indonesia. Penyusunan kamus memerlukan keterampilan yang sangat khusus dan oleh sebab itu tidak banyak yang menggeluti pekerjaan ini secara profesional. Namun demikian, perkembangan ilmu pengetahuan dan kebutuhan masyarakat terhadap bahan rujukan akan membuka peluang pekerjaan pekamus menjadi makin dibutuhkan. Akhir-akhir ini makin bervariasi kamus-kamus yang beredar di pasaran, baik itu kamus saku, kamus pelajar, maupun kamus bidang ilmu, dan sebagainya. Media kamus pun makin beragam, tidak hanya dalam bentuk cetak, tetapi juga dalam bentuk elektronik, bahkan dalam jaringan (daring).

Keberagaman produk kamus tentu saja membutuhkan keahlian yang makin beragam pula. Oleh sebab itu, profesi pekamus tidak lagi hanya diisi oleh ahli bahasa, tetapi juga oleh ahli di bidang ilmu tertentu dan ahli teknologi informasi.

Untuk mewadahi orang-orang yang terlibat dalam penyusunan kamus, kami berharap dari seminar ini akan muncul pemikiran untuk membentuk asosiasi penyusun kamus atau pekamus di Indonesia. Wadah ini diperlukan sebagai ajang komunikasi dan bertukar pikiran untuk kemajuan perkamusan di Indonesia dan sebagai wadah untuk menyampaikan informasi terbaru tentang tren leksikografi dunia dan perkembangan leksikografi teoretis.

Akhirnya, semoga dari seminar ini akan muncul pemikiran-pemikiran baru yang berguna untuk kemajuan leksikografi Indonesia, khususnya dalam penyusunan kamus bahasa daerah. Selamat berseminar.

Daftar Pustaka

- Kurniawati, Wati (Ed.). akan terbit. *Bahasa dan Peta Bahasa*. Edisi ketiga. Jakarta: Badan Pengembangan dan Pembinaan Bahasa.
- Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.). 2014. *Ethnologue: Languages of the World, Seventeenth edition*. Dallas, Texas: SIL International.
- Pusat Pengembangan dan Pelindungan. 2016. *Rencana Strategis Pusat Pengembangan dan Pelindungan*. Jakarta: Badan Pengembangan dan Pelindungan.
- Spolsky, Bernard. 2004. *Language Policy*. Cambridge: Cambridge University Press.

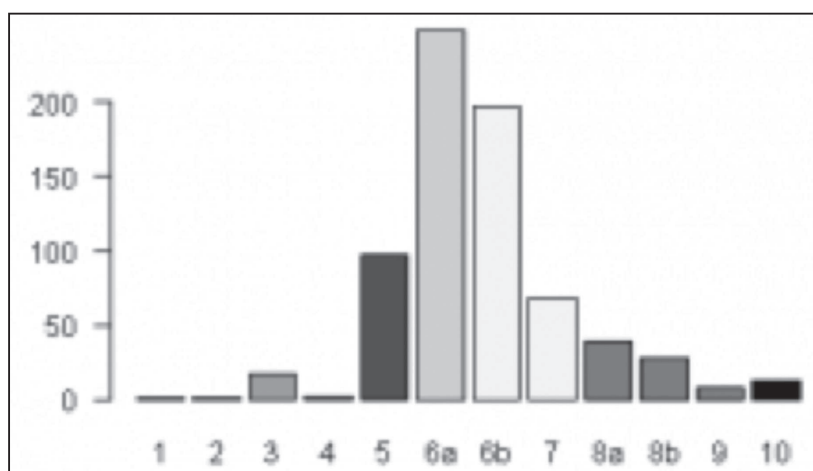
Tantangan Leksikografis Bahasa-Bahasa Daerah di Indonesia

Dr. Sugiyono

Badan Pengembangan dan Pembinaan Bahasa

I. KONDISI KEBAHASAAN DI INDONESIA

Profil kebahasaan di Indonesia ditandai dengan banyaknya bahasa yang terancam punah. Seperti yang dikutip dari Ethnologue.com bahwa ada sekitar 719 bahasa yang tercatat di Indonesia. Sebanyak 706 di antaranya adalah bahasa yang masih hidup, dengan perincian 21 bahasa sudah dibakukan, 97 bahasa sedang berkembang, 248 bahasa kuat, 265 bahasa terancam punah, dan 75 bahasa hampir punah.



Gambar 1. Profil kebahasaan di Indonesia
(dikutip dari: Ethnologue.com)

Bilah grafik 1-5 pada gambar di atas menunjukkan status bahasa yang baku dan digunakan secara luas, bahkan digunakan dan diajarkan melalui institusi. Adapun bilah 6a menunjukkan kondisi bahasa yang masih digunakan secara luas, tetapi belum dibakukan. Bilah grafik 6a-7 menunjukkan kondisi yang ditandai oleh mulai tersendatnya transmisi bahasa antargenerasi. Pada tahap inilah perlu dilakukan upaya revitalisasi. Status bahasa yang mulai menuju ke kepunahan, ditunjukkan oleh bilah grafik 8a-9 yang ditandai oleh bahasa yang hanya dikuasai oleh penutur tua saja, sedangkan bilah 9-10 menunjukkan bahasa yang sudah benar-benar punah.

Bahasa-bahasa daerah yang jumlahnya ratusan di Indonesia tersebut sangat kental dengan tradisi lisan. Artinya, bahasa-bahasa daerah tersebut biasanya ditandai dengan tidak adanya tradisi tulis dan hal itu berarti bahwa tidak ada korpus bahasa daerah yang tersedia. Hampir semua data berasal dari tradisi lisan, baik itu berupa percakapan lisan, lagu, dongeng, nasihat yang harus dijangkau melalui metode pengumpulan data lapangan. Oleh karena itu, penyusunan kamus untuk bahasa-bahasa daerah tersebut masuk dalam cabang leksikografi lapangan (*field lexicography*) atau leksikografi bahasa hampir punah (*endangered languages lexicography*). Produk dari leksikografi lapangan adalah kamus lapangan (*field dictionary*), yaitu “*dictionaries exist only for ‘minor’ languages, spoken in small communities without any tradition of written literature*” (Béjoint, 1983: 70).

Kondisi lain yang menandai profil kebahasaan di Indonesia adalah banyaknya bahasa yang statusnya terancam punah, terutama bahasa-bahasa daerah yang ada di Indonesia bagian timur. Kondisi seperti itu membutuhkan penanganan yang cepat sebelum bahasa-bahasa tersebut punah. Penyusunan kamus sebagai salah satu usaha dokumentasi bahasa harus berpacu dengan waktu dan harus berhadapan dengan ratusan bahasa daerah yang membutuhkan penanganan segera. Oleh karena itu, perlu ditemukan metode yang tepat dan sekaligus cepat untuk mengumpulkan data, menyusun korpus, dan menyusun kamus. Hal-hal tersebut akan dibicarakan dalam makalah ini.

II. SEJARAH LEKSIKOGRAFI INDONESIA

Leksikografi di Indonesia ditandai dengan penyusunan kamus dwibahasa. Diawali dengan hanya leksikografi tertua di Indonesia yang berupa daftar kata Cina-Melayu yang disusun pada awal abad XV yang memuat 500 lema. Setelah itu Pigafetta, seorang berkebangsaan Italia, membuat daftar kata Italia-Melayu pada tahun 1522. Selanjutnya pada tahun 1603 Frederick de Houtman menyusun kamus *Spraeck ende woordboek, Inde Malaysche ende Madagaskarsche Talen met vele Arabische ende Turksche Woorden*. Beberapa tahun kemudian, yaitu tahun 1623, Casper Wiltens dan Sebastian Danckaerts menyusun kamus Belanda-Melayu dan Melayu-Belanda yang berjudul *Vocabularium ofte woordboek naer order vanden Aphabet in't Duytsch-Maleysche ende Malaysche-Duytsch. Lexicon Javanum* (anonim) terbit tahun 1706 dan naskahnya sekarang tersimpan di perpustakaan Vatikan. Lalu setelah itu ada tiga kamus yang terbit tahun 1800-an, yaitu *Dictionary of the Malayan Language* (Marsden, 1812); *Nederduitsch-Maleisch en Soendasch Woordenboek* (A. De Wilde, 1841), dan *Bataksch-Nederduitsch Woordenboek* (van der Tuuk, 1861).

Minat pada bahasa dan perkamusan pada zaman itu (kolonial) terbatas pada orang asing saja. Kamus-kamus tersebut disusun pada umumnya untuk keperluan penyebaran agama dan kepentingan dagang untuk berkomunikasi dengan penduduk setempat. Walaupun ada beberapa dari penyusun kamus tersebut adalah ahli bahasa, seperti van der Tuuk dan A. De Wilde, sebagian besar mereka mempunyai latar belakang yang beragam.

Barulah pada awal abad ke-19, kamus mulai disusun oleh orang pribumi. Raja Ali Haji dari Riau menyusun *Kitab Pengetahuan Bahasa* yang memuat daftar kata bahasa Melayu. Kamus ini diterbitkan oleh Al Ahmadiyah Press, Singapura tahun 1928. Setelah itu pada tahun 1930 disusunlah *Baoesastra Djawa* oleh W.J.S. Poerwadarminta, C.S. Hardjasoedarma, and J.C. Poedjasoedira. Adapun kamus bahasa Sunda modern disusun oleh R. Satjadibrata pada tahun 1948 dengan judul *Kamoes Bahasa Soenda*.

Dari sejarah perkamusan di Indonesia tersebut, dapat disimpulkan bahwa leksikografi di Indonesia pada awalnya diwarnai dengan penyusunan kamus bahasa daerah. Beberapa bahasa daerah, seperti bahasa Jawa, telah mempunyai aksara dan naskah-naskah yang dapat dijadikan sumber data, tetapi tidak demikian halnya dengan bahasa-bahasa daerah yang lain yang tidak memiliki tradisi tulis. Penyusunan kamus tersebut dilakukan dengan mencatat semua bentuk kebahasaan secara langsung dari penutur bahasa daerah. Pada proses ini, penyusun kamus telah melakukan metode leksikografi lapangan.

III. STRATEGI LAMA PENYUSUNAN KAMUS

Penyusunan kamus bahasa daerah yang dilakukan oleh Badan Pengembangan dan Pembinaan Bahasa, dahulu Pusat Bahasa, telah dimulai pada tahun 1970-an. Terbitan kamus bahasa daerah yang tercatat paling awal, yaitu kamus yang terbit pada tahun 1975. Selanjutnya, kamus-kamus bahasa daerah terus disusun sampai dengan sekarang sehingga berjumlah 123 kamus, baik bahasa daerah-bahasa Indonesia, maupun bahasa Indonesia-bahasa daerah.

Jika dibandingkan dengan jumlah bahasa daerah hasil inventarisasi Badan Bahasa sebesar 617 bahasa (data tahun 2014), penyusunan kamus bahasa daerah baru mencapai angka 17% saja dari semua bahasa daerah di Indonesia. Hal tersebut berarti ada sekitar 83% atau sekitar 500-an bahasa daerah yang belum disusun kamusnya. Oleh karena itu, pekerjaan penyusunan kamus bahasa daerah tersebut tidak mungkin dilakukan oleh Badan Bahasa semata.

Penyusunan kamus bahasa daerah selama ini dilakukan dengan strategi sebagai berikut.

1. Penyusunan dilakukan oleh pekamus dari Pusat Bahasa.

Penyusunan kamus bahasa daerah selama ini dilakukan oleh pekamus dari Badan Bahasa dulu Pusat Bahasa dan UPT Balai/Kantor yang tersebar di daerah. Ada beberapa kamus bahasa daerah yang disusun oleh masyarakat, tetapi jumlahnya sangat sedikit. Itu juga sebagian besar dilakukan oleh penyusun yang pernah mengikuti pelatihan di Pusat Bahasa. Kamus Bahasa Betawi dan Kamus Bahasa Gorontalo, sebagai contoh. Terdapat juga beberapa kamus bahasa daerah yang disusun oleh masyarakat dan beberapa peneliti asing untuk keperluan penelitian atau karena kecintaan terhadap bahasa tertentu.

2. Masyarakat pemilik bahasa diposisikan hanya sebagai narasumber dalam penyusunan kamus.

Pekamus dari Pusat Bahasa turun ke daerah atau menghadirkan penutur bahasa daerah untuk menjadi narasumber dalam penyusunan kamus. Melalui narasumber digali kekayaan bahasa daerah tertentu. Pekamus dibekali dengan ilmu leksikografi, fraseologi, semantik leksikal, fonologi, dan sebagainya dalam menggali kekayaan bahasa daerah dan menyajikannya dalam bentuk kamus. Hasil penyusunan berupa kamus belum tentu akan sampai ke hadapan narasumber. Biasanya hanya menjadi bahan penelitian di kampus-kampus atau Balai/Kantor Bahasa.

3. Hasil dari penyusunan kamus dicetak

Belum terdapat teknologi penyebarluasan hasil penyusunan kamus selain dari pencetakan dan pengiriman. Kita mendapati kamus-kamus berbahasa daerah terbitan Pusat Bahasa di berbagai perpustakaan kampus dalam kondisi masih bagus dan tua karena jarang dibuka. Kamus tersebut hanya sesekali dibuka untuk penelitian mahasiswa atau tugas kuliah. Jarang sekali kamus bahasa daerah dibuka selain untuk keperluan penelitian.

4. Belum menggunakan aplikasi teknologi informasi

Cara penyusunan kamus mulai dari pemilihan sampai penyuntingan masih dilakukan dengan cara manual. Penulisan entri di slip terkadang masih dilakukan. Terdapat tempat khusus untuk penyimpanan data yang dapat berupa ruangan tersendiri. Penyimpanan memerlukan ruangan dan cara pengolahan dokumen tersendiri. Cara ini tidak lagi efektif dan efisien saat ini. Belum terdapat pangkalan data yang sudah diolah secara modern yang memudahkan konversi data ke dalam berbagai aplikasi nantinya.

Strategi penyusunan kamus seperti tersebut di atas hanya mampu menghasilkan sedikit kamus bahasa daerah. Oleh karena itu, perlu dicari strategi baru dapat mengatasi masalah kepunahan bahasa dengan mempercepat langkah dokumentasi bahasa melalui penyusunan kamus. Diperlukan cara pendokumentasian bahasa melalui penyusunan kamus yang cepat, praktis, dan mudah.

IV. USULAN STRATEGI PENYUSUNAN KAMUS BAHASA DAERAH

Untuk menjawab kebutuhan penyusunan kamus yang cepat diperlukan strategi baru yang sesuai dengan perkembangan teknologi. Strategi baru tersebut sebagai berikut.

1. Pelibatan masyarakat

Sejalan dengan kebijakan Kementerian Pendidikan dan Kebudayaan yang menjadikan masyarakat sebagai ekosistem pendidikan, penyusunan kamus juga harus melibatkan masyarakat pemilik bahasa. Pelibatan masyarakat tidak hanya dimanfaatkan sebagai

narasumber, tetapi sebagai penyusun sekaligus. Dalam penyusunan tersebut masyarakat dibantu dengan aplikasi dan penyunting dari Badan Bahasa. Aplikasi diperlukan untuk memudahkan penyusunan, sedangkan penyunting bertugas memperbaiki dan menyempurnakan definisi, contoh penggunaan, kelas kata dan sebagainya. Dengan cara ini, kalau setiap penutur membuat kamus bahasanya sendiri, kamus 600-an bahasa daerah akan cepat terwujud sekaligus dengan dialek-dialeknya yang tak kalah banyak juga. Namun cara ini bukannya tidak memiliki kendala, seperti kurangnya penyunting kamus.

Usaha ke arah ini sudah dilakukan. Tahun ini Badan Bahasa sudah melakukan dua kali bengkel penyusunan kamus untuk Indonesia bagian Bbarat dan Timur. Untuk bagian barat diadakan di Medan, sedangkan untuk bagian timur diadakan di Makassar. Pada bengkel itu diajarkan cara praktis penyusunan kamus. Dikenalkan pula aplikasi penyusunan kamus, aplikasi pemilihan contoh, dsb. Pengenalan aplikasi disertai dengan praktik langsung. Peserta bengkel terdiri atas pekamus dari Balai/Kantor, dosen/mahasiswa perguruan tinggi, dan masyarakat umum.

2. Penggunaan sarana teknologi informasi yang tepat.

Teknologi informasi berkembang dengan cepat, bahkan sangat cepat. Teknologi baru bermunculan mengungguli teknologi baru lain. Terdapat banyak sekali aplikasi penyusunan kamus, baik dari penyediaan basis data, pembangunan korpus, bahkan pembuatan antarmuka yang bagus. Baik yang berbayar, maupun gratis. Dalam pemilihan aplikasi, Badan Bahasa memilih membeli aplikasi, alih-alih menggunakan aplikasi gratis karena berkaitan dengan keamanan data kamus yang sangat penting.

3. Pemanfaatan jaringan internet

Dunia sudah menjadi kampung kecil. internet menghubungkan setiap orang. Memudahkan lalu lintas data. Penyebaran kamus yang sebelumnya dilakukan melalui pencetakan dan pengiriman sekarang dapat dilakukan dengan lebih cepat, supercepat bahkan, jauh lebih murah, dan mudah, selama terdapat jaringan internet. Setiap orang dapat memiliki akses terhadap kamus, bahkan dapat membuat aplikasi dari data yang ada. Sebagai contoh, KBBI edisi III memiliki puluhan aplikasi yang dibuat oleh masyarakat sendiri. Setiap aplikasi memiliki kekurangan dan kelebihan masing-masing. Masyarakat memiliki kebebasan dalam menentukan aplikasi KBBI yang sesuai dengan kebutuhan mereka. Saat ini, lebih banyak masyarakat membuka KBBI dibanding dulu.

4. Pemanfaatan Korpus

Korpus, sederhananya, merupakan kumpulan beberapa teks bahasa yang dapat digunakan untuk penelitian linguistik. Saat ini, korpus tidak hanya dimanfaatkan oleh para peneliti. Pekamus dan penyusun tata bahasa juga sudah menyandarkan hasil karyanya pada korpus agar teori dan kamus yang dihasilkan tidak bersifat priskriptif belaka, tetapi juga memenuhi unsur deskriptifnya atau perpadua dari

keduanya yang biasa disebut dengan proskriptif. Melalui penelitian korpus, informasi yang tersaji dalam kamus tidak hanya bersifat ‘bagaimana seharusnya’ saja, tetapi juga ‘bagaimana kenyataannya’. Dengan cara ini keberterimaan sebuah kata berikut definisi dan kelengkapan lain di tengah pengguna kamus tinggi.

Korpus merupakan sumber data otentik pemakaian bahasa oleh pengguna. Melalui data yang ada di korpus, konteks, definisi, kelas kata, ranah penggunaan, ragam, register, contoh, dan sebagainya dapat diketahui dengan pasti. Pekamus tinggal memadukan antara kenyataan penggunaan di lapangan dengan kaidah yang disepakati.

Penggunaan korpus bersifat melengkapi kemampuan intuitif pekamus. Melalui korpus, pekamus dapat menemukan realisasi dari makna suatu kata. Pekamus tidak perlu lagi meraba makna, kelas kata, ragam dan sebagainya kemudian mengarang contohnya. Semua sudah tersedia dalam korpus. Pekamus tinggal menyempurnakannya dan memilih yang paling sesuai. Korpus dapat mengurangi subjektivitas pekamus. Semua berasal dari sumber data yang ada yang berasal dari masyarakat.

Korpus bahasa daerah masih asing di telinga. Belum terdapat korpus tersendiri yang berisi kekayaan bahasa daerah tertentu, tetapi sudah ada beberapa laman yang berisi bahasa daerah tertentu. Surat kabar berbahasa daerah, wikipedia berbahasa daerah, komunitas lokal, blog pribadi, dan sebagainya dapat dijadikan korpus sebagai dasar pembuatan definisi, kelas kata, contoh dsb kamus bahasa daerah.

Daftar Pustaka

- Béjoint, Henry. 1983. “Field Dictionaries” dalam *Lexicography: Principles and Practice*. London: Academic Press.
- Hartmann, R.R.K. 2001. *Teaching and Researching Lexicography*. Edinburg: Pearson Education Limited.
- Moe, Ronald. 2007. ”Dictionary Development Program” dalam *SIL Forum for Language Fieldwork 2007-009*. Ron Moe and SIL International.
- Svensen, Bo. 2009. “Lexicography” dalam *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.

Sumber Daring:

Ethnologue: Languages of The World. Diakses dari <http://www.ethnologue.com/profile/> ID, tanggal 20 Juli 2016.

Pemerian Makna¹

Hasan Alwi

hasanalwi40@yahoo.com

Dalam kegiatan penyusunan kamus, menyusun deskripsi makna lema/sublema merupakan tahapan yang utama dan sangat menentukan bagi keberhasilan kamus yang diterbitkan. Oleh karena itu, sering dikatakan bahwa mutu kamus ditentukan oleh ketepatan perumusan definisinya, di samping ditentukan juga oleh kepraktisan contoh pemakaian dari lema/sublema yang dideskripsikan maknanya tersebut. Karena lema/sublema yang didefinisikan itu pada dasarnya harus ditinjau dari atau dihubungkan dengan kelas katanya, maka perumusan makna itu harus memperhatikan atau bahkan didasarkan pada kelas kata lema/sublema yang bersangkutan. Dengan perkataan lain, cara merumuskan kata yang tergolong nomina, misalnya, haruslah berbeda dari cara merumuskan kata yang tergolong verba atau adjektiva.

Demikian pula halnya dengan kata yang tergolong partikel harus dirumuskan dengan cara yang tidak sama dengan perumusan kata yang termasuk adverbial atau kelas kata lainnya. Sehubungan dengan hal itu, berikut ini dicontohkan perumusan kata *rumah*, *sekolah*, *kantor*, dan *masjid* yang kesemuanya berkelas kata nomina dan “diperkirakan” memiliki ciri-ciri semantis yang bermiripan. Kutipan pemerian makna di bawah ini sesuai dengan yang tercantum di dalam *Kamus Besar Bahasa Indonesia* Edisi IV (KBBI IV, 2008).

ru.mah *n* **1** bangunan untuk tempat tinggal; **2** bangunan pd umumnya (spt gedung); ... (KBBI IV, 2008: 1188)

se.ko.lah *n* **1** bangunan atau lembaga untuk belajar dan mengajar serta tempat menerima dan memberi pelajaran (KBBI IV, 2008: 1244)

kan.tor *n* **1** balai (gedung, rumah, ruang) tempat mengurus suatu pekerjaan (perusahaan dsb); **2** tempat bekerja; (KBBI IV, 2008: 619)

mas.jid *n* rumah atau bangunan tempat bersembahyang orang Islam (KBBI IV, 2008: 883)

Pada definisi keempat nomina di atas terdapat tiga kata kunci yang juga diperkirakan memiliki kemiripan dalam ciri-ciri semantisnya, yaitu kata *bangunan*, *balai* dan *gedung*. Berikut adalah kutipan deskripsi makna dari ketiga kata tersebut.

¹ Makalah Seminar Leksikografi Indonesia dengan tema “Tantangan Leksikografis Bahasa-bahasa Daerah di Indonesia”, Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan, Hotel Santika, Taman Mini Indonesia Indah, 26–29 Juli 2016

ba.ngun.an *n* sesuatu yg didirikan; sesuatu yg dibangun (spt rumah, gedung, menara); (KBBI IV, 2008: 134)

ba.lai *n* **1** gedung; rumah (umum); kantor; **2** *kl* rumah (dl lingkungan istana); ...

ge.dung *n* **1** bangunan tembok dsb yg berukuran besar sbg tempat kegiatan, spt perkantoran, pertemuan, perniagaan, pertunjukan, olahraga, dsb; **2** rumah tembok yg berukuran besar; (KBBI IV, 2008: 425)

Sehubungan dengan *masjid* yang berfungsi sebagai ‘tempat bersembahyang orang Islam’, deskripsi makna *masjid* ini perlu pula dibandingkan dengan nama bangunan bagi para pemeluk agama yang lain.

gereja/ge-re-ja/ /geréja/ *n* **1** gedung (rumah) tempat berdoa dan melakukan upacara agama Kristen: *di situ ada -- yang besar*; **2** badan (organisasi) umat Kristen yang sama kepercayaan, ajaran, dan tata cara ibadahnya (KBBI IV, 2008: 445)

pu-ra *n* tempat beribadat (bersembahyang) umat Hindu Dharma; (KBBI IV, 2008: 1119)

ke.len.teng /kelenténg/ *n* bangunan tempat memuja (berdoa, bersembahyang) dan melakukan upacara keagamaan bagi penganut Konghucu (KBBI IV, (2008: 655)

Dalam takrif masjid terdapat frasa *tempat bersembahyang orang Islam*. Dalam takrif *pura* digunakan frasa *umat Hindu Dharma*, sementara frasa penganut Konghucu muncul dalam takrif *kelenteng*. Makna *umat* pada *umat Hindu Dharma* atau *penganut* pada *penganut Konghucu* pasti berhubungan dengan agama atau kepercayaan.

umat *n* **1** para penganut (pemeluk, pengikut) suatu agama; penganut nabi; **2** makhluk manusia; (KBBI IV, 2008: 1524)

pe-me-luk *n* **1** orang yg memeluk; **2** *ki* penganut ajaran: (KBBI IV, 2008: 1042)

peng-a-nut *n* pengikut (aliran politik); pemeluk (agama, kepercayaan): (KBBI IV, 2008: 79)

peng-i-kut *n* **1** penganut; **2** peserta (KBBI IV, 2008 (KBBI IV, 2008: 523)

Berbeda dengan *umat*, *pemeluk*, *penganut*, dan *pengikut*, makna orang tidak harus selalu memiliki ciri semantis ‘agama’. Adapun orang pada frasa *orang Islam*, jika dilihat dari deskripsi makna *orang* yang dikodifikasikan dalam kamus, pasti berhubungan dengan ciri semantis ‘agama’, sebagaimana yang dimaksudkan dengan *manusia dalam arti khusus*. Dengan ciri makna ini, para pemeluk agama Kristen, Hindu, dan Konghucu pun masing-masing dapat pula dinyatakan lewat frasa *orang/umat Kristen*, *orang/umat Hindu*, dan *orang/umat Konghucu*.

orang *n* **1** manusia (dl arti khusus); **2** manusia (ganti diri ketiga yg tidak tentu): *jangan lekas percaya pd mulut --*; **3** dirinya sendiri; manusianya sendiri: *saya tidak bertemu dng -- nya*; **4** kata penggolong untuk manusia: *lima -- nelayan*; **5** anak buah (bawahan): *mereka itu -- nya Pak Camat*; **6** rakyat (dr suatu negara); warga negara: -- *Pakistan*; **7** manusia yg berasal dr atau tinggal di suatu daerah (desa, kota, negara, dsb): *dia -- Bogor*; *suaminya -- Eropa*; **8** suku bangsa; **9** manusia lain; bukan diri sendiri; bukan kaum (golongan, kerabat) sendiri: *jangan anak*

sendiri, anak -- pun saya tolong; negeri -- , negeri lain (bukan negeri kita); 10 cak karena (sebenarnya): mana dapat membayar, -- belum gajian; (KBBI IV, 2008: 986)

Yang dilakukan oleh umat Islam/Kristen/Hindu/Konghucu di tempat peribadatan masing-masing adalah *bersembahyang, beribadat, berdoa*, atau *memuja*. Sepintas lalu, keempat perbuatan/kegiatan ritual memiliki makna yang kurang lebih sama. Untuk itu, simaklah apa yang dikodifikasikan kamus mengenai keempat kegiatan itu.

ber·sem·bah·yang *v* **1** melakukan sembahyang (salat): *mereka- hari raya di lapangan dekat masjid Al-Azhar; 2* berdoa (memohon) kpd Tuhan: - *di gereja; 3* upacara selamatan untuk menghormati (memuliakan) para leluhur dsb: - *Tahun Baru; (KBBI IV, 2008: 1259)*

iba·dat *n* **1** ibadah; **2** segala usaha lahir dan batin sesuai dng perintah Tuhan untuk mendapatkan kebahagiaan dan keseimbangan hidup, baik untuk diri sendiri, keluarga, masyarakat maupun thd alam semesta; **3** upacara keagamaan;

ber·i·ba·dat *v* menunaikan ibadah; (KBBI IV, 2008: 515)

ber·doa *v* mengucapkan (memanjatkan) d(KBBI IV, 2008) kpd Tuhan (KBBI IV, 2008: 337)

me·mu·ja *v* **1** menghormati dewa-dewa dsb dng membakar dupa, membaca mantra, dsb; **2** memuja-muja; **3** menjadikan sesuatu dng mantra: *ia ~ anak yg tidak tahu budi itu hingga menjadi batu; (KBBI IV, 2008: 1112)*

Kata *bersembahyang, beribadat, berdoa*, dan *memuja* tergolong verba. Untuk makna untuk lema/sublema yang berkelas kata verba berikut ini dicontohkan verba yang erat kaitannya dengan perbuatan yang dilakukan dengan menggunakan tangan dan kaki, yaitu *melempar, memukul, menyepak*, dan *menendang*.

me·lem·par *v* **1** membuang jauh-jauh; **2** melontari (dng) (KBBI IV, 2008: 811).

me·mu·kul *v* **1** mengenakan suatu benda yg keras atau berat dng kekuatan (untuk mengetuk, memalu, meninju, menohok, menempa, dsb) ...

me·nye·pak *v* memukul dng kaki; menendang; mendepak; (KBBI IV, 2008: 1278)

me·nen·dang *v* **1** menyepak; mendepak (dng kaki): *ia berusaha ~ bola ke arah gawang lawan; 2* mengusir; memecat; mengeluarkan (dr kantor, perusahaan, organisasi, dsb): *berdasarkan undang-undang perburuhan, setiap pemimpin tidak dapat ~ bawahannya begitu saja tanpa alasan yg kuat; 3* mendesak (mendorong) kuat-kuat: *bocoran pd pipa gas disebabkan oleh tekanan gas yg kuat yg ~ klep pipa hingga meletup; (KBBI IV, 2008: 1438)*

Dari definisi itu terlihat bahwa pada *melempar* dan *memukul* peran tangan sama sekali tidak dimunculkan secara eksplisit. Hal ini berbeda dengan definisi *menyepak* dan *menendang* yang secara jelas menyebutkan bahwa perbuatan yang dimaksudkan dilakukan dengan menggunakan kaki. Pada verba *membakar* dan *memotong* berikut ini yang berperan tidak hanya tangan, tetapi ada semacam alat yang digunakan untuk melaksanakan perbuatan yang bersangkutan, yaitu *api* untuk *membakar* dan *barang tajam* untuk *memotong*.

mem.ba.kar *v* menghanguskan (menyalakan, meru-sak-kan) dng api (*KBBI IV*, 2008:121)

me.mo.tong *v* **1** memutuskan dng barang tajam; me-ngerat; memenggal. (*KBBI IV*, 2008: 1096)

Pada contoh berikut ditampilkan deskripsi makna tentang lema/sublema yang tergolong adjektiva. Kata *sedih* dan *senang* pasti menggambarkan keadaan atau suasana hati seseorang, sedangkan kata *gelap* dan *terang* ada atau erat kaitannya dengan ada atau tidak adanya cahaya di suatu tempat tertentu.

se.dih *a* **1** merasa sangat pilu dl hati; susah hati; ...; **2** menimbulkan rasa susah (pilu dsb) dl hati; duka (*KBBI IV*, 2008: 1238)

se.nang *a* **1** puas dan lega, tanpa rasa susah dan kecewa, dsb ...; **2** betah ...; **3** berbahagia (tidak ada sesuatu yg menyusahkan, tidak kurang suatu apa dl hidupnya) ...; **4** suka; gembira (*KBBI IV*, 2008: 1267)

ge.lap *a* **1** tidak ada cahaya; kelam; tidak terang ...; **2** malam ...; **3** tidak atau belum jelas (tt perihal, perkara, dsb); samar ...; **4** rahasia (tidak secara terang-terangan); tidak halal atau tidak sah; tidak menurut aturan (undang-undang hukum) yg berlaku (*KBBI IV*, 2008: 428)

te.rang *a* **1** dl keadaan dapat dilihat (didengar); nyata; jelas ...; **2** cerah; bersinar ...; **3** siang hari ...; **4** bersih (tt halaman atau kebun) krn dibersihkan (disiangi, ditebangi, dsb) ...; **5** jernih; bersih (tt udara, langit) ...; **6** sah (tt bukti, barang-barang, dsb) ...; **7** terbukti (kebenarannya, kesalahannya, dsb) ...; **8** cahaya; sinar (*KBBI IV*, 2008: 1448)

Sehubungan dengan deskripsi makna kata yang berkategori adjektiva ini, secara khusus perlu dikemukakan catatan mengenai *tinggi* dan *panjang* yang masing-masing lawan katanya ialah *rendah* dan *pendek*. Akan tetapi, dalam kenyataan berbahasa di Indonesia *tinggi* tidak dikontraskan dengan *rendah*, tetapi dengan *pendek*. Bahwa dalam BI *pendek* merupakan lawan kata *tinggi*, hal itu sudah merupakan kelaziman berbahasa yang tidak perlu dipertanyakan lagi. Namun, jika kita mencermati takrif kamus, baik kamus bahasa Indonesia maupun kamus bahasa Melayu, dan mengingat-ingat kembali ungkapan lama, termasuk peribahasa, kita akan menyadari bahwa *tinggi* berkontras dengan *rendah*, bukan dengan *pendek* (yang sudah lazim sebagai lawan kata *panjang*). Perhatikan contoh pemakaiannya sbb.

- Suaranya terlalu **tinggi**, **rendahkan** satu oktaf lagi.
- Pangkatnya sudah **tinggi**, pangkat saya sih masih **rendah**.
- Bagaimanapun kita mesti duduk sama **rendah** dan berdiri sama **tinggi**.

Contoh berikut jelas memperlihatkan bahwa dalam bahasa Indonesia digunakan pasangan *tinggi* dan *pendek*, sementara dalam bahasa Melayu kelazimannya adalah pasangan *tinggi* dan *rendah* (khususnya dalam mendeskripsikan ukuran tubuh manusia).

Bila pria memiliki **tubuh pendek**, jangan merasa minder. Pasalnya, mereka yang **bertubuh pendek** ternyata lebih panjang umur.

(<http://lifestyle.okezone.com/read/2014/05/19/482/986931/pria-pendek-tak-perlu-minder-anda-lebih-panjang-umur>)

Senario sekarang, apabila profesion seperti juruterbang, pramugari, polis, dan askar turut menjadikan ketinggian antara syarat utama, menyebabkan mereka yang **bertubuh rendah** mempunyai peluang yang agak terhad dalam memilih pekerjaan.

(<http://www.utusan.com.my/gaya-hidup/keluarga/tubuh-berhenti-meninggi-1.24161>)

Takrif *pendek* dan *rendah* menurut *Kamus Besar Bahasa Indonesia Edisi IV (KBBI IV)* dan *Kamus Dewan Edisi IV (KD IV)* adalah sebagai berikut.

pen-dek /péndék/ **a** **1** *dekat* jaraknya dr ujung ke ujung: *penggaris --*; **2** *dekat* jaraknya dr sebelah bawah; tidak tinggi: *buah jambu dapat dipetik anak-anak krn pohonnya masih --*; **3** *sebentar* (tt waktu): *dl waktu yg -- kemelut itu dapat diatasi*; **4** *ringkas*; *singkat* (tt cerita dsb): *cerita --*; (KBBI IV, 2008: 1044)

pendek (péndék) **1.** tidak panjang, tidak tinggi, pandak, rendah, ringkas; **2.** tidak lama, singkat; (KD IV, 2007:1171)

¹ren-dah **a** **1** *dekat ke bawah*; tidak tinggi; **2** *hina*; *nista*: -- *budi*; (KBBI IV, 2008: 1163)

rendah **I** **1.** tidak jauh jaraknya dr bawah, tidak tinggi (lwn tinggi); sedang (saja) rendahnya (badan orang)² (KD IV, 2007: 1313)

Telah dikemukakan bahwa ketepatan mendeskripsikan makna lema/sublema merupakan syarat utama dalam penyusunan kamus. Tingkat ketepatan pendeskripsian makna itu akan sangat ditentukan oleh seberapa jauh definisi yang dirumuskan itu menggambarkan komponen makna yang secara inheren merupakan konsep inti yang terkandung di balik lema/sublema yang bersangkutan. Untuk jenis permainan dalam bidang olahraga, misalnya, deskripsi makna yang disusun haruslah menyinggung sekurang-kurangnya tiga hal, yaitu pemain, alat yang digunakan, dan tempat permainan atau cara bermain. Untuk itu, contoh tentang *bulu tangkis*, *sepak bola* dan *tenis* di bawah ini perlu disimak untuk mendefinisikan ketiga hal tersebut.

bu.lu tang.kis *n* cabang olahraga yg berupa permainan yang dimainkan dng memakai raket dan kok yg dipukul melampaui jaring yg direntangkan di tengah lapangan; badminton; (KBBI IV, 2008: 221)

² Selain merupakan antonim dari *tinggi*, makna *rendah* yang lainnya adalah (1) tidak nyaring, lembut, perlahan (suara dll); (2) ~ *hati* suka merendahkan diri, tidak angkuh (sombong, bongkak), baik hatinya; (3) tidak mulia (budi hati), hina, aib, nista; (4) tidak mahal (harga), murah; (5) tidak banyak, kecil, sedikit (pendapatan dll); (6) (taraf) permulaan; (7) tidak tinggi atau tidak penting (kedudukan, pangkat, dll), tidak istimewa, kecil; (8) kurang baik (mutu dll); (9) tidak maju lagi; (KD IV, 2007: 1313)

se.pak bo.la *n* Olr permainan beregu di lapangan, menggunakan bola sepak dr dua kelompok yg berlawanan yg masing-masing terdiri atas sebelas pemain, berlangsung selama 2 X 45 menit, kemenangan ditentukan oleh selisih gol yg masuk ke gawang lawan; (KBBI IV, 2008: 1278)

te.nis *n* permainan olahraga yg menggunakan bola (sebesar kepalan) sbg benda yg dipukul dan raket sbg pemukulnya, dimainkan oleh dua pemain (dua pasang), di lapangan yg dibatasi oleh jaring setinggi kira-kira satu meter (KBBI IV, 2008: 1442)

Ketiga cabang olahraga di atas ditandai secara berbeda. Pada *bulu tangkis* dan *tenis* masing-masing ditandai dengan *cabang olahraga* dan *permainan olahraga*, sementara penanda tersebut pada *sepak bola* dinyatakan dengan singkatan *Olr*. Seharusnya, demi ketaatasaan, singkatan *Olr* itu digunakan juga pada *bulu tangkis* dan *tenis*.

Di samping ketaatasaan seperti yang dimaksudkan di atas, pencantuman beberapa cabang olahraga berikut ini perlu dicermati lebih lanjut.

ping-pong *n* tenis meja (KBBI IV, 2008: 1077)

sepak raga permainan bola dr rotan yg dianyam yg menggunakan tendangan kaki dan kepala; (KBBI IV, 2008: 1278)

sepak takraw sepak raga; (KBBI IV, 2008: 1278)

te-nis /ténis/ *n* permainan olahraga yg menggunakan bola (sebesar kepalan) sbg benda yg dipukul dan raket sbg pemukulnya, dimainkan oleh dua pemain (dua pasang), di lapangan yg dibatasi oleh jaring setinggi kira-kira satu meter: *setiap minggu pagi ayah bermain --*;

-- **meja** permainan olahraga yg menggunakan bola pingpong dan bet (berlapis karet) sbg pemukulnya dan meja yg dirancang khusus sbg lapangannya (KBBI IV, 2008: 1442)

vo.li *n* **1**. bola voli; **2** bola yang terbang atau melayang sebelum menyentuh tanah (spt dl bola poli atau tenis); **3** tendangan atau pukulan (dl sepak bola, bola voli) sebelum bola menyentuh tanah; (KBBI IV, 2008: 1550)

bo.la *n* **1** benda bulat yg dibuat dr karet dsb untuk bermain-main: *ia senang sekali bermain-main dng --*; **2** barang yg bentuknya menyerupai bulatan -- voli *Olr* **1** bola yg digunakan dl permainan olahraga bola voli; **2** permainan bola voli (KBBI IV, 2008: 204)

Dari kutipan tentang beberapa nama cabang olahraga di atas, ada tiga hal yang perlu dicatat. Pertama, tentang pingpong dan tenis meja pilihannya sudah sesuai dengan cara kerja leksikografis : yang diberi takrif ialah tenis meja, bukan pingpong. Hampir semua orang tahu bahwa nama induk organisasinya ialah Persatuan Tenis Meja Seluruh Indonesia atau PTMSI (bukan PPSI atau Persatuan Pingpong Seluruh Indonesia).

Kedua, sepak raga dan sepak takraw yang kedua-duanya dicantumkan sebagai sublema dari lema sepak. Yang benar dan resmi penggunaannya ialah sepak takraw, bukan sepak raga. Induk olahraganya bernama Persatuan Sepak Takraw Seluruh Indonesia

(PSTI), bukan Persatuan Sepak Raga Seluruh Indonesia (PSRI). Dengan demikian, pemberian takrifnya bukan pada sepak raga, tetapi seharusnya pada sepak takraw.

Ketiga, pencantuman bola voli baik sebagai sublema dari lema bola maupun sebagai lema, kedua cara itu tidak menyalahi kaidah leksikografis.

Sehubungan dengan hal itu, ketaatasasan dalam arti yang luas harus benar-benar dijadikan sebagai pemandu utama di dalam merumuskan atau mendeskripsikan makna. Oleh karena itu, pada Edisi III ketaatasasan perumusan makna itu diupayakan agar tetap terjaga, terutama untuk lema/sublema yang sudah pasti memiliki ciri makna yang sama, seperti pada

1. penyebutan bilangan: *puluh, ratus, ribu*, dst.
2. nama hari: *Ahad, Kamis, Sabtu*, dst.
3. nama bulan tahun Masehi: *Januari, Februari, Maret*, dst.
4. nama bulan tahun Kamariah: *Muharam, Safar, Rabiulawal*, dst.
5. nama Zodiak: *Akuarius, Aries, Leo*, dst.
6. satuan berat: *gram, ons, ton*, dst.

Mengenai hal itu, pada bagian akhir paparan tentang pemerian makna ini dicontohkan beberapa rumusan yang didasarkan pada ketaatasasan tersebut.

Bilangan

¹ra·tus *n* satuan bilangan kelipatan seratus yg dilambangkan dng dua nol (00) di belakang angka 1—9; (*KBBI IV*, 2008: 1147)

ri·bu *n* satuan bilangan kelipatan seribu yg dilambangkan dng tiga nol (000) di belakang angka 1—999: tiga --;(*KBBI IV*, 2008: 1173)

ju·ta *n* satuan bilangan kelipatan sejuta yg dilambangkan dng enam nol (000.000) di belakang angka 1 s.d. 999; (*KBBI IV*, 2008: 595)

Nama Hari

Se·nin *n* hari ke-2 dl jangka waktu satu minggu (*KBBI IV*, 2008: 1273)

Sab·tu *n* hari ke-7 dl jangka waktu satu minggu (*KBBI IV*, 2008: 1197)

Satuan Berat

ki·lo·gram *n* satuan ukuran berat (massa) 1.000 g (disingkat kg) (*KBBI IV*, 2008: 699)

ton *n* satuan ukuran berat 1.000 kg (*KBBI IV*, 2008: 1479)

Nama Bulan Tahun Masehi

Ja·nu·a·ri *n* bulan pertama tahun Masehi (31 hari) (*KBBI IV*, 2008: 567)

Agus·tus *n* bulan ke-8 tahun Masehi (31 hari) (*KBBI IV*, 2008: 18)

Nama Bulan Tahun Kamariah:

Mu ha·ram *n* bulan pertama tahun Hijriah (*KBBI IV*, 2008: 934)

Ra·bi·ul·a·wal *n* bulan ke-3 tahun Hijriah (30 hari); bulan Maulid (*KBBI IV*, 2008: 1127)

Nama Zodiak

A.ries /Ariés/ *n* **1** *Astron* nama gugus bintang di belahan langit sebelah utara khatulistiwa; **2** *Astrol* **Ari-es** zodiak pertama yg digambarkan dng domba jantan sbg lambang bagi orang yg lahir antara tanggal 21 Maret dan 19 April; Mesa; Hamal (*KBBI IV*, 2008: 85)

Leo /Léo/ *n* **1** *cak singa*; **2** *Astron* nama bintang di belahan langit sebelah utara khatulistiwa; **3** *Astrol* lambang ke-5 zodiak yg digambarkan dng singa sbg lambang bagi orang yg dilahirkan antara tanggal 23 Juli—22 Agustus (*KBBI IV*, 2008: 817)

Vir-go *n* **1** *Astron* nama gugus bintang yg terletak di antara Leo dan Libra pd deklinasi 00; **2** *Astrol* zodiak ke-6 yg digambarkan dng wanita kembar sbg lambang bagi orang lahir antara tanggal 23 Agustus—22 September; Kanya; Mayang; Sunbulat (*KBBI IV*, 2008: 1548)

Daftar Pustaka

Alwi Hasan, “Bahasa Indonesia dan Bahasa Melayu” Sekelumit Catatan”, Makalah Konvensi Bahasa dan Budaya dengan tema “Jejak Raja Ali Haji dalam Bahasa Media” dalam rangka kegiatan Hari Pers Nasional (HPN) 2015 yang dipusatkan di Provinsi Kepulauan Riau (Kepri) pada tanggal 5–9 Februari 2015.

Dewan Bahasa dan Pustaka, 2007, *Kamus Dewan* Edisi Keempat, Kuala Lumpur: Dewan Bahasa dan Pustaka

Pusat Bahasa, Departemen Pendidikan Nasional, 2008, *Kamus Besar Bahasa Indonesia* Edisi Keempat, Jakarta: PT Gramedia Pustaka Utama

(<http://lifestyle.okezone.com/read/2014/05/19/482/986931/pria-pendek-tak-perlu-minder-anda-lebih-panjang-umur>)

(<http://www.utusan.com.my/gaya-hidup/keluarga/tubuh-berhenti-meninggi-1.24161>)

Kamus sebagai Kitab Undang-Undang

Deny A. Kwary, Ph.D.

Universitas Airlangga

kwary@yahoo.com

Abstrak

Di beberapa negara yang memiliki kajian leksikografi yang cukup maju, banyak peneliti dan institusi yang telah memberikan perhatian besar terhadap perkembangan kamus. Hal ini berbeda dengan keadaan di Indonesia. Perkamusan di Indonesia nampaknya masih kurang berkembang dan kurang mendapatkan perhatian serius dari para peneliti dan institusi. Berbagai penelitian di negara-negara lain telah menunjukkan peran penting kamus di berbagai bidang, termasuk dalam pengambilan keputusan hukum. Dalam makalah ini, kamus diusulkan untuk dapat disejajarkan dengan kitab undang-undang. Kamus yang disusun dengan serius dan dengan prinsip leksikografi yang tepat akan menghasilkan rujukan yang sah dan berkualitas. Untuk mencapai tujuan tersebut, perlu ada fitur-fitur khusus yang dimasukkan ke dalam entri sehingga kamus tersebut dapat menjadi rujukan dalam memutuskan kasus hukum.

Kata kunci: kamus hukum; kamus umum; definisi; makna kata; entri kamus

I. PENDAHULUAN

Di beberapa negara, meskipun sudah tersedia kamus hukum, kamus umum tetap dijadikan rujukan dalam beberapa kasus hukum. Hal ini terkait dengan kenyataan bahwa kasus hukum cukup sering melibatkan permasalahan makna kata umum, bukan kata yang khusus berkaitan dengan ilmu hukum. Calhoun (2014: 248) mencatat bahwa di Amerika Serikat terdapat peningkatan penggunaan kamus umum di pengadilan selama dua puluh lima tahun terakhir ini. Pada tahun 2010, Pengadilan Tinggi (*Supreme Courts*) mengutip kamus empat kali lebih banyak dari pada di tahun 1985. Pengadilan cukup

banyak merujuk pada kamus umum karena penggunaan suatu kata di dalam produk hukum atau kasus hukum mengacu pada makna kata tersebut secara umum.

Berdasarkan data yang dikumpulkan oleh Brudney dan Baum (2013: 529), selama tahun 1986-2010, menunjukkan bahwa kamus umum adalah yang paling banyak dirujuk (74%), kamus hukum hanya dirujuk 45%, dan kamus kosakata teknis (misalnya kamus Akuntansi) hanya dirujuk 6% dari semua kasus hukum yang ada. Secara lebih rinci lagi, 21% pendapat mengutip kamus umum dan kamus hukum, 53% hanya kamus umum namun tidak kamus hukum, dan 24% hanya kamus hukum namun tidak kamus umum. Dengan demikian, dapat dikatakan bahwa lebih dari separuh kasus hukum hanya mengutip kamus umum. Hal ini menunjukkan peran penting kamus umum dalam dunia peradilan, khususnya di Amerika Serikat. Data lebih lanjut menunjukkan bahwa ada empat kamus umum yang paling sering dirujuk, yaitu *Webster's Third New International Dictionary* (36%), *the Oxford English Dictionary* (20 percent), *Webster's Dictionary Second Edition* (19 percent), and *the American Heritage Dictionary of the English Language* (15 percent).

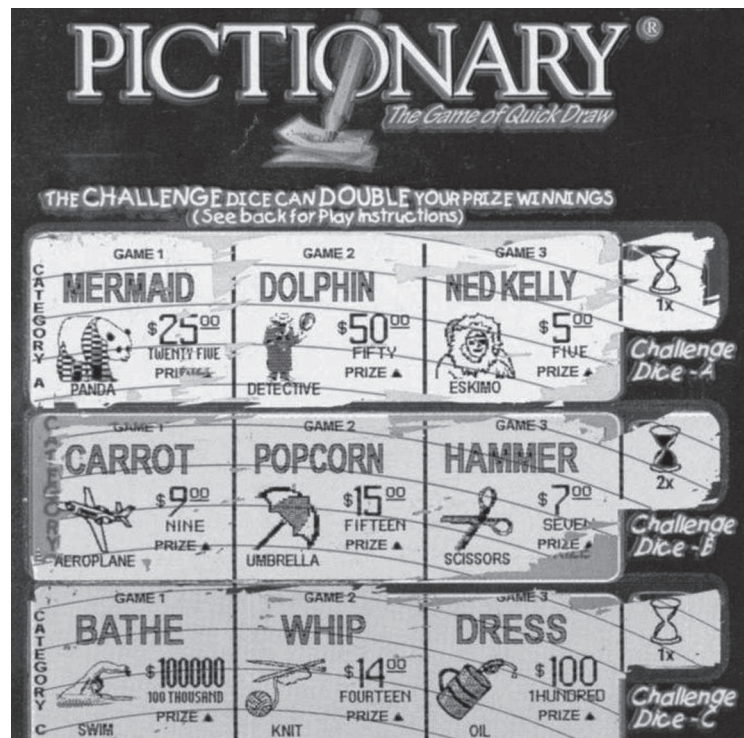
Di Indonesia, beberapa hakim dan pengacara sudah mulai merujuk pada kamus umum, khususnya Kamus Besar Bahasa Indonesia (KBBI). Sebenarnya KBBI (hingga edisi ke empat) masih kurang tepat untuk dijadikan rujukan mengenai makna kata, karena berbagai hal. Salah satunya adalah entri yang terlalu singkat, misalnya lema kepala yang hanya didefinisikan dengan sinonim, tanpa penjelasan dan contoh penggunaan. Sinonim kurang tepat digunakan untuk mendefinisikan lema di kamus besar, khususnya jika ingin dijadikan rujukan hukum, karena tidak ada dua kata yang maknanya persis sama. Dengan kata lain, tidak pernah ada sinonim seluruh (*true synonyms*). Setiap kata pasti memiliki perbedaan, meskipun kecil, dengan kata lain dalam hal makna dan penggunaannya. Oleh karena itu, jika kita ingin membuat kamus yang bisa dijadikan rujukan dalam kasus hukum, kamus yang kita buat harus memiliki fitur-fitur yang dibutuhkan dalam menyelesaikan kasus hukum.

Pembahasan dalam makalah ini dibagi dalam dua bagian utama. Bagian pertama membahas salah satu contoh kasus hukum yang melibatkan penafsiran makna suatu kata umum. Bagian kedua merumuskan usulan mengenai fitur-fitur yang perlu ada dalam kamus yang bisa berfungsi sebagai kitab undang-undang kata.

II. CONTOH KASUS HUKUM YANG MELIBATKAN KAMUS UMUM

Salah satu kasus hukum yang melibatkan makna kata dan cukup populer adalah kasus pengadilan federal antara Kuzmanovski dan New South Wales Lotteries Corporation pada tahun 2010 (The Australian Professional Liability Blog, 19 Oktober 2010). Keputusan apakah penggugat bisa memenangkan 100.000 dolar Australia bergantung pada apakah kata *bathe* berarti berenang. Penggugat merasa telah memenangkan lotere

karena mendapatkan kata BATHE yang berkaitan dengan gambar orang berenang. Akan tetapi, perusahaan lotere menganggap bahwa kata BATHE tidak sama artinya dengan berenang. Potongan dari lotere tersebut dapat dilihat di Gambar 1.



Gambar 1. Potongan lotere kasus makna kata

Dari potongan lotere yang ditunjukkan di Gambar 1, kita bisa melihat di ujung kiri bawah terdapat kata BATHE dengan hadiah sebesar 100.000 dolar Australia jika kata tersebut sesuai dengan gambar di bawahnya. Pada gambar di bawah kata tersebut, terlihat gambar orang berenang dan kata *swim*. Dengan demikian, jika *bathe* memiliki makna yang sama dengan *swim*, pemilik lembar lotere tersebut akan memenangkan 100.000 dolar Australia. Akan tetapi, perusahaan lotere menyatakan bahwa kata *bathe* tidak memiliki makna yang sama dengan *swim*. Kasus ini kemudian dibawa ke pengadilan.

Pihak perusahaan lotere menjelaskan pendapatnya dengan mengutip *Macquarie Dictionary* yang menyebutkan bahwa kata *bathe* hanya bermakna *swim* dalam bahasa Inggris Britania (*chiefly British*). Dengan demikian, *bathe* tidak bermakna *swim* dalam bahasa Inggris Australia. Untuk menyangkal pendapat tersebut, hakim yang bertugas saat itu mengutip definisi kata *bathe* dengan pemarkah taktransitif dari empat kamus, sebagai berikut:

- *Collins English Dictionary* (edisi ketiga) (Konsultan Khusus Australia: G.A. Wilkes dan W.A. Krebs): makna pertamanya adalah “to swim or paddle in a body of open water or a river”.

- *Australian Concise Oxford Dictionary of Current English* (edisi ketiga): makna pertamanya adalah “*immerse oneself in water; exp to swim*”.
- *Collins Australian Dictionary* (edisi kedelapan) yang awalnya dikutip oleh Kuzmanovski (sebagai penggugat): makna pertamanya adalah “*swim in open water for pleasure*”.
- *New Shorter Oxford English Dictionary* (Volume 1; 1993): makna keempatnya (yang merupakan makna pertama dengan pemarkah taktransitif) adalah “*Immerse oneself in water, esp in the sea, a river, a swimming pool etc, for recreation*”. Entri kamus ini juga menyebutkan sublema *bathing beauty* dengan definisi *an attractive woman in a swimsuit*, dan sublema *bathers (esp Austral.)* yang didefinisikan *swimtrunks, a swimming costume*.

Ke empat definisi di atas menunjukkan bahwa kata *bathe* memang bisa dianggap berkaitan dengan gambar orang berenang dan kata *swim*. Untuk menanggapi pendapat hakim yang mengutip definisi dari empat kamus di atas, pihak perusahaan lotere menyatakan bahwa tetap ada perbedaan antara *bathe* dan *swim* dalam konsep makna. Perusahaan lotere ini menyampaikan bahwa makna utama dari *swim* adalah menggerakkan tubuh melalui air (*propel the body through water*), sedangkan *bathe* hanya bermakna memasukkan tubuh ke air (*immersing the body in water*). Tanggapan tersebut juga dibantah oleh hakim, karena inti dari permainan lotere tersebut adalah apakah kata *bathe* sesuai dengan gambar orang berenang. Makna kata *swim* bukan bagian dari permainan lotere ini. Dengan demikian, hakim memutuskan untuk memenangkan penggugat, yaitu Kuzmanovski, dan meminta New South Wales Lotteries Corporation untuk membayar sesuai tuntutan Kuzmanovski.

III. FITUR ENTRI KAMUS SEBAGAI KITAB UNDANG-UNDANG

Beberapa kritikus telah menyampaikan permasalahan mengenai kamus umum yang seharusnya digunakan sebagai rujukan dalam kasus hukum. Jika hakim hanya menggunakan satu kamus, kemungkinan definisi yang dipilih akan memberatkan salah satu pihak. Jika ada beberapa kamus yang digunakan, kemungkinan ada pertentangan juga dalam definisinya. Hingga saat ini, kamus umum memang dibuat bukan untuk menjadi rujukan di pengadilan. Secara tradisional, kamus dibuat untuk memberikan informasi mengenai makna kata. Dalam perkembangan selanjutnya, kamus dibuat berdasarkan kebutuhan pengguna sasaran. Dengan demikian, jika kita ingin membuat kamus yang bisa dijadikan rujukan di pengadilan atau kamus yang bisa dijadikan Kitab Undang-undang Hukum Kata (agar sejajar dengan rujukan lain yang digunakan dalam kasus hukum, misalnya Kitab Undang-undang Hukum Pidana), ada beberapa fitur yang perlu dimasukkan ketika menyusun entri kamus tersebut.

Fitur suatu entri atau struktur internal entri kamus bervariasi antara satu kamus dengan kamus yang lain. Hanks (2006) menyebutkan delapan fitur atau unsur yang

umumnya ada di dalam entri kamus, yaitu lema kepala, pengucapan, gramatika, definisi, contoh, frasa, penggunaan, dan etimologi. Untuk kamus yang bisa digunakan sebagai rujukan di pengadilan, dalam makalah ini diusulkan minimal ada enam fitur, yaitu lema, informasi gramatika, definisi, contoh penggunaan, label, dan masa berlaku.

3. 1. Lema

Kamus untuk fungsi reseptif biasanya memiliki lebih banyak lema (lema kepala dan sublema) daripada kamus untuk fungsi produktif. Oleh karena fungsi utama kamus ini adalah reseptif, maka lema kepala yang dimasukkan sebisa mungkin mencakup semua jenis butir leksikal (*lexical item*). Atkins dan Rundell (2008: 164) membagi jenis butir leksikal ke dalam empat kelompok, yaitu: kata sederhana, singkatan, kata parsial, dan kata rangkap (*multiword expressions*). Ke empat kelompok butir leksikal tersebut perlu dimasukkan sebagai lema kepala atau sublema di kamus ini.

3. 2. Informasi Gramatika

Lema kepala dan sublema perlu disertai dengan informasi gramatika yang bisa menunjukkan perbedaan makna. Informasi ini bisa mencakup kelas kata, transitivitas, kolokasi, dll. Informasi kelas kata diperlukan karena suatu kata bisa memiliki makna yang berbeda jika kelas katanya berbeda. Informasi transitivitas diperlukan jika makna suatu kata bisa berbeda jika digunakan dalam bentuk transitif dan taktransitif. Contohnya adalah kata *bathe* di kasus yang dibahas sebelumnya. Makna kata *bathe* sebagai verba transitif agak berbeda dari makna kata *bathe* sebagai verba taktransitif. Menurut *Oxford Advanced Learner's Dictionary* edisi kesembilan (OALD9), kata *bathe* sebagai verba transitif berarti *to wash something with water, especially a part of your body*; sedangkan sebagai verba taktransitif definisinya adalah *to go swimming in the sea, a river, etc. for enjoyment*. Hal tersebut menunjukkan bahwa informasi transitivitas bermanfaat dalam menentukan makna kata. Pola kolokasi, jika membedakan makna kata, juga perlu dimasukkan. Contohnya verba *hump*. OALD9 menyebutkan bahwa jika verba *hump* diikuti oleh *something*, maknanya adalah *to carry something heavy* (mengangkat sesuatu yang berat); sedangkan jika diikuti oleh *someone*, maknanya adalah *to have sex with somebody* (melakukan hubungan badan dengan seseorang).

3. 3. Definisi

Ada beberapa cara untuk menulis definisi suatu kata, misalnya dengan menggunakan genus-diferensia dan sinonim. Dalam perkembangan lebih lanjut, Sinclair (1987) mengusulkan cara pendefinisian yang disebut FSD (*full sentence definition* 'definisi kalimat lengkap'). Cara pendefinisian ini cukup banyak dikritik karena membuat entri kamus menjadi lebih panjang. Untuk kamus ini, definisi yang lebih panjang itulah yang dibutuhkan. Bahkan, alih-alih menyebutnya definisi, kita bisa menyebutnya penjelasan.

Jika kita kembali ke kamus COBUILD (1987), di bagian petunjuk penggunaan kamus ini ada subjudul *Explanation of meaning and use* ('penjelasan makna dan penggunaan'), dan tidak ada kata 'definisi'. Penggunaan kata 'penjelasan', alih-alih kata 'definisi', sebenarnya sudah disebutkan juga jauh sebelumnya oleh Samuel Johnson di bagian Prakata dari *Dictionary of the English Language* (1755). Dalam makalah ini, kata 'definisi' dan 'penjelasan' untuk lema atau suatu kata akan digunakan secara bergantian, namun merujuk pada hal yang sama.

Seperti yang disebutkan dalam subbagian penjelasan mengenai lema, fungsi utama dari kamus ini adalah reseptif. Akan tetapi, dalam pembuatan definisi, kita sebaiknya tidak mengikuti prinsip pembuatan definisi seperti yang digunakan di kamus reseptif. Secara prinsip, definisi di kamus reseptif bertujuan untuk menyediakan petunjuk dan asosiasi yang mengaitkan sesuatu (kata) yang tidak diketahui dengan sesuatu yang diketahui (Bolinger 1965: 572). Cara ini akan memungkinkan pengguna kamus untuk mengetahui makna dari suatu kata yang dicari di kamus. Akan tetapi, pengguna kamus tidak akan mengetahui penggunaan kata tersebut. Oleh karena itu, di kamus ini, meskipun fungsi utamanya adalah reseptif, cara penulisan definisinya mengacu pada fungsi produktif. Salah satu contoh penerapannya adalah pada kata bahasa Inggris *carrion* yang disebutkan oleh Fillmore (2003).

Kamus yang menggunakan definisi dengan fungsi reseptif, akan mendefinisikan kata *carrion* sebagai berikut: *the rotting meat of a dead animal* ('daging yang sudah membusuk dari hewan yang sudah mati'). Definisi tersebut bisa memenuhi kebutuhan pengguna untuk mengetahui makna kata *carrion*. Dengan kata lain, fungsi reseptif telah terpenuhi. Akan tetapi, jika kita ingin menggunakan kata tersebut secara produktif dalam konteks yang tepat, perlu ada penjelasan lebih lanjut. Kata *carrion* tidak berlaku untuk daging yang membusuk di dalam kulkas, dan juga tidak berlaku untuk daging hewan yang tidak sengaja terinjak ketika seseorang berada di hutan. Oleh karena itu, perlu ada penjelasan tambahan bahwa *carrion* hanya digunakan untuk makanan dari hewan pemakan bangkai. Informasi semacam ini juga diperlukan untuk kamus yang akan digunakan untuk rujukan di pengadilan. Jadi, pemilihan lema kamus ini merujuk pada fungsi reseptif, namun penulisan definisinya merujuk pada fungsi produktif.

3. 4. Contoh penggunaan

Kamus untuk fungsi reseptif tidak terlalu banyak menggunakan contoh penggunaan. Akan tetapi, untuk kamus ini, penyediaan contoh penggunaan sangat disarankan. Contoh penggunaan merupakan landasan untuk mendeskripsikan suatu kata dan menjadi bukti cara penggunaan kata tersebut oleh suatu masyarakat (Atkins dan Rundell 2008: 2). Oleh karena itu, contoh yang otentik perlu digunakan, bukan contoh yang dikarang oleh leksikografer. Dalam hal ini, disarankan mengikuti penggunaan contoh di *Oxford English Dictionary*. Semua contoh penggunaan kata di entri kamus tersebut diambil dari teks otentik untuk menunjukkan secara objektif bahwa kata tersebut dan makna yang disebutkan memang ditemukan dalam bahasa tersebut (Simpson 2003: 268).

3. 5. Label

Selain definisi dan contoh penggunaan, kita juga perlu menggunakan label sebagai pemarah penggunaan suatu kata. Kata-kata yang tidak diberi label akan dianggap sebagai kata yang memiliki penggunaan standard (Kipfer 1984: 140). Atkins dan Rundell (2008: 228-231) menyebutkan delapan jenis label yang biasanya digunakan di dalam kamus. Dari delapan jenis label tersebut, tujuh di antaranya relevan untuk dimasukkan ke dalam kamus yang akan berfungsi sebagai kitab undang-undang kata. Label mengenai waktu penggunaan, yang menunjukkan apakah suatu kata sudah obsolet atau belum, tidak perlu dimasukkan sebagai label, namun dimasukkan dalam fitur yang dijelaskan di subbab 3.6. dalam makalah ini. Penjelasan mengenai tujuh label yang dimasukkan ke dalam kamus ini adalah sebagai berikut:

- Label domain: Label ini menunjukkan bidang ilmu atau bidang khusus, misalnya teknik, kedokteran, tenis, puisi, dll. Label ini penting karena suatu kata yang sama bisa memiliki makna yang berbeda di domain yang berbeda. Misalnya nomina *love* tidak bermakna cinta di domain tenis, melainkan skor nol.
- Label wilayah: Dalam bahasa Inggris, label wilayah sangat penting karena suatu kata mungkin hanya digunakan atau lebih sering digunakan di suatu wilayah tertentu, namun tidak di wilayah yang lain. Suatu kata juga bisa memiliki makna yang berbeda di wilayah yang berbeda. Contohnya makna kata *hotel* dalam bahasa Inggris Australia bisa berarti pub, dan dalam bahasa Inggris India bisa berarti restoran. Di Indonesia, label wilayah ini juga penting karena negara Indonesia memiliki berbagai bahasa daerah yang masing-masing memberikan kontribusi ke bahasa Indonesia.
- Label register: Label ini memiliki berbagai fungsi, misalnya penanda formalitas, penanda slang, dll. Kamus sebagai kitab undang-undang kata perlu memiliki label register karena makna suatu kata bisa berbeda pada register yang berbeda. Misalnya kata *acid* yang secara umum berarti ‘senyawa kimia/asam’, memiliki makna obat-obatan terlarang jika diberi label slang (lihat OALD9).
- Label genre: Label ini menunjukkan penggunaan kata pada genre tertentu (misalnya karya sastra, surat kabar, dll.). Contohnya kata *cunning*. Kata tersebut umumnya merujuk pada seseorang yang pandai menipu. OALD9 mendefinisikannya *able to get what you want in a clever way, especially by tricking or cheating somebody* (‘bisa memperoleh apa yang diinginkan dengan cara yang cerdik, khususnya dengan cara mencurangi atau menipu seseorang’). Akan tetapi, di genre karya sastra menurut OED2, *cunning* berarti *knowledge how to do a thing; ability, skill, expertness, dexterity, cleverness* (‘pengetahuan mengenai cara melakukan sesuatu; kemampuan, keterampilan, keahlian, kecekatan, dan kepandaian’).
- Label sikap. Label ini menunjukkan apakah suatu kata memiliki makna yang disetujui atau tidak disetujui (*approval* atau *disapproval*). Contohnya kata *snazzy*

dan *jazzy*. Kedua kata tersebut memiliki makna yang sama, yaitu ‘bergaya yang menarik perhatian’. Akan tetapi, *snazzy* memiliki label *approval* (lihat CALD3), dan *jazzy* memiliki label *disapproving* (lihat OALD9).

- Label jenis makna. Label ini menyatakan apakah suatu kata perlu dimaknai secara figuratif atau tidak. Misalnya kata *screech*. Menurut CALD3, makna dasar dari kata ini adalah *to make an unpleasant loud high noise* (‘membuat suara yang gaduh’), namun jika diberi label figuratif, maknanya menjadi *stop very suddenly* (‘berhenti mendadak’).

3. 6. Masa berlaku

Kamus pemelajar biasanya hanya berisi kata dan makna kata yang masih digunakan di masyarakat. Hal ini berbeda dari kamus umum yang mencatat semua kata dan makna kata yang masih digunakan dan yang sudah tidak digunakan. OED2 menggunakan simbol salib untuk menyatakan bahwa suatu kata atau makna kata sudah tidak digunakan lagi. Cara tersebut kurang spesifik karena kita tidak mengetahui secara jelas kapan makna kata tersebut tidak digunakan lagi. Misalnya, jika seseorang dituntut karena menyebut orang lain *gay*, kita perlu mengetahui kapan sebutan tersebut disampaikan dan kapan makna *gay* berubah. Kata *gay* bisa berarti ‘homoseksual’, namun juga bisa berarti ‘riang gembira’ (*happy and full of fun*). OALD9 hanya memberikan label *old-fashioned* untuk kata *gay* yang berarti *happy and full of fun*. Pemberian rincian tahun penggunaan akan sangat membantu hakim memutuskan apakah dalam kasus di atas telah terjadi pencemaran nama baik atau tidak. Untuk kamus yang berfungsi sebagai kitab undang-undang, kita perlu menyediakan informasi masa berlaku, yaitu kapan suatu makna mulai berlaku pada suatu kata, dan kapan makna tersebut tidak digunakan lagi.

IV. SIMPULAN

Berdasarkan pembahasan di atas, dapat disimpulkan bahwa kamus sebenarnya bisa dijadikan kitab undang-undang hukum. Akan tetapi, perlu ada penyesuaian dalam informasi yang disediakan di kamus tersebut. Misalnya, meskipun fungsi kamus ini adalah reseptif dan pemilihan lema juga merujuk pada prinsip yang diterapkan pada kamus dengan fungsi reseptif, namun pembuatan definisi harus merujuk pada prinsip yang diterapkan pada kamus dengan fungsi produktif. Di makalah ini, diusulkan agar kamus ini minimal memiliki enam fitur pada struktur internal entrinya, yaitu lema, informasi gramatika, definisi, contoh penggunaan, label, dan masa berlaku.

Daftar Pustaka

- AHD – *American Heritage Dictionary of the English Language*. 1982. Boston: Houghton Mifflin.
- Atkins, B. T. S. dan Rundell, M. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Australian Concise Oxford Dictionary of Current English* (edisi 3). 1987. Oxford: Oxford University Press.
- Bolinger, D. 1965. The Atomization of Meaning. *Language* 41: 555–573.
- Brudney, J. J. dan Baum, L. 2013. Oasis or Mirage: The Supreme Court’s Thirst for Dictionaries in the Rehnquist and Roberts Eras. *William & Mary Law Review* 55(2): 483-580.
- CALD3 – *Cambridge Advanced Learner’s Dictionary*. 2008. Cambridge: Cambridge University Press.
- Calhoun, J. 2014. Measuring the Fortress: Explaining Trends in Supreme Court and Circuit Court Dictionary Use. *Yale Law Journal* 124(2): 284-526.
- Collins Australian Dictionary* (edisi 8). 2005. Sydney: HarperCollins.
- Collins English Dictionary* (edisi 3). 2004. Glassgow: HarperCollins.
- Fillmore, C. J. 2003. Double-decker Definitions: the Role of Frames in Meaning Explanations. *Sign Language Studies* 3(3): 263–295.
- Hanks, P. 2006. Lexicography: Overview. Dalam *Encyclopedia of Language and Linguistics*. Elsevier.
- KBBI – *Kamus Besar Bahasa Indonesia*. 2008. Jakarta: Gramedia Pustaka Utama.
- Kipfer, B. A. 1984. *Workbook on Lexicography*. Exeter: University of Exeter.
- Macquarie Dictionary*. 1982. Sydney: Macquarie Library.
- New Shorter Oxford English Dictionary*. 1993. Oxford: Oxford University Press.
- OED2 – *Oxford English Dictionary* (edisi 2). 1989. Oxford: Oxford University Press.
- OALD9 – *Oxford Advanced Learner’s Dictionary* (edisi 9). 2015. Oxford: Oxford University Press.
- Simpson, J. 2003. The Production and Use of Occurrence Examples. Dalam P. van Sterkenburg (ed.). *A Practical Guide to Lexicography*. Amsterdam: John Benjamins. 260–272.

Sinclair, J. M. 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London: Collins.

The Australian Professional Liability Blog, 19 Oktober 2010.

Webster's Dictionary (edisi 2). 1939. Springfield: G. & C. Merriam Co.

Webster's Third New International Dictionary. 1961. Springfield: G. & C. Merriam Co.

Pemanfaatan Korpus dalam Analisis Makna Kata Bersinonim *mau, ingin, hendak, dan akan*

Dewi Puspita

Pusat Pengembangan dan Pelindungan
Badan Pengembangan dan Pembinaan Bahasa
dewi.puspita@kemdikbud.go.id

Abstrak

Linguistik korpus sudah lama digunakan untuk membantu penyusunan kamus berbahasa Inggris, khususnya dalam pendefinisian lema karena selain dapat memberi definisi yang lebih subjektif, korpus juga dapat memberikan data faktual tentang sebuah kata. Data tersebut dapat juga digunakan untuk membedakan dua atau lebih kata yang berdekatan makna. Sayangnya, lema-lema dalam Kamus Besar Bahasa Indonesia masih didefinisikan berdasarkan intuisi penyusunnya, belum dengan bantuan korpus.

Makalah ini akan mengulas pendefinisian kata bersinonim *mau, ingin, hendak, dan akan* dalam Kamus Besar Bahasa Indonesia edisi 4 dengan menggunakan pendekatan korpus linguistik. Dengan menggunakan korpus bahasa Indonesia berbasis web dan kompilasi berbagai teks, hasil analisis dari penelitian ini diharapkan dapat digunakan untuk membantu pemutakhiran Kamus Besar Bahasa Indonesia.

I. PENDAHULUAN

Latar belakang penulisan makalah ini sebenarnya adalah rasa ketidaknyamanan saya ketika mendengar anak saya bercerita bahwa temannya “ingin” jatuh pada saat ia tidak dapat mengendalikan sepedanya. Spontan saya mengoreksinya dengan mengatakan bahwa yang betul adalah “mau jatuh” bukan “ingin jatuh”. Anak saya kemudian bertanya “apa bedanya?” Pada saat itu saya menjawab, “ingin” muncul dari diri sendiri, sedangkan “mau” adalah sesuatu yang akan terjadi. Tentu saja jawaban saya tadi adalah jawaban yang didasarkan pada pengetahuan dan nilai rasa pribadi terhadap kedua kata itu,

sehingga jawabannya sangat subjektif. Orang lain, termasuk anak saya sendiri, mungkin mempunyai persepsi lain mengenai definisi kata tersebut.

Kami lalu mencoba mencari makna kedua kata tersebut dalam KBBI. *Kamus Besar Bahasa Indonesia Edisi Keempat* mendefinisikan *ingin* dan *mau* sebagai berikut.

ingin *adv* hendak; mau; berhasrat: *dia -- mencoba apakah telur merpati juga enak dimakan* (2008; 536)

mau **1** *adv* sungguh-sungguh suka hendak; suka akan; sudi: *ia -- datang kalau dijemput*; **2** *adv* akan; hendak: *ibu -- berangkat ke Surabaya*; **3** *n* kehendak; maksud: *apa --mu datang ke sini?*; --nya begini (2008; 890)

Merujuk pada definisi seperti yang disajikan di atas, pemahaman saya mengenai kata *ingin* sepertinya sudah tepat, yaitu ‘berhasrat’, begitu pula dengan pemahaman saya tentang kata *mau* yang didefinisikan dengan ‘akan’ dan ‘hendak’, bukan dengan ‘ingin’.

Namun, pertanyaan selanjutnya muncul ketika saya kemudian juga memeriksa definisi kata *hendak* dan *akan* di kamus yang sama seperti berikut.

hendak *adv* mau; akan; bermaksud akan: *pamannya -- hendak pergi jauh*; *ia -- mengadakan pesta pernikahan* (2008; 491)

akan *adv* (untuk menyatakan sesuatu yg hendak terjadi, berarti) hendak: *disangkanya hari – hujan* (2008; 25)

Jika *ingin*, *mau*, dan *akan* didefinisikan dengan ‘hendak’ sedangkan definisi dari *hendak* adalah ‘mau’ dan ‘akan’, lalu apa perbedaan dari keempat kata tersebut?

Dari situlah, ide untuk menganalisis kata-kata tersebut kemudian muncul. Bagaimanapun, pasti ada perbedaan makna dari setiap kata yang bersinonim mengingat sinonim mutlak sangatlah sedikit (Cruse, 2004). Penelitian kecil yang terangkum dalam makalah ini dilakukan untuk mendapatkan jawaban dari pertanyaan-pertanyaan di atas. Lebih jelasnya, tujuan dari kajian ini adalah untuk mengetahui perbedaan makna *mau*, *ingin*, *hendak*, dan *akan* berdasarkan data korpus yang hasilnya diharapkan dapat memberi pencerahan bagi pengguna bahasa juga untuk membantu perbaikan definisi keempat kata tersebut (jika ada) dalam Kamus Besar Bahasa Indonesia. Selain itu, kajian ini juga bermaksud untuk menunjukkan bagaimana data korpus dan pendekatan linguistik korpus dapat digunakan untuk mendefinisikan sebuah kata pada khususnya dan membantu penyusunan kamus pada umumnya.

Kajian ini akan dibatasi ruang lingkupnya pada empat kata bersinonim seperti yang telah disebutkan sebelumnya. Walaupun sebetulnya masih ada sinonim atau bentuk lain dari keempat kata yang dijadikan objek kajian, seperti *bakal*, *pingin*, *kepingin*, atau *pengen* (dalam bentuk cakapan). Bentuk turunan dari kata-kata tersebut, seperti *kemauan*, *keinginan*, atau *kehendak* juga tidak akan diikutsertakan dalam kajian ini.

Pusat Bahasa (2007) dalam Buku Praktis Bahasa Indonesia 2, Bab Pilihan Kata sudah banyak menerangkan perbedaan penggunaan kata-kata bersinonim. Di antara kata-kata bersinonim yang dijelaskan adalah *masing-masing dan tiap-tiap, jam dan pukul, setengah dan separo, pertandingan dan perlombaan, di dan pada*, serta *sudah dan telah*. Kata-kata yang bermiripan itu diterangkan persamaan dan perbedaannya secara semantik dan pragmatik disertai dengan contoh penggunaannya dalam kalimat. Penjelasan yang diberikan sudah cukup dapat dimengerti dan diterima, tetapi argumennya akan lebih kuat lagi jika didukung oleh data korpus. Sayangnya, keterangan mengenai persamaan dan perbedaan dari kata-kata bersinonim di atas tidak ditampilkan dalam definisi entri kata tersebut dalam Kamus Besar Bahasa Indonesia yang juga terbitan Badan Bahasa.

Mengenai persamaan dan perbedaan kata bersinonim *sudah* dan *telah*, Kaswanti (2011) telah lama mengkajinya. Penelitian tentang kedua kata tersebut telah dimulai sejak tahun 1978 dan masih berlanjut hingga saat ini. Pada awalnya, Kaswanti menggunakan metode pencatatan manual untuk setiap penggunaan kata *sudah* dan *telah* dalam konteks yang ditemui hingga akhirnya kini beliau menggunakan korpus.

Sementara itu, kajian-kajian berbasis korpus untuk kata bersinonim dalam bahasa Inggris sudah banyak sekali dilakukan. Dalam kelas linguistik korpus yang pernah penulis ikuti, kajian kata bersinonim dengan pendekatan korpus bahkan dijadikan tugas akhir untuk mata kuliah tersebut. Mahasiswa diminta untuk memilih dua hingga lima kata atau frasa dan melakukan studi korpus dari kata atau frasa tersebut untuk mencari persamaan dan/atau perbedaannya dilihat dari berbagai sisi, baik sisi semantik, pragmatik, maupun tata bahasanya. Kata atau frasa bahasa Inggris yang sudah pernah dikaji sebelumnya diantaranya adalah *between* dan *through*; *immense*, *enormous* dan *massive*; *reason to* dan *reason for*; *on the other hand* dan *on the contrary*. Hasil yang didapat dari kajian-kajian di atas sebagian besar menunjukkan bahwa pendekatan korpus linguistik dapat secara efektif memperlihatkan pola persamaan dan/atau perbedaan dari kata-kata yang dijadikan objek kajian.

II. KERANGKA TEORI DAN METODE PENELITIAN

Penggunaan korpus untuk berbagai keperluan kebahasaan telah menjadi tren pada saat ini. Metode penganalisisan gejala-gejala linguistik dengan menggunakan korpus dikenal sebagai linguistik korpus. Beberapa akademisi menganggap linguistik korpus sebagai pendekatan teoretis untuk mempelajari bahasa (Tognini-Bonelli, 2001; Teubert, 2005; Gries, 2009; McEnery dan Hardie, 2012). Sebagai metodologi, karakteristik utama korpus linguistik terdapat pada pendekatan empirisnya, fokus pada data otentik, pendekatan makna, dan penggunaan perangkat digital untuk memproses data bahasa yang memberikan hasil yang objektif. Namun demikian, hasil yang diberikan oleh

perangkat korpus masih perlu ditafsirkan untuk dapat difahami. Ketika data-data tersebut ditafsirkan, pada saat itulah teori dan panduan hipotesis linguistik diperlukan (Hunston, 2002).

Selain untuk pengajaran bahasa, korpus linguistik juga dapat diaplikasikan untuk penyusunan kamus, penerjemahan, analisis wacana, dan berbagai keperluan lainnya. Hunston (2002) menyatakan bahwa kontribusi utama korpus linguistik adalah di bidang pengajaran bahasa dan penerjemahan, aplikasi lainnya yaitu pada bidang leksikografi dan terminologi, studi ideologi dan budaya, karakterisasi register dan genre, serta di bidang linguistik forensik. Dalam penyusunan kamus, korpus dan linguistik korpus sangat membantu dalam proses pendefinisian terutama untuk mendeskripsikan kata sesuai dengan konteks yang ada dan digunakan oleh masyarakat. Korpus linguistik merupakan pendekatan lain yang lebih mutakhir dalam penyusunan kamus dan dianggap lebih faktual untuk dapat mengetahui makna sebuah kata. Hal tersebut juga berlaku untuk pendefinisian kata-kata bersinonim.

Selama ini, metode yang biasa digunakan oleh para linguist untuk mengetahui perbedaan makna kata yang bersinonim dan mengetahui perbedaan ciri semantik antara satu kata dengan kata yang lain adalah metode analisis komponen (Subroto, 2011). Dengan mempertimbangkan beberapa keunggulan korpus dan linguistik korpus seperti yang telah disampaikan sebelumnya, metode yang digunakan dalam kajian ini adalah metode korpus karena dalam korpus terdapat sekumpulan data autentik dalam jumlah besar yang diharapkan dapat memberi informasi lebih jelas mengenai perbedaan makna kata yang bermiripan.

Kajian ini akan menggunakan dua perangkat korpus. Data korpus pertama yang akan digunakan adalah data korpus bahasa Indonesia dari Leipzig Corpora yang dapat diakses secara bebas melalui jaringan internet dengan mengklik tautan berikut: http://corpora.uni-leipzig.de/en?corpusId=ind_mixed_2013. Data korpus bahasa Indonesia yang terdapat dalam Leipzig Corpora berisi teks-teks berbahasa Indonesia yang diambil dari web pada tahun 2013. Korpus tersebut memuat 74.329.815 kalimat, 7.964.109 tipe, dan 1.206.281.985 token. Salah satu kesulitan menggunakan Leipzig Corpora adalah tidak tersedianya menu KWIC yang dapat memudahkan pengguna melihat pola kata secara vertikal. Akan tetapi, Leipzig Corpora menyediakan grafik yang memperlihatkan hubungan suatu kata dengan kata-kata terdekat yang berkolokasi di depan dan di belakang kata tersebut.

Selain data korpus dari Leipzig Corpora, data juga diambil dari korpus IndonesianWaC dalam Sketch Engine. IndonesianWaC memiliki 109.281.359 token atau kata yang sudah dilabeli sehingga dapat memberi informasi lebih dibanding korpus yang lain. Sketch Engine juga memiliki keunggulan dibanding perangkat korpus lainnya, yaitu tersedianya menu *word sketch* dan *sketch diff*. Menu *word sketch* menampilkan ringkasan dari perilaku gramatikal dan kolokasional sebuah kata (Kilgariff et al, 2004). Sementara itu, hasil dari *sketch diff* akan menampilkan perbedaan dari dua kata sehingga dapat terlihat bagaimana sepasang kata yang bersinonim berbeda satu sama lain.

Prosedur pengambilan data dari kedua korpus tersebut beserta analisisnya akan berbeda karena fasilitas yang disediakan oleh kedua laman korpus tadi juga berbeda. Data yang diperoleh dari Leipzig Corpora berupa data konkordansi, kolokasi, dan grafik. Sementara itu, data yang diambil dari Sketch Engine berupa data *word sketch* dan *sketch diff*.

Data yang diperoleh dari kedua korpus tersebut akan dianalisis melalui kolokasinya. Hal ini dilakukan karena beberapa kolokasi dapat memberi informasi yang tepat mengenai konsep suatu kata dalam konteksnya. Kolokasi dapat menyediakan “analisis semantik dari sebuah kata” (Sinclair, 1991) dan juga dapat “menyampaikan pesan secara implisit” (Hunston, 2002).

III. ANALISIS PENELITIAN

3.1. Analisis data dari Leipzig Corpora

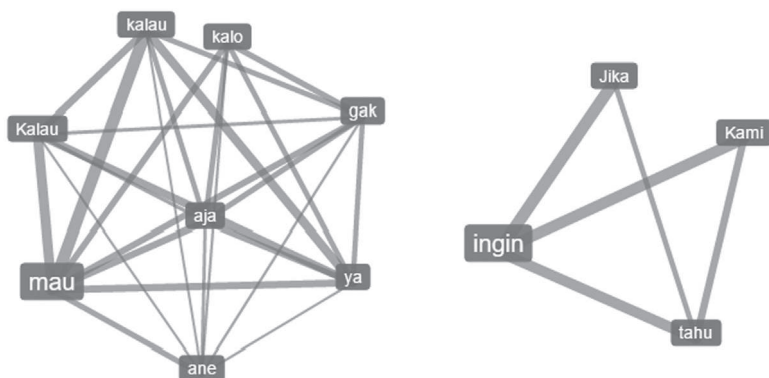
Sebelum masuk ke analisis, terlebih dahulu akan dijelaskan mengenai profil kata *mau*, *ingin*, *hendak*, dan *akan* dalam Leipzig Corpora berdasarkan frekuensi kemunculannya yang dapat dilihat dalam Tabel 1. Dari keseluruhan 1.206.281.985 token dalam korpus bahasa Indonesia di Leipzig Corpora, kata *hendak* muncul sebanyak 100.270 kali, *mau* muncul sebanyak 699.957 kali, *ingin* sebanyak 840.371 kali, dan terbanyak *akan* sebanyak 6.676.654 kali.

Kata	Frekuensi kemunculan
mau	699.957
ingin	840.371
hendak	100.270
akan	6.676.654

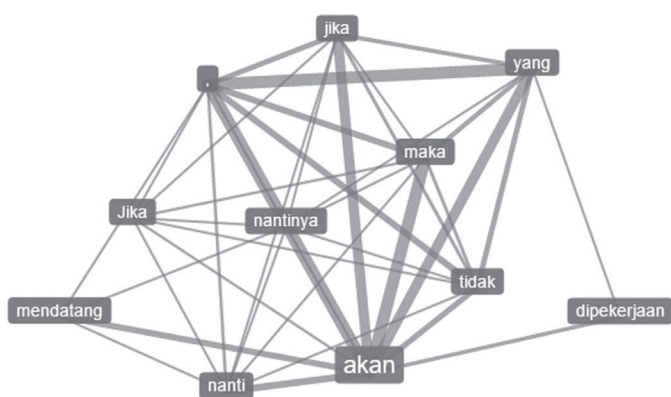
Tabel 1. Frekuensi kemunculan kata *mau*, *ingin*, *hendak*, dan *akan*

Kemunculan kata *akan* jauh lebih banyak dibandingkan tiga kata lainnya karena akan yang muncul dalam korpus bukan hanya *akan* yang memiliki makna mirip dengan *mau*, *ingin*, dan *hendak* saja. Sayangnya, dalam Leipzig Corpora pengguna tidak dapat menapis pencarian kata berdasarkan kelas katanya.

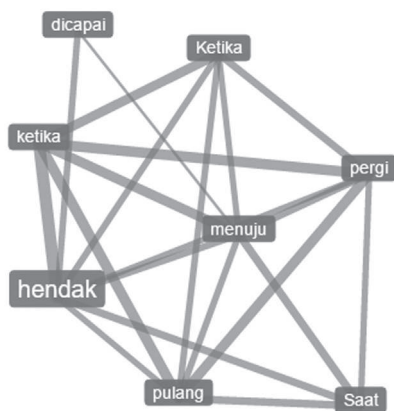
Selanjutnya, seperti yang telah disinggung sebelumnya, agak sulit menganalisis keempat kata yang dikaji melalui baris konkordans karena data disajikan per kalimat. Kata yang akan dianalisis tidak diletakkan dalam satu kolom yang sejajar. Oleh karena itu, data dari korpus ini dianalisis berdasarkan grafik kolokasi yang disajikan dalam gambar-gambar berikut.



Gambar 1. Kolokasi dari kata mau dan ingin dalam Leipzig Corpora



Gambar 2. Kolokasi dari kata akan dalam Leipzig Corpora



Gambar 3. Kolokasi dari kata hendak dalam Leipzig Corpora

Gambar 1 menunjukkan bahwa kata *mau* banyak berkolokasi dengan kata cakapan seperti *kalo*, *gak*, *aja*, dan *ane*. Hal ini menunjukkan bahwa kata tersebut dalam korpus Leipzig berasal dari konteks nonformal. Sementara itu, kata *ingin* digambarkan hanya berkolokasi dengan tiga kata, yaitu *jika*, *kami*, dan *tahu*. *Jika* dalam KBBI adalah kata penghubung untuk menandai syarat.

Walaupun agak sulit untuk menjabarkan Gambar 2 karena banyak kata yang kemungkinan berkolokasi dengan *akan* yang bukan bersinonim dengan *mau*, *ingin*, dan *hendak*, ada satu kata yang dicurigai berkolokasi dengan *akan* dimaksud yang tidak ditemukan pada grafik kolokasi ketiga kata lainnya, yaitu kata *tidak*. Sementara itu, Gambar 3 memperlihatkan bahwa kata *hendak* berkolokasi dengan keterangan waktu (ketika, saat) dan kata kerja yang menunjukkan tujuan (pulang, pergi, menuju).

3.2. Analisis data dari IndonesianWaC Sketch Engine

Dalam korpus IndonesianWaC dalam Sketch Engine, yang jumlah tokennya lebih sedikit dibanding dalam korpus Leipzig, frekuensi kemunculan kata *mau*, *ingin*, *akan* dan *hendak* dapat dilihat pada Tabel 2 berikut.

Kata	Frekuensi kemunculan
mau	49.462
ingin	61.628
hendak	14.297
akan	452.763

Tabel 2. Frekuensi kemunculan kata *mau*, *ingin*, *hendak*, dan *akan*

Seperti halnya Tabel 1, data pada Tabel 2 juga menunjukkan bahwa frekuensi kemunculan kata *hendak* paling rendah sementara kata *akan* paling tinggi. Penjelasan untuk tingginya kemunculan kata *akan* sama seperti sebelumnya, yaitu karena kata *akan* memiliki homonim dan banyak polisem. Dari data juga dapat dilihat bahwa kata *ingin* lebih banyak muncul dibanding kata *mau*.

Data yang didapat melalui fasilitas *word sketch* sedianya dapat memperlihatkan perilaku gramatikal dan kolokasional dari kata yang dianalisis. Sayangnya, mungkin karena program yang belum sempurna atau kesalahan teknis, data yang diperoleh tidak sesuai dengan yang diharapkan. Dalam *word sketch*, diperlihatkan kata apa saja yang berkolokasi dengan kata yang dianalisis berdasarkan kelas katanya. Namun ternyata, banyak kelas kata yang salah. Banyak verba yang masuk dalam kolom nomina, seperti yang dapat dilihat pada tabel di Gambar 4 berikut.

<div> <div>mau</div> <div>IndonesianWaC freq = 49,462 (452.61 per million)</div> </div>									
nextright		nextleft		verb_right		adv_left		noun_right	
	49,423 1.00		47,755 0.97		33,110 0.67		25,822 0.52		22,233 0.45
belajar +	480 6.98	kalah +	1,782 8.02	tanya +	431 7.08	tak +	2,821 6.15	dikata	96 6.88
mau belajar		kalah mau		saya mau tanya		tak mau		apa mau dikata	
jujur +	240 6.95	nggak +	419 7.61	ketinggalan +	207 6.90	tidak +	11,739 5.75	berkorban	78 6.22
mau jujur ,		nggak mau		tidak mau ketinggalan		tidak mau		mau berkorban	
kemana +	190 6.84	kau +	425 7.10	menerima +	922 6.41	jangan +	259 4.55	memaafkan	75 6.11
mau kemana		kau mau		tidak mau menerima		jangan mau		mau memaafkan	
tanya +	413 6.81	kalo +	279 7.09	beli +	236 6.34	nanti +	103 4.86	silahkan	93 5.78
saya mau tanya		kalo mau		mau beli		saja +	745 4.29	mengerti +	129 5.17
ketinggalan +	198 6.47	gak +	247 6.94	tunduk +	162 6.28	pernah +	491 4.28	tidak mau mengerti	
tidak mau ketinggalan		gak mau		tidak mau tunduk		tidak pernah mau		berteman	30 5.11
kalah +	318 6.27	ga +	157 6.43	mendengarkan +	199 6.17	sekarang +	371 4.24	mau berteman dengan	
tidak mau kalah		ga mau		tidak mau mendengarkan		masih +	607 4.00	mengorbankan	47 5.04
tau +	136 6.24	aku +	738 6.41	diajak +	129 6.08	masih mau		mau mengorbankan	
mau tau		aku mau		mau diajak		belum +	331 3.95	berperang	56 5.04
menerima +	841 6.21	kamu +	373 6.28	bikin +	155 5.99	belum mau		tidak mau berperang di jalan Allah	
tidak mau menerima		kamu mau		mau bikin		tidakkah	14 3.79	mengampuni	30 4.88
nanya +	111 6.17	saya +	1,544 6.08	ikut +	468 5.97	. Tidakkah mereka mau		mau mengampuni	
mau nanya		saya mau		mau ikut		dulu +	112 3.73	maaf	69 4.63
berbagi +	121 6.08	cuma +	143 5.96	ambil +	165 5.91	seolah-olah	32 3.65	mau minta maaf	
mau berbagi		cuma mau		mau ambil		seolah-olah mau		berusaha +	179 4.59
menikah +	137 6.06	tak +	2,502 5.95	cari +	139 5.84	sengaja	41 3.64	mau berusaha	
mau menikah dengan		tak mau		mau cari		sengaja tidak mau		namanya	92 4.54
bilang +	133 6.04	dia +	634 5.77	menyerah +	118 5.75	tadi	73 3.54	mau disebutkan namanya	
mau bilang apa		dia mau		tidak mau menyerah		barangkali	25 3.54	merubah	45 4.52
dikata	99 5.91	kita +	1,332 5.75	pulang +	247 5.70	barangkali mau		mau merubah	
apa mau dikata ,		kita mau		mau pulang		sekali +	219 3.51	bersikap	63 4.50
beli +	192 5.81	gimana	86 5.74	bicara +	199 5.67	sama sekali tidak mau		mau bersikap	
mau beli		, gimana mau		mau bicara		tanpa +	259 3.50	zakat	48 4.48
mengakui +	251 5.79	ngga	84 5.72	pakai +	134 5.58	tanpa mau		tidak mau membayar zakat	
tidak mau mengakui		ngga mau		mau pakai		disini	35 3.45	dioperasi	17 4.46
tunduk +	142 5.78	tidak +	10,473 5.59	bertobat	66 5.56	lagi +	427 3.44	mau dioperasi .	

Gambar 4. Tabel *word sketch* kata mau

Sehubungan dengan itu, data *word sketch* yang akan dianalisis hanya data kolokasi *nextright* dan *nextleft*.

Data kolokasi *word sketch* menunjukkan bahwa *mau* lebih banyak berkolokasi dengan kata cakapan. Kata cakapan yang berada di sebelah kiri yaitu *nggak* (419 kali), *kalo* (279 kali), *gak* (247 kali), *ga* (157 kali), *gimana* (86 kali), *ngga* (84 kali), *enggak* (65 kali), dan *gue* (69 kali), di sebelah kanan di antaranya *bikin* (132 kali), *bilang* (133 kali), dan *nanya* (111 kali). Sementara itu, pada *word sketch* kata ingin terlihat ada pronomina formal yang berkolokasi di sebelah kiri kata tersebut seperti *anda* (2.374 kali), *engkau* (163 kali), dan *beliau* (145 kali). Perbedaan formal-nonformal ini juga terlihat pada kemunculan kata *tau* dan *tahu*. *Tau* hanya ditemukan berkolokasi dengan *mau* sebanyak 136 kali, sedangkan kata *tahu* hanya ditemukan berkolokasi dengan *ingin* sebanyak 1.993 kali.

Kolokasi kata *akan* memperlihatkan pola yang berbeda. Data menunjukkan bahwa pada kolokasi sebelah kiri kata tersebut terdapat kata *tidak* dengan jumlah kemunculan 34.036 kali dan *tak* sebanyak 4.204 kali. Hal ini mengonfirmasi analisis sebelumnya dari korpus Leipzig. Selain kata-kata negasi, *akan* juga berkolokasi dengan kata-kata yang menunjukkan tingkatan kepastian atau kemungkinan seperti *pasti* (4.675 kali), *tentu* (2.121 kali), dan *mungkin* (3.172 kali) di sebelah kiri. Di sebelah kanan, kata-kata

yang menunjukkan kepastian itu di antaranya *terus* (5.067 kali), *segera* (2.970 kali), *selalu* (2.900 kali), dan *tetap* (2.978 kali).

Data *hendak* justru memperlihatkan bahwa dalam korpus IndonesianWaC terdapat teks berbahasa Melayu. Dalam data kolokasi kata *hendak* terdapat kata *kerana*, *bercadang*, dan *cuba* yang bukan merupakan kosakata bahasa Indonesia. Selain dari itu, kolokasi kata *hendak* menunjukkan banyak pengandaian yang ditunjukkan oleh kata *seakan* (28 kali), *seakan-akan* (24 kali), *seolah* (19 kali), *seolah-olah* (50 kali), *apabila* (62 kali), dan *bila* (68 kali).

Mengingat masih banyak kesalahan pelabelan kelas kata pada korpus IndonesianWaC, data *sketch diff* yang sedianya juga akan dianalisis dengan terpaksa diabaikan karena dengan data yang kurang valid dikhawatirkan hasil analisisnya menjadi salah.

IV. SIMPULAN

Dari analisis data yang berasal dari dua korpus di atas dapat diambil kesimpulan sebagai berikut.

- *mau* lebih banyak digunakan dalam konteks nonformal, sementara ingin digunakan dalam konteks yang lebih formal.
- *akan* banyak berkolokasi dengan penyangkalan, kemungkinan, dan kepastian.
- *hendak* berkolokasi dengan keterangan waktu, pengandaian, dan banyak digunakan dalam bahasa Melayu.

Kajian ini tidak bermaksud menawarkan definisi baru dari keempat kata di atas. Namun, keterangan-keterangan yang dihasilkan dari analisis yang telah dilakukan diharapkan dapat membantu para penyusun kamus untuk memutakhirkan definisinya.

Untuk bisa mendapatkan hasil yang maksimal diperlukan korpus yang lebih komprehensif, memuat semua jenis, ranah dan genre teks, dan didukung oleh perangkat lunak yang mendukung. Sketch Engine sudah sangat mendukung penyusun kamus untuk dapat membedakan kata-kata yang bermiripan makna, hanya saja untuk korpus IndonesianWaC *tagging*-nya belum sempurna.

Masalah penggunaan kata memang masuk dalam ranah pragmatik, tetapi tidak ada salahnya jika kamus juga memuat keterangan mengenai hal tersebut dalam definisinya daripada hanya memberikan definisi sinonim. Ini dapat membantu penutur terutama pemelajar bahasa.

Daftar Pustaka

- Gries, Stefan Th. (2009). What is Corpus Linguistics? *Language and Linguistics Compass* Vol. 3 No. 5, pages 1225–1241, September 2009. The Author Journal Compilation @ 2009 Blackwell Publishing Ltd
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press
- Kilgarriff, A., Rychly, P., Smrz, P., and Tugwell, D. (2004). The Sketch Engine. Information Technology Research Institute Technical Report Series University of Brighton.
- McEnery, T., Hardie, A., (2012). *Corpus Linguistics: Method, Theory, and Practice*. Cambridge: CUP
- Sinclair, J.M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Subroto, Edi. (2011). *Pengantar Studi Semantik dan Pragmatik*. Surakarta: Cakrawala Media
- Sugono, D. (Penyunting) (2007). *Buku Praktis Bahasa Indonesia Jilid 2*. Jakarta: Departemen Pendidikan Nasional
- Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10 (1), 1-13.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Tim Redaksi KBBI. (2008). *Kamus Besar Bahasa Indonesia Edisi Keempat*. Jakarta: Gramedia Pustaka Utama.
- http://corpora.uni-leipzig.de/en?corpusId=ind_mixed_2013 diakses tanggal 9 Juli 2016
- <https://the.sketchengine.co.uk/auth/corpora/> diakses tanggal 9 Juli 2016

Menuju Pemanfaatan Leksikografis Korpus Bahasa Daerah: Keanekaragaman Tantangan dan Peluang

Yanti, Ph.D.

Nany Setyono Kurnia, Ph. D.

Universitas Katolik Indonesia Atma Jaya

reginayanti@gmail.com; regina.yanti@atmajaya.ac.id

nanykurnia2003@yahoo.com; nany.kurnia@atmajaya.ac.id

Abstrak

Sejarah panjang kegiatan penyusunan kamus sejumlah bahasa yang mendunia seperti bahasa Inggris menunjukkan bagaimana pemanfaatan data pemakaian bahasa di dunia nyata (bahasa alami) membuka celah peningkatan perbaikan mutu kamus. Penggunaan bahasa secara alami yang terhimpun dalam korpus bahasa, perkembangan pelbagai perangkat lunak yang mampu makin tajam menangguk data tertentu dari suatu korpus, dan peningkatan mutu korpus bahasa itu sendiri – dipadukan dengan intuisi penutur (jati) pekamus yang memanfaatkan semua ini – telah memperkaya kamus yang dihasilkan dalam rangka menjawab kebutuhan pengguna kamus yang disasar. Salah satu contoh, dalam hal kamus untuk pemelajar asing, adalah kemajuan dalam hal informasi frekuensi, fraseologi, dan penggunaan korpus pemelajar. Selain itu, tersedianya korpus bahasa berukuran besar dengan berbagai rancang bangun juga mendorong menjamurnya penelitian korpus yang hasilnya juga dapat mengilhami penyusunan kamus.

Pola penyusunan dan pemanfaatan korpus bahasa seperti di atas tentu dapat pula diterapkan pada Bahasa Indonesia dan bahasa daerah yang penutur (jati) dan bahan-bahan ragam tulisnya relatif tidak sukar ditemukan. Akan tetapi, sebagaimana kita ketahui, bahasa-bahasa daerah di Indonesia mempunyai rentang keanekaragaman yang sangat lebar, mulai dari yang masih hidup dengan subur sampai yang terancam punah. Makalah ini, menapak dari pengamatan dan pengalaman menyangkut upaya dokumentasi sejumlah bahasa daerah di Indonesia, selanjutnya membahas berbagai tantangan yang dihadapi dan peluang yang dimungkinkan oleh kemajuan teknologi untuk membuka jalan yang lebih lapang ke arah tersusunnya korpus bahasa-bahasa

daerah. Penyusunan korpus bahasa daerah tanpa tradisi tulis dan yang belum dikenal hal ihwal tata bahasanya, misalnya, berbeda tantangannya dengan penyusunan korpus bahasa daerah yang banyak karya tulisnya dan yang sistem ejaan serta tata bahasanya sudah banyak dipelajari dan dikenal. Tantangan-tantangan ini mengisyaratkan perlunya pengerahan daya dan kerja keras untuk penyusunan korpus bahasa yang tertata baik. Makalah ini mengusulkan agar upaya mewujudkan tersedianya korpus bahasa daerah – yang nantinya dapat digunakan bersama – dilakukan dengan pendekatan kegotongroyongan, yaitu dengan melibatkan pihak-pihak yang sudah melakukan dokumentasi bahasa daerah di Indonesia dan masyarakat luas. Kemajuan teknologi sangat memungkinkan penggalangan data bahasa yang memenuhi ciri-ciri atau persyaratan yang ditetapkan dari khalayak ramai yang awam sekalipun. Hasilnya ditampung untuk diolah lebih lanjut dan berangsur-angsur mewujudkan menjadi korpus bahasa daerah yang dapat terus ditingkatkan mutunya.

I. PENDAHULUAN

Penutur suatu bahasa seringkali mempunyai keterikatan yang kuat dengan bahasanya, seperti terlihat dalam kegaduhan di Perancis gara-gara perubahan ejaan yang mulai diimplementasikan tahun 2016 ini, sekalipun telah 26 tahun berselang sejak pencanangan reformasi ejaan itu. Sikap berapi-api menentang perubahan itu antara lain dicetuskan dalam bentuk kampanye *#Je suis circonflexe*—meniru unjuk solidaritas *Je suis Charlie* setelah penyerangan kantor majalah “Charlie Hebdo” di Paris awal 2015 silam—untuk bersama-sama menolak penghapusan tanda aksent sirkonfleks yang menyerupai topi di atas huruf hidup itu (Johnson 2016; Lichfield 2016). Kenyataan bahwa seseorang dapat sangat mengebu-gebu dalam hal-hal yang berkaitan dengan bahasa dan bahwa pekerjaan besar dapat diselesaikan secara gotong royong melalui penggalangan partisipasi khalayak ramai/sukarelawan (urun daya/*crowdsourcing*) seperti misalnya gerakan yang dikenal dengan nama Kawal Pemilu pada pemilihan Presiden (Pilpres) Republik Indonesia 2014 yang lalu (<https://id.wikipedia.org/wiki/KawalPemilu.org>) mengilhami gagasan utama yang diajukan dalam makalah ini.

Sungguh tepat Seminar *Leksikografis Indonesia* yang diselenggarakan *Badan Bahasa* kali ini mengambil tema “Tantangan Leksikografis Bahasa-Bahasa daerah di Indonesia”, karena menurut catatan www.ethnologue.com terdapat 713 bahasa daerah di Indonesia, 701 di antaranya masih hidup dengan berbagai tingkat kebelumpunahan. Suatu tantangan besar bila untuk setiap bahasa daerah diharapkan akan tersusun sebuah kamus yang baik dan kamus yang sudah ada ditingkatkan mutunya.

Bagian kedua makalah ini akan membahas kontribusi korpus bahasa dalam penyusunan kamus. Dengan mengambil contoh dari Bahasa Inggris (selanjutnya disebut BING) dan bahasa Indonesia (selanjutnya disebut BI) sejumlah manfaat yang dapat diperoleh penggunaan korpus bahasa dalam penyusunan kamus diketengahkan.

Penyusunan suatu korpus bahasa untuk keperluan penyusunan kamus melibatkan penentuan rancang bangun korpus yang berimbang dan pengumpulan data berupa sampel pemakaian bahasa sesuai dengan rancang bangun yang sudah ditentukan. Bertolak dari pengalaman lapangan dalam pengumpulan data untuk menyusun basis data bahasa daerah di Indonesia termasuk dinamika pelatihan pengumpul data, bagian ketiga makalah ini membahas keanekaragaman tantangan yang dihadapi.

Makalah diakhiri dengan bagian keempat yang membahas peluang-peluang yang ada serta kemungkinan-kemungkinan pemanfaatannya.

II. KAMUS DAN KORPUS BAHASA

Penggunaan data dari korpus bahasa untuk penyusunan kamus kini sudah menjadi kelaziman, sebagaimana dikatakan oleh Sue Atkins dan Michael Rundell (2008: 3): “In the twenty-first century, all good dictionaries take corpus data as their starting point,...”. Tulisan Adi Budiwiyanto di laman Badan Bahasa (<http://badanbahasa.kemdikbud.go.id/lamanbahasa/artikel/1580>) mengulas peran korpus dalam penyusunan kamus secara ringkas bernas. Intinya, dalam penyusunan kamus pekamus tidak hanya menggunakan introspeksi yang bertumpu pada intuisinya selaku penutur bahasa yang bersangkutan tetapi juga melakukan observasi bagaimana bahasa tersebut digunakan oleh penuturnya di dunia nyata. Penggunaan bahasa secara alami ini dikumpulkan sampelnya secara sistematis menjadi suatu korpus bahasa untuk ditelaah oleh pekamus agar dapat menggambarkan kosakata bahasa yang bersangkutan dengan tepat. Sejumlah hal yang tercakup dalam tulisan tersebut akan diulas di sini untuk menekankan pentingnya penggunaan korpus bahasa dalam penyusunan kamus: frekuensi dan senarai/daftar kata, ungkapan multikata, serta korpus pemelajar.

Contoh-contoh yang digunakan dalam pembahasan hal di atas diambil dari kamus BING dan BI serta sebuah korpus BI. Kamus BING yang dikutip adalah kamus ekabahasa untuk pemelajar tingkat lanjut dari Macmillan (Rundell 2007), sementara kamus BI yang dikutip adalah *Kamus Besar Bahasa Indonesia Pusat Bahasa* edisi ke 4 (Pusat Bahasa 2008), yang selanjutnya akan disebut KBBI. Data korpus yang digunakan berasal dari korpus BI susunan Hananto, Markus, dan Kurnia (2011) sebesar sekitar 10 juta kata meliputi teks-teks di bawah ini:

- buku sekolah (pendidikan dasar dan menengah) daring yang disediakan oleh Kementerian Pendidikan Nasional
- cerita untuk anak-anak
- cerita untuk orang dewasa

- bacaan non-fiksi untuk orang dewasa dari empat surat kabar daring, berita dan bukan-berita.

Karena cerita untuk anak-anak ukurannya kecil sekali (total hanya sekitar 120 ribuan kata), korpus sekitar 10 juta ini dipilah menjadi tiga subkorpus:

1. Buku sekolah dan cerita untuk anak-anak (sekitar 4,75 juta kata)
2. Bacaan dewasa fiksi (sekitar 3 juta kata)
3. Bacaan dewasa bukan-fiksi (sekitar 2,25 juta kata).

2.1 Frekuensi dan Senarai / Daftar Kata

Frekuensi atau kekerapan digunakannya suatu kata atau ungkapan hanya dapat diperoleh dari suatu korpus, sebagaimana dikatakan oleh Geoffrey Leech (2011: 7):

If asked what is the one benefit that corpora can provide and that cannot be provided by other means, I would reply ‘information about frequency’.

Informasi tunggal tentang frekuensi penggunaan suatu kata atau ungkapan dalam suatu korpus, misalnya kata X muncul Y kali dalam korpus Z, tidaklah banyak gunanya, tetapi dapat menjadi sangat berguna jika digandengkan dengan informasi frekuensi kata atau ungkapan lain dalam korpus yang sama atau informasi frekuensi kata yang sama di korpus lain. Berdasarkan data frekuensi ini telah disusun berbagai senarai atau daftar kata (Coxhead 2000; Gardner dan Davies 2014; Leech, Rayson, dan Wilson 2001; Paquot 2010; West 1953; Xue dan Nation 1984). Kata yang kerap dipakai dalam suatu bahasa penting dikuasai oleh pemelajar bahasa tersebut karena akan kerap ditemuinya. Kamus BING ekabahasa untuk pemelajar tingkat lanjut telah lama menggunakan apa yang disebut ‘controlled defining vocabulary’ (Cowie 1999: 110), artinya semua definisi atau batasan makna yang ada dalam kamus itu sedapat mungkin menggunakan sejumlah (misalnya dua ribu) kata berfrekuensi tertinggi dalam BING berdasarkan korpus yang disusun oleh penerbit kamus. Dengan cara ini diharapkan pengguna kamus yang mencari makna suatu kata tidak menemui batasan makna yang mengandung kata-kata berfrekuensi rendah (kata-kata yang jarang dipakai) yang tidak dikenalnya sehingga ia harus mencari pula makna kata-kata tersebut. Berdasarkan senarai kata yang dikembangkan oleh penerbit yang bersangkutan, kamus BING ekabahasa untuk pemelajar tingkat lanjut (terbitan Longman, Oxford, Macmillan, dan lain-lain) biasanya juga memberi tanda tertentu pada lema yang merupakan kata yang kerap kali dipakai (baik pada umumnya atau dalam ragam tertentu, seperti misalnya bahasa akademik) agar pemakai kamus memberikan perhatian khusus pada lema tersebut.

2.2 Ungkapan Multikata

Kekayaan kosakata suatu bahasa tidak hanya berbentuk kata-kata tunggal tetapi juga ungkapan-ungkapan multikata dalam berbagai bentuk, seperti idiom, kolokasi, ungkapan-ungkapan rutin untuk hal-hal tertentu dan lain-lain (Granger dan Meunier

2008; Moon 1998; Nattinger dan DeCarrico 1992; Pawley dan Syder 1983). Penguasaan ungkapan multikata sangat membantu seorang pemelajar untuk berbahasa secara natural dan meningkatkan kefasihan bertuturnya. Berdasarkan suatu korpus atau sejumlah korpus tertentu yang digabung atau dikontraskan telah disusun berbagai senarai ungkapan multikata (Biber dan Barbieri 2007; Biber, Conrad, dan Cortes 2004; Hyland 2008; Martinez dan Schmitt 2012; Shin dan Nation 2008; Simpson-Vlach dan Ellis 2010). Data korpus dapat membantu pekamus melihat pemakaian ungkapan multikata dalam berbagai ragam bahasa, misalnya penggunaan gugus kata atau gabungan kata BARANG TENTU yang dapat kita temukan di bawah lema BARANG yang kedua di KBBI (yakni adverba BARANG) seperti tersalin pada Gambar 1 di bawah ini. Sesuai patokan yang digunakan KBBI (hal. XXXV) gabungan kata ini diletakkan di bawah lema atau sublema yang menjadi unsur pertama yaitu BARANG, dan diberi batasan makna ‘tentu saja’:

²ba.rang *adv* 1 sedikit banyak; sekadar; kira-kira: *beri aku nasi -- sesuap*; 2 *kl* mudah-mudahan: -- *disampaikan Allah kiranya surat ini kpd anakku*; 3 kata yg digunakan untuk menambah ketidakpastian; -- **apa** apa saja; segala sesuatu; -- **bila** suatu waktu; bilamana saja; sewaktu-waktu; -- **sesuatu** segala sesuatu; -- **siapa** siapa saja; -- **tentu** tentu saja

Gambar 1. Gugus kata BARANG TENTU di bawah lema BARANG (adverba) di KBBI (Pusat Bahasa 2008: 140)

Dengan bantuan perangkat lunak ANTCONC (Anthony 2014) kita akan mendapati 17 gugus kata BARANG TENTU dalam korpus BI sebesar sekitar sepuluh juta kata susunan Hananto, Markus, dan Kurnia (2011) yang telah diuraikan di atas. Gambar 2 di bawah ini menunjukkan ketujuhbelas larik konkordans BARANG TENTU dalam korpus tersebut, masing-masing dengan subkorpus asalnya.

Dengan bantuan fitur pengurutan pada ANTCONC (Anthony 2014) ketujuhbelas larik pemakaian BARANG TENTU ini dapat disusun berdasarkan kata pertama di sebelah kiri gugus kata BARANG TENTU tersebut yang diurut secara alfabetis (Gambar 3). Dengan pengurutan seperti pada Gambar 3 di bawah ini jelas terlihat bahwa 15 dari 17 kali pemunculan BARANG TENTU dalam korpus tersebut didahului dengan kata SUDAH, sementara pada dua pemakaian BARANG TENTU lainnya BARANG tidak berkaitan (tidak dalam satu frasa) dengan TENTU sebagaimana terlihat dari adanya koma setelah kata BARANG pada larik yang kedua, sementara tanpa koma pun, larik pertama *Setiap manusia menggunakan barang tentu ada tujuannya* jelas menunjukkan bahwa BARANG tidak terkait dengan TENTU. Perlu ditegaskan di sini bahwa sekalipun cukup memadai untuk tujuan utama pembuatannya, korpus sekitar 10 juta kata yang dipakai

sebagai sumber data di sini tergolong kecil ukurannya dan jauh dari sempurna rancang bangunnya untuk merepresentasikan BI. Namun demikian, hasil 17 larik konkordans ini, dikombinasikan dengan intuisi penutur jati BI, cukup kuat mengindikasikan perlunya kita melakukan observasi dengan data bahasa alami yang cakupannya lebih luas dan berimbang untuk menentukan apakah gugus kata yang baku adalah BARANG TENTU atau SUDAH BARANG TENTU.

sangat mengesankan. Sudah barang tentu kamu tidak akan pernah	BukuSekolah&FiksiAnak.
ajanya memeluk Islam, sudah barang tentu rakyatnya dengan mudah	BukuSekolah&FiksiAnak.
usaha dalam menghasilkan barang, tentu harus mempertimbangk	BukuSekolah&FiksiAnak.
tiap manusia menggunakan barang tentu saja ada tujuan- nya,	BukuSekolah&FiksiAnak.
ghargaan yang kurang, sudah barang tentu akan menimbulkan rasa t	BukuSekolah&FiksiAnak.
arah tujuan yang sama. Sudah barang tentu bukan pelaku dalam arti	BukuSekolah&FiksiAnak.
bangsa dan bernegara, sudah barang tentu perlu memiliki dasar neg	BukuSekolah&FiksiAnak.
sukan dari masyarakat, sudah barang tentu diperlukan keterbukaan	BukuSekolah&FiksiAnak.
terapkan di Indonesia. Sudah barang tentu tidak semua kemajuan y	BukuSekolah&FiksiAnak.
diri dari globalisasi ini. Sudah barang tentu globalisasi ini akan berda	BukuSekolah&FiksiAnak.
ipun jenisnya. Namun, sudah barang tentu tidak setiap perubahan ya	BukuSekolah&FiksiAnak.
nerangan. Kegiatan ini sudah barang tentu membuat hati Amran dar	BukuSekolah&FiksiAnak.
h membasahi bajunya. Sudah barang tentu kegaduhan ini membuat	BukuSekolah&FiksiAnak.
in dari seberang sana. Sudah barang tentu kondektur itu tak dapat	Dewasa_Fiksi.txt
ensial,. Oleh sebab itu sudah barang tentu untuk kemajuannya dibu	Dewasa_Non_Fiksi.txt
na reshuffle. Sehingga sudah barang tentu masalah pencopotan Fad	Dewasa_Non_Fiksi.txt
1 pemikiran manusia). Sudah barang tentu dengan klasifikasi itu tak	Dewasa_Non_Fiksi.txt

Gambar 2. Larik konkordans BARANG TENTU

1	g? Setiap manusia menggunakan barang tentu saja ada tujuan- nya,	BukuSekoli
2	i pengusaha dalam menghasilkan barang, tentu harus mempertimbangk	BukuSekoli
3	penghargaan yang kurang, sudah barang tentu akan menimbulkan rasa t	BukuSekoli
4	arah tujuan yang sama. Sudah barang tentu bukan pelaku dalam arti	BukuSekoli
5	dan pemikiran manusia). Sudah barang tentu dengan klasifikasi itu tak	Dewasa_N
6	masukn dari masyarakat, sudah barang tentu diperlukan keterbukaan c	BukuSekoli
7	diri dari globalisasi ini. Sudah barang tentu globalisasi ini akan berda	BukuSekoli
8	ang sangat mengesankan. Sudah barang tentu kamu tidak akan pernah	BukuSekoli
9	ecah membasahi bajunya. Sudah barang tentu kegaduhan ini membuat I	BukuSekoli
10	ingan dari seberang sana. Sudah barang tentu kondektur itu tak dapat	Dewasa_Fil
11	vacana reshuffle. Sehingga sudah barang tentu masalah pencopotan Fade	Dewasa_N
12	u penerangan. Kegiatan ini sudah barang tentu membuat hati Amran dan	BukuSekoli
13	berbangsa dan bernegara, sudah barang tentu perlu memiliki dasar nega	BukuSekoli
14	ja-rajanya memeluk Islam, sudah barang tentu rakyatnya dengan mudah	BukuSekoli
15	k diterapkan di Indonesia. Sudah barang tentu tidak semua kemajuan ya	BukuSekoli
16	maupun jenisnya. Namun, sudah barang tentu tidak setiap perubahan ya	BukuSekoli
17	potensial,. Oleh sebab itu sudah barang tentu untuk kemajuannya dibu	Dewasa_N

Gambar 3. Larik konkordans gugus kata BARANG TENTU

(disusun berdasarkan urutan alfabetis kata pertama di sebelah kirinya)

Contoh lain adalah larik konkordans untuk gugus kata ASAL TAHU SAJA (Gambar 4):

Jerman lebih unggul. Asal tahu saja, kecepatan Maglev menc	BukuSekolah&FiksiAnak
tahu berterima kasih. Asal tahu saja, kau hidup kembali	BukuSekolah&FiksiAnak
kalian di ruangan ini. Asal tahu saja, malapetaka tak akan	Dewasa_Fiksi.txt
gggal di sini.\x94 \x93Asal tahu saja,\x94 ujar Ross. \	Dewasa_Fiksi.txt
sangar dan berkata, "Asal tahu saja, Weinhard, kau tidak	Dewasa_Fiksi.txt
n," balas si Perlente. "Asal tahu saja, Lohmann, aku mengaw	Dewasa_Fiksi.txt
ra 2,\x94 ungkapinya. Asal tahu saja, kenaikan saham HRUM	Dewasa_Non_Fiksi.txt
terakhir," tandasnya. Asal tahu saja, kurs rupiah di	Dewasa_Non_Fiksi.txt
tur tubuh yang buruk. Asal tahu saja, punggung sangat sensiti	Dewasa_Non_Fiksi.txt
i dari berbagai pihak. Asal tahu saja, 250 nasabah Diamond I	Dewasa_Non_Fiksi.txt
ai investasi alternatif. Asal tahu saja, pagi tadi, kontrak	Dewasa_Non_Fiksi.txt

Gambar 4. Sembilan larik konkordans ASAL TAHU SAJA dan subkorpus asalnya

¹**asal 1** n keadaan (tempat, wujud, rupa, dsb) yg semula; pangkal permulaan: *patung-patung itu akan dikembalikan ke --nya*; **2** a mula-mula sekali; semula: *batas-batasnya yg -- sudah tidak diketahui lagi*; -- *ayam pulang ke lumbung*; -- *itik pulang ke pelimbanan*, *pb* tabiat orang tidak akan berubah; *usul menunjukkan --*, *pb* kelakuan (budi bahasa) seseorang menunjukkan asal keturunannya;
-- **jadi** asal-asalan: *memang cara kerja developer belakangan ini agak sembarangan, RS dan RSS (rumah sangat sederhana) dibuat -- jadi*; -- **mula** mula-mula sekali; keadaan (sebab-sebab dsb) yg semula: -- *mula pertikaian itu krn salah paham saja*; --**usul 1** asal keturunan; silsilah; susur galur: *kalau dilihat dr --nya, dia masih keturunan Raja Mataram*; **2** cerita (secara urut dr awal sampai terjadinya suatu peristiwa); riwayat; **3** yg menjadi sebab-sebabnya (tt suatu peristiwa atau kejadian); sebab mulanya;
meng.a.sal.kan v memulangkan (mengembalikan) kpd keadaan semula;
ber.a.sal v **1** bermula: *peperangan itu ~ dr sengketa perbatasan*; **2** bersumber: *berita itu dapat dipercayai krn ~ dr pihak resmi*; **3** bertempat asal (dibuat, dilahirkan, dsb): *barang-barang yg ~ dr Jepang terdapat di mana-mana*; *dia ~ dr Jawa Timur*; **4** keturunan: *dia ~ dr orang baik-baik*
²**asal p 1** dng syarat; apabila: *engkau boleh pergi -- pekerjaanmu beres*; **2** pokoknya; yg penting: *biar lambat -- selamat*;
asal.kan p dng syarat
³**asal a** cak sembarangan; seenaknya saja: *kalau bekerja jangan -- saja, ikutilah petunjuk yg ada*;
asal-asal.an a sembarang; seenaknya
asal.an Bjr n barang yg belum diolah: *getah --*, karet mentah

Gambar 5. Lema ASAL di KBBI
(Pusat Bahasa 2008: 90)

Berbeda dengan larik konkordans SUDAH BARANG TENTU yang terutama muncul di sub-korpus Buku Sekolah dan Fiksi Anak, larik konkordans ASAL TAHU SAJA tampak muncul merata di ketiga sub-korpus. Selain itu, sekalipun masih diperlukan telaah lebih lanjut, larik ASAL TAHU SAJA yang terlihat di sini mengesankan berasal dari ragam lisan, indikasi yang tidak muncul pada larik konkordans SUDAH BARANG TENTU. Di lain pihak, sebagaimana terlihat pada lema ASAL di KBBI (Gambar 5), gugus kata ASAL TAHU SAJA tidak tercantum di bawah lema ini, sama halnya dengan SUDAH BARANG TENTU yang tidak tertera di bawah lema BARANG.

Tentu saja suatu kamus, terlebih-lebih kamus cetak, punya keterbatasan ruang yang dapat sangat mengungkung, sehingga pekamus harus benar-benar ketat menyeleksi ungkapan mana yang akan dimasukkan di bawah suatu lema atau sub-lema. Artinya, walau pun seandainya kedua gugus kata yang baru dibahas itu dianggap sebagai ungkapan multikata BI yang sah, belum tentu pekamus dapat memasukkannya ke dalam KBBI, misalnya karena ungkapan itu kalah bersaing dengan ungkapan lain yang frekuensi pemakaiannya lebih tinggi. Akan tetapi, bila KBBI dimaksudkan untuk mencakup semua kosakata BI, serupa dengan bagaimana, seperti digambarkan oleh Simon Winchester (2003: 41), awal mulanya di abad ke-19 The Oxford English Dictionary itu diidam-idamkan:

A dictionary of the English language in its totality...that would give, in essence and in fact, the meaning of everything.
maka penyusun KBBI perlu mempertimbangkan seluruh ungkapan BI yang ada, termasuk kedua gugus kata tadi.

2.3 Korpus Pemelajar

Tulisan Adi Budiwiyanto dalam laman *Badan Bahasa* mengenai peran korpus dalam penyusunan kamus yang telah disebut di atas juga membahas tipologi korpus. Di antaranya adalah korpus pemelajar (bahasa asing) yang dibedakan atau dikontraskan dengan korpus penutur jati. Tipologi berdasarkan status penutur bahasa ini juga mengenal korpus perkembangan dan korpus lingua franca (Cheng 2012). Korpus perkembangan adalah korpus penutur jati yang masih dalam tahap pemerolehan bahasa ibu mereka, seperti misalnya anak-anak, atau mahasiswa yang sedang dalam proses pemerolehan bahasa akademik yang matang. Adapun korpus lingua franca biasanya adalah korpus BING yang penuturnya sehari-hari dengan cukup fasih menggunakan BING sebagai lingua franca dalam komunikasi tetapi bahasa-ibu mereka bukan BING.

Salah satu contoh korpus pemelajar BING adalah ICLE (*The International Corpus of Learner English*) yang dikembangkan oleh *Centre for English Corpus Linguistics*, Universitas Katolik Louvain di Belgia (<http://www.uclouvain.be/en-cecl-icle.html>). Korpus ini berisikan esai argumentatif karya pemelajar BING dengan tingkat kemahiran menengah atas dan lanjut dari enam belas latarbelakang bahasa-ibu yang berbeda. Pusat Korpus Linguistik BING tersebut telah mengundang pihak luar untuk ikut serta mengisi korpus pemelajar BING ini dan esai yang berhasil dikumpulkan dalam korpus hasil kerjasama internasional dengan berbagai universitas rekanan ini memenuhi ketentuan ciri-ciri yang sama karena diterapkannya petunjuk pengumpulan korpus yang sama.

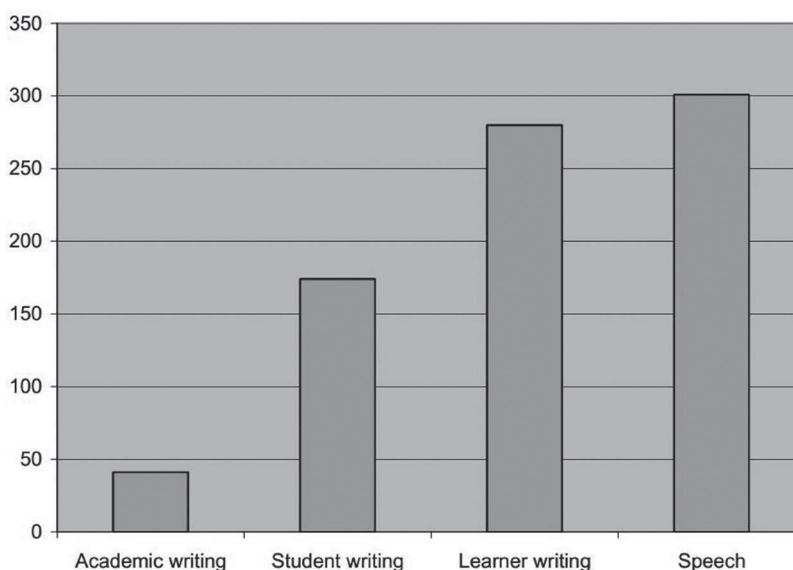
Data korpus pemelajar seperti ini dapat menunjukkan kesulitan-kesulitan pemelajar yang terjadi berulang kali dan dialami oleh pemelajar dengan berbagai latarbelakang bahasa-ibu (Rundell dan Granger 2007). Kerjasama penerbit kamus Macmillan dengan Pusat Linguistik Korpus BING di Louvain telah memanfaatkan data dari korpus pemelajar yang disusun pusat tersebut dalam upaya peningkatan isi kamus ekabahasa Macmillan

untuk pemelajar BING (Rundell 2007; Rundell dan Granger 2007). Contoh di bawah ini menggambarkan bagaimana hasil penelitian yang menunjukkan kurangnya kesadaran pemelajar bahasa akan perbedaan ragam lisan dengan ragam akademik tulis yang formal sifatnya dapat dimanfaatkan untuk memperkaya kamus dengan menambahkan catatan khusus mengenai hal ini.

Gaëtanelle Gilquin dan Magali Paquot (2007) menunjukkan perbedaan dan kemiripan yang signifikan frekuensi pemakaian kata MAYBE (Gambar 6) dalam empat korpus BING:

- 10 juta kata ragam lisan penutur jati dari BNC (*The British National Corpus*)
- 15 juta kata teks akademik penutur jati yang diambil dari BNC
- Hampir 1,5 juta kata dari korpus pemelajar ICLE dengan 16 bahasa-ibu
- Sekitar 300,000 kata LOCNESS (*Corpus of Native English Essays*), teks akademik karya mahasiswa Inggris dan Amerika penutur jati BING, sebuah korpus yang bisa digolongkan ke dalam korpus perkembangan.

Dari perbedaan frekuensi yang tampak pada Gambar 6 di bawah dapat disimpulkan bahwa karya tulis formal pemelajar asing BING maupun penutur jati BING yang masih dalam fase perkembangan lebih mendekati ciri ragam lisan dalam hal kekerapan penggunaan kata MAYBE. Hasil ini, diperluas dengan perbandingan frekuensi pemakaian PERHAPS yang merupakan padanan MAYBE tetapi lebih formal, dimanfaatkan oleh kamus Macmillan ekabahasa bagi pemelajar tingkat lanjut BING untuk memberikan catatan khusus pada lema MAYBE (Gambar 7) maupun dalam sisipan *Improve your Writing Skills* kamus tersebut (Rundell 2007: IW17).

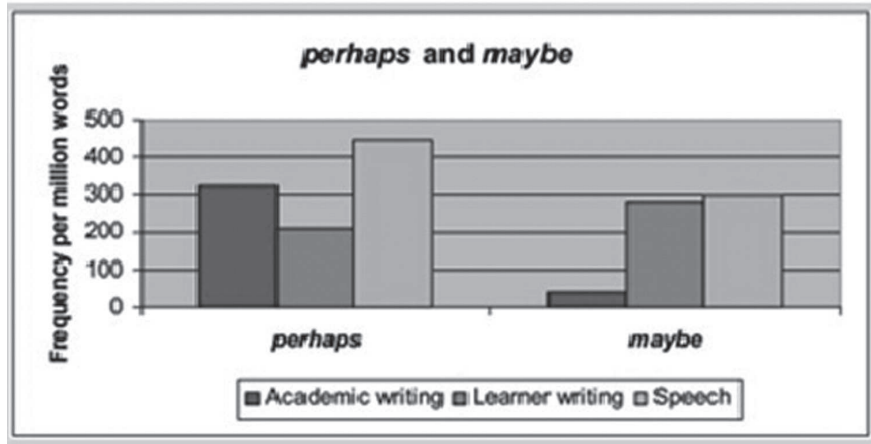


Gambar 6. Perbandingan frekuensi (per 1 juta kata) pemakaian kata MAYBE dalam 4 korpus (Gilquin dan Paquot 2007: 8)

Maybe and **perhaps** have the same meaning, but **maybe** is used mainly in spoken English and informal writing. In more formal English, **perhaps** is far more common:

Now, **maybe** I haven't explained myself very well.

There are, **perhaps**, three principles which must be followed.



Gambar 7. Catatan khusus di bawah lema MAYBE kamus Macmillan untuk pemelajar asing tingkat lanjut (Rundell 2007: 931)

III. TANTANGAN-TANTANGAN PENGEMBANGAN BASIS DATA KORPUS BAHASA DAERAH

Tantangan yang dihadapi untuk membuat korpus bahasa-bahasa daerah di Indonesia (selanjutnya, BD) sebagai basis data untuk pengolahan lebih lanjut tidaklah ringan dan sangat beraneka ragam. Selain jumlahnya yang besar, kondisi kehidupannya sebagai alat komunikasi juga berbeda-beda. Sebagian sudah ada data dan kajiannya, sedang yang lain belum tersentuh sama sekali.

3.1. Keanekaragaman Kondisi Kekinian Bahasa Daerah di Indonesia

Seperti yang telah disampaikan di bagian pendahuluan, Ethnologue mencatat sebanyak 719 bahasa di Indonesia dan hanya 707 yang masih digunakan. 701 dari bahasa yang masih digunakan ini, dikategorikan sebagai bahasa asli, yaitu bahasa daerah. Rentang keanekaragaman bahasa-bahasa daerah di Indonesia (selanjutnya, BD) sangat lebar. Seperti yang juga dilaporkan oleh Ethnologue, 76 dari bahasa tersebut berstatus mendekati kepunahan, sementara 266 bahasa lainnya dinyatakan dalam status rawan punah dan 261 masih cukup kuat, dalam arti masih berperan sebagai bahasa sehari-hari semua generasi penutur jati (<https://www.ethnologue.com/country/ID>).

Selain memiliki tingkat kebelumpunahan yang beraneka ragam, tradisi dari BD juga berbeda-beda. Ada bahasa yang sudah memiliki tradisi tulis dan sistem ejaannya

jelas, seperti bahasa Jawa dan bahasa Sunda, ada pula bahasa yang tidak ada tradisi tulisnya, apalagi sistem ejaannya. Bahasa-bahasa yang masuk dalam kategori kedua, hanya memiliki tradisi lisan dan BD yang masuk dalam kategori ini mungkin jumlahnya lebih besar.

Dari segi tersedia tidaknya literatur BD juga berbeda antara yang satu dengan yang lain. Ada bahasa yang hal ikhwal tata bahasanya sudah banyak dipelajari dan literturnya pun sudah cukup banyak, seperti bahasa Jawa dan bahasa Sunda. Sementara itu ada bahasa yang jumlah literatur dan deskripsi bahasanya masih terbatas, seperti bahasa-bahasa dalam kelompok bahasa Kerinci di Provinsi Jambi. Akhirnya, pasti ada pula bahasa yang sama sekali belum dipelajari dan tidak ada literturnya.

3.2 Produk-produk tentang Bahasa Daerah di Indonesia

3.2.1 Produk cetak

Produk-produk tentang BD, terutama dalam bentuk deskripsi tata bahasa, kamus, dan karya ilmiah, sudah cukup banyak diterbitkan, khususnya dalam versi cetak. Pada tahun 1970-1980an, Pusat Pembinaan dan Pengembangan Bahasa (sekarang Badan Bahasa) di bawah Departemen Pendidikan dan Kebudayaan, misalnya, menerbitkan sejumlah buku tentang tata bahasa dan kamus bahasa daerah di Indonesia, antara lain:

- Morfologi dan Sintaksis Bahasa Kaili (Sofyan dkk. 1979)
- Morfologi dan Sintaksis Bahasa Rejang (Napsin dkk. 1981)
- Kamus Melayu Langkat-Indonesia (Masindan dkk. 1985)
- Kosa Kata Bahasa Melayu Riau (Sulaiman dkk. 1986)
- Struktur Bahasa Pekal (Nikelas dkk. 1986)
- Struktur Bahasa Muko-Muko (Manan dkk. 1986)
- Morfologi dan Sintaksis Bahasa Komodo (Troebos dkk. 1987)
- Struktur Bahasa Baru (Elbar dkk. 1987)

Arka (2013: 89) mencatat 1556 judul publikasi yang diterbitkan Badan Bahasa dalam kurun waktu 1975-2007 dan sepertiga di antaranya membahas BD dan bahasa yang paling banyak ditulis adalah bahasa Jawa, bahasa daerah yang jumlah penutur jatinya paling banyak¹.

Selain itu, dari tahun 1975 hingga sekitar 2004, Pusat Kajian Bahasa dan Budaya (PKBB – dahulu Lembaga Bahasa) Unika Atma Jaya melalui serial NUSA telah menerbitkan tulisan yang berisi deskripsi bahasa dan artikel ilmiah dari sejumlah BD, antara lain *Miscellaneous Studies in Indonesian and Languages in Indonesia*, bagian I hingga bagian X (Verhaar 1975, Halim 1977, Poedjosoedarmo 1977, Suharno

¹ Perlu dicatat bahwa publikasi-publikasi ini, khususnya yang diterbitkan sebelum tahun 2000 tidak mudah diakses tidak semua perpustakaan memiliki koleksinya dan juga karena tidak ada versi elektroniknya.

1977, Verhaar 1978, Halim 1979, Dardjowidjojo dan Verhaar 1983, Verhaar 1984, Dardjowidjojo 1986, 1987, Purwo 1990) serta sebuah kamus berjudul *Asilulu-Englsih Dictionary* (Collins 2003)². Produk tentang BD, baik berupa buku tata bahasa maupun kamus, juga diterbitkan oleh pihak-pihak lain, misalnya:

- *Wolio Dictionary: Wolio, English, Indonesian* (Anceaux 1987)
- *Muna-English dictionary* (Berg dan Sidu 1996)
- *Kamus Muna-Indonesia* (Berg dan Sidu 2000)
- *A grammar of the Pendau language of central Sulawesi*, Indonesia (Phil 2007)
- *A Grammar of Tukang Besi* (Donohue 2009)
- *A Grammar of Madurese* (Davies 2010)
- *Kamus Bahasa Rongga* (Arka 2012)

3.2.2 Produk-produk daring

Selain produk-produk cetak, publikasi-publikasi yang relatif lebih baru sudah banyak yang tersedia secara daring. SIL Internasional, misalnya telah memublikasikan sedikitnya 444 artikel dan buku linguistik yang terkait dengan BD (<http://www.sil.org/resources/publications/search/country/Indonesia/domain/linguistics>)³. SEALANG library (<http://sealang.net/library/>) memiliki koleksi kamus bahasa Indonesia dan beberapa BD, yaitu kamus bahasa Bali, kamus bahasa Bugis, kamus bahasa Jawa, dan kamus bahasa Melayu. Dalam perkembangannya, NUSA telah menjadi jurnal berkala sejak 2013 dan masih menerbitkan karya-karya ilmiah mengenai bahasa-bahasa di Indonesia dan sekitarnya yang versi elektroniknya dapat diakses secara gratis melalui <http://www.aa.tufts.ac.jp/en/publications/nusa>⁴. Austronesian basic vocabulary database yang dapat diakses di <http://language.psy.auckland.ac.nz/austronesian/> merupakan pangkalan data yang memuat kosa kata dari sekitar 1.369 bahasa yang digunakan wilayah Pasifik, termasuk cukup banyak BD.

3.3. Basis Data Korpus Bahasa Daerah

Walaupun sejumlah produk telah dipublikasikan baik cetak maupun elektronik, data dalam publikasi-publikasi itu merupakan data yang sudah diseleksi dari korpus yang dimiliki oleh peneliti. Penyeleksian data ini biasanya erat kaitannya dengan

² Versi elektronik terbitan NUSA periode 1975-2004 ini kini disimpan dan dapat diakses di <http://sealang.net/nusa/>.

³ SIL International jugalah yang mengembangkan www.ethnologue.com, sebuah situs yang menyediakan semacam katalog yang menyeluruh mengenai bahasa-bahasa yang masih digunakan di dunia, termasuk di Indonesia.

⁴ Jurnal NUSA merupakan salah satu wujud sama PKBB Unika Atma Jaya dan Institute for Languages and Cultures of Asia and Afrika (ILCAA), Tokyo University of Foreign Studies. Versi cetak dapat diperoleh di PKBB Unika Atma Jaya.

tujuan penulisan dan terbatasnya ruang yang tidak memungkinkan para penulis untuk memasukkan semua data yang ada. Selain itu, data yang dipublikasikan pun acapkali sudah mengalami penyuntingan, seperti yang disampaikan oleh Budiwiyanto (Badan Bahasa). Sementara itu, banyak bahan mentah berupa korpus dari publikasi-publikasi tersebut tidak didapat diakses untuk dilakukan penelitian lanjutan.

Ada sebagian kecil BD yang data mentahnya dipublikasikan secara daring. SEALANG library (<http://sealang.net/library/>), misalnya, selain memiliki koleksi kamus juga memiliki korpus dari bahasa Indonesia, bahasa Melayu, bahasa Bali, bahasa Bugis, dan bahasa Jawa⁵. Ada juga The DOBES Archive (<http://dobes.mpi.nl/#>), sebuah portal yang menampung basis data dari berbagai bahasa di dunia, termasuk beberapa BD, seperti bahasa Totoli di Sulawesi dan empat bahasa di Kepulauan Aru. Korpus dari The DOBES Archive ini sudah diintegrasikan di dalam *The Language Archive (TLA)* (<https://corpus1.mpi.nl/ds/asv/?19>), tempat penyimpanan basis data bahasa-bahasa di dunia yang diprakarsai oleh the Max Planck Institute for Psycholinguistics di Nijmegen. Basis data korpus yang dipublikasikan di TLA ada yang aksesnya terbuka untuk umum, ada juga yang aksesnya harus dimintakan kepada penyumbang data, serta ada yang aksesnya tertutup.

Basis data korpus dari BD yang sudah ada tentu masih dapat ditingkatkan dalam hal ukuran dan cakupan datanya. Semakin besar ukuran dan cakupan data sebuah basis data korpus, maka pemahaman akan bahasa tersebut menjadi lebih baik dan produk yang dapat dihasilkan berdasarkan korpus tersebut juga diharapkan baik. Sementara itu, untuk BD yang belum ada basis data korpusnya, produk-produk yang sudah dipublikasikan dapat memberikan gambaran untuk rancang bangun sebuah basis data korpus bahasa yang terkait. Akhirnya, untuk BD yang belum pernah dipelajari tentu saja upaya pengembangan basis data korpusnya menjadi jauh lebih sulit, meskipun bukan berarti tidak dimungkinkan.

3.4 Kesulitan-kesulitan Pengumpulan Sampel Bahasa Daerah di Indonesia

Dalam prakteknya, pengumpulan sampel BD dalam rangka penyusunan basis data korpus memang bukan perkara yang mudah. Pertama, seperti yang disampaikan oleh Budiwiyanto (Badan Bahasa), daerah-daerah di mana bahasa-bahasa yang terancam punah itu sulit diakses karena letaknya jauh di pedalaman, di pegunungan, atau di pulau terpencil sehingga diperlukan waktu dan upaya yang tidak sedikit untuk ke sana dan juga bagaimana masyarakat setempat dapat menerima kedatangan peneliti.

⁵ Keempat bahasa daerah yang ada korpusnya ini merupakan bahasa yang sudah cukup banyak ditulis di dalam literatur dan yang jumlah penuturnya besar.

Selain masalah geografis di atas, ada sejumlah kesulitan yang dihadapi dalam upaya pengumpulan sampel BD untuk penyusunan korpusnya, mulai dari pengumpulan data, pengantosan data, hingga penyuntingan data. Untuk bahasa-bahasa yang belum pernah diteliti, bagaimana seorang peneliti akan memulainya? Apakah menggunakan daftar kata dan bagaimana berkomunikasi dengan penuturnya? Selain itu, seringkali bahasa alami tidak serta merta keluar dari penutur jati dengan hadirnya peneliti yang notabene “orang asing” dan mungkin tidak fasih atau bahkan tidak dapat berkomunikasi dalam bahasa setempat pada suatu kegiatan berbahasa. Penutur jati mungkin menjadi lebih formal dalam berbahasa atau mungkin banyak menggunakan istilah-istilah bahasa yang dianggap lebih bisa dipahami peneliti, misalnya bahasa Indonesia atau bahasa yang lebih besar di sekitar wilayah tersebut. Masih dalam konteks pengumpulan sampel bahasa, data dikumpulkan idealnya berasal dari berbagai praktek kebahasaan komunitas penutur bahasa yang diteliti. Namun ada praktek-praktek kebahasaan yang tidak selalu atau sering dilakukan. Hal ini mengakibatkan sulitnya mendapatkan sampel bahasa demikian.

Setelah data rekaman terkumpul, tentu ada isu-isu yang muncul ketika data tersebut akan dianotasi. Sistem ejaan yang digunakan tentu penting sekali dalam upaya mendapatkan teks data yang konsisten. Penentuan apakah dua buah bunyi merupakan dua fonem yang berbeda atau hanya variasi bunyi misalnya sangatlah penting. Penentuan batas morfem dari suatu kata dan bagaimana morfem-morfem diartikan juga tidak mudah dilakukan. Untuk bahasa yang sudah ada sistem ejaan dan sistem bahasanya sudah dilaporkan sebelumnya, masalah ini bukan lagi menjadi masalah.

IV. PELUANG PEMANFAATAN SUMBER DAYA DAN TEKNOLOGI

Pekerjaan besar untuk menyusun basis data korpus bahasa daerah tidak harus ditangani sendiri; sub-bab 2.3 telah menggambarkan suksesnya *Centre for English Corpus Linguistics*, di Belgia mengembangkan korpus pemelajar ICLE (*The International Corpus of Learner English*) dengan mengundang partisipasi sukarela sejumlah pihak tanpa mengorbankan keseragaman data yang terkumpul karena pihak-pihak yang berkontribusi menjalankan petunjuk-petunjuk pengumpulan data yang telah ditentukan. Cara crowdsourcing atau urun daya yang penuh gotong royong seperti ini sangat mungkin untuk dilakukan dalam hal pengembangan basis data korpus bahasa daerah mengingat semangat berkobar-kobar yang sering ditunjukkan terhadap persoalan kebahasaan. Selain itu, kemajuan teknologi memungkinkan peran serta khalayak ramai memberikan kontribusi kepada pengembangan korpus bahasa daerah tanpa batas waktu dan tempat. Yang berminat untuk ikut serta memberikan kontribusi bahan korpus bisa mendapatkan petunjuk-petunjuk pengumpulan data elektronik yang harus diikuti, atau bahkan mendapat pelatihan, mendapat pinjaman alat yang diperlukan (alat perekam, misalnya) dan nantinya mengunggah hasil pengumpulan bahannya ke dalam wadah penampungan data yang telah disiapkan.

Bahan-bahan korpus yang sudah terkumpul dari urun data tersebut kemudian disunting dan ditampilkan dalam bentuk basis data korpus BD yang terbuka untuk diakses khalayak ramai. Sudah terbukti bahwa banyak penelitian kebahasaan telah dihasilkan dari terbukanya akses terhadap berbagai korpus bahasa daring, seperti misalnya dapat dilihat dari daftar panjang penelitian (<http://corpus.byu.edu/researchers.asp>) yang berkaitan dengan korpus-korpus di laman <http://corpus.byu.edu/>. Bila basis data korpus BD yang sudah ada dibuka aksesnya untuk umum, dapat diharapkan akan muncul penelitian-penelitian tentang BD yang dapat meningkatkan pemahaman kita akan BD terkait.

Daftar Pustaka

- Anceaux, J.C. 1987. *Wolio dictionary: Wolio, English, Indonesian*. Dordrecht: Foris Publications Holland.
- Anthony, Laurence. 2014. AntConc (Versi 3.4.4). Tokyo: Waseda University. Diunduh dari <http://www.antlab.sci.waseda.ac.jp/software.html#antconc>
- Arka, I. Wayan. 2012. *Kamus bahasa Rongga-Indonesia: dengan pelacak kata Bahasa Indonesia-Rongga*, Penerbit Universitas Atma Jaya, Jakarta Indonesia.
- Arka, I. Wayan. 2013. Language management and minority language maintenance in (eastern) Indonesia: strategic issues *Language Documentation and Conservation*. Vol. 7, 74-105.
- Atkins, Sue, & Rundell, Michael. 2008. *The Oxford guide to practical lexicography*. Oxford: OUP.
- Berg, R. van den. & Sidu, La Ode. 1996. *English dictionary*. Leiden: KITLV Press.
- Berg, R. van den. & Sidu, La Ode. 2000. *Kamus Muna-Indonesia*. Kupang: Artha Wacana Press.
- Biber, Douglas, & Barbieri, Federica. 2007. Lexical bundles in university spoken and written registers. *English For Specific Purposes*, 26(3): 263-286.
- Biber, Douglas, Conrad, Susan, & Cortes, Viviana. 2004. If you look at. . . : Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3): 371-405.
- Budiwiyanto, Adi. Korpus dalam penyusunan kamus. <http://badanbahasa.kemdikbud.go.id/lamanbahasa/artikel/1580> Diakses 1 Juli 2016.
- Budiwiyanto, Adi. Pendokumentasian bahasa dalam upaya revitalisasi bahasa daerah yang terancam punah di Indonesia. <http://badanbahasa.kemdikbud.go.id/lamanbahasa/artikel/1823>. Diakses tanggal 16 Juni 2016.
- Cheng, Winnie. 2012. *Exploring corpus linguistics: language in action*. London and New York: Routledge.

- Collins, J. T. 2003. *Asilulu-English dictionary*. NUSA, Vol. 52-53. Jakarta: Pusat Kajian Bahasa dan Budaya Unika Atma Jaya.
- Cowie, A. P. 1999. *English Dictionaries for Foreign Learners: A History*. Oxford: OUP.
- Coxhead, Averil. 2000. A new academic word list. *TESOL Quarterly*, 34(2): 213-238.
- Darjowidjojo, Soenjono & John W.M. Verhaar (Ed.). 1983. *Miscellaneous Studies in Indonesian and Languages in Indonesia*, Part VII. NUSA, Vol. 15. Jakarta: Lembaga Bahasa Atma Jaya.
- Dardjowidjojo, Soenjono (Ed.). 1986. *Miscellaneous Studies in Indonesian and Languages in Indonesia*, Part VIII. NUSA, Vol. 25. Jakarta: Lembaga Bahasa Atma Jaya.
- Dardjowidjojo, Soenjono (Ed.). 1987. *Miscellaneous Studies in Indonesian and Languages in Indonesia*, Part IX. NUSA, Vol. 27. Jakarta: Lembaga Bahasa Atma Jaya.
- Davies, Williem. 2010. *A grammar of Madurese*. Berlin: Mouton de Gruyter.
- Donohue, Mark. 2009. *A grammar of Tukang Besi*. Berlin: Mouton de Gruyter.
- Elbar, Lambertus, Admojo, Wihadi, D., & Dominicus, Nanang. 1987. *Struktur bahasa Baru*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Halim, Amran (Ed.). 1977. *Miscellaneous studies in Indonesian and languages in Indonesia*, Part II. NUSA, Vol. 3. Jakarta: Lembaga Bahasa Atma Jaya.
- Halim, Amran (Ed.). 1979. *Miscellaneous studies in Indonesian and languages in Indonesia*, Part VI. NUSA, Vol. 7. Jakarta: Lembaga Bahasa Atma Jaya.
- Himmelman N.P. 1998. Documentary and descriptive linguistics. *Linguistics* 36: 161-195.
- Gardner, Dee, & Davies, Mark. 2014. A new academic vocabulary list. *Applied Linguistics*, 35(3): 305-327.
- Gilquin, Gaëtanelle, & Paquot, Magali 2007. *Spoken features in learner academic writing: identification, explanation and solution*. Makalah disajikan pada Fourth Corpus Linguistics Conference, University of Birmingham.
- Granger, Sylviane, & Meunier, Fanny. 2008. *Phraseology: an interdisciplinary perspective*. Amsterdam/Philadelphia: John Benjamins.
- Hananto, Markus, I Made, & Kurnia, Nany Setyono. 2011. *Frequency-based Indonesian word lists*. Makalah disajikan pada The second AISOFOLL (Annual International Symposium of Foreign Language Learning) Jakarta.
- Hyland, Ken. 2008. As Can Be Seen : Lexical Bundles And Disciplinary Variation. *English For Specific Purposes*, 27(1): 4-21.
- Johnson. 2016, 27 Feb. Je suis circonflexe: Why a minor fiddling with French spelling causes such anguish, *The Economist*. <http://www.economist.com/news/books-and->

arts/21693551-why-minor-fiddling-french-spelling-causes-such-anguish-je-suis-circonflexe Diakses 15 Juni 2016.

- Leech, Geoffrey. 2011. Frequency, corpora and language learning. Dalam Fanny Meunier, Sylvie De Cock, Gaëtanelle Gilquin & Magali Paquot (Ed.), *A taste for corpora: in honour of Sylviane Granger* (hal. 7-31). Amsterdam/Philadelphia: John Benjamins.
- Leech, Geoffrey, Rayson, Paul, & Wilson, Andrew. 2001. *Word frequencies in written and spoken English*. Harlow: Longman Pearson Education Limited.
- Lichfield, John. 2016, 4 Feb. Circumflex: Uproar over France's decision to implement 26-year-old reform of spelling rules: The accent is on caution in Education Ministry's new guide for text books, *The Independent*. <http://www.independent.co.uk/news/world/europe/french-grammar-uproar-over-decision-to-implement-26-year-old-reform-of-spelling-rules-a6854441.html> Diakses 15 Juni 2016.
- Manan, Umar, Amir, Zainuddin, Malano, Nasroel, Syafei, Anas, & Agustar Surin. 1986. *Struktur bahasa Muko-Muko*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Martinez, Ron, & Schmitt, Norbert. 2012. A Phrasal Expressions List. *Applied Linguistics*, 33(3): 299-320.
- Moon, Rosamund. 1998. *Fixed expressions and idioms in English: a corpus-based approach*. Oxford: Clarendon Press.
- Napsin, Syahrul, Naning, Zainal Abidin, Abdullah, Slamet, Sjamsuddin, Sjafran, Arsyad Mohammad, & Tarmizi. 1980/1981. *Morfologi dan sintaksis bahasa Rejang*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Nattinger, James R., & DeCarrico, Jeanette S. 1992. *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Nikelas, Syahwin, Asni Ayun, Zainil, Marah Rusmali, & Syofyan Adam. 1986. *Struktur Bahasa Pekal*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Masindan, Dra., Sinar, Dra. T. Silvana, Zulkifli, Drs., Muchtar, Drs. Muhizar, & Harahap, Dra. Oliviana. 1985. *Kamus Melayu Langkat – Indonesia*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Paquot, Magali. 2010. *Academic vocabulary in learner writing: from extraction to analysis*. London: Continuum.
- Pawley, Andrew, & Syder, Frances Hodgetts. 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. Dalam J. C. Richards & R. W. Schmidt (Ed.), *Language and communication* (hal. 191-226). New York: Longman.

- Poedjosoedarmo, Soepomo (Ed.). 1977. *Miscellaneous studies in Indonesian and languages in Indonesia*, Part III. Vol. 4. Jakarta: Lembaga Bahasa Atma Jaya.
- Phil, Quick. 2007. *A Grammar of the Pendau language of central Sulawesi, Indonesia*. Canberra: ANU.
- Purwo, Bambang Kaswanti (Ed.). 1990. *Miscellaneous studies in Indonesian and languages in Indonesia*, Part X. *NUSA*, Vol. 32. Jakarta: Lembaga Bahasa Atma Jaya.
- Pusat Bahasa. 2008. *Kamus Besar Bahasa Indonesia Pusat Bahasa* (Edisi keempat). Jakarta: Gramedia.
- Rundell, Michael (Ed.). (2007). *Macmillan English Dictionary for Advanced Learners* (Edisi Kedua). Oxford: Macmillan Publishers Limited.
- Rundell, Michael, & Granger, Sylviane. 2007. From corpora to confidence. *English Teaching professional*(50): 15-18.
- Shin, Dongkwang, & Nation, Paul. 2008. Beyond single words: the most frequent collocations in spoken English. *ELT Journal*, 62(4): 339-348.
- Simpson-Vlach, Rita, & Ellis, Nick C. 2010. An academic formulas list: new methods in phraseology research. *Applied Linguistics*, 31(4): 487-512.
- Sofyan, Inghuong Alias, Kaseng, Syahrudin, Sikki, M., & Pepy, Patuko. 1979. *Morfologi dan sintaksis bahasa Kaili*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Suharno, Ignatius (Ed.). 1977. *Miscellaneous studies in Indonesian and languages in Indonesia*, Part IV *NUSA*, Vol. 5. Jakarta: Lembaga Bahasa Atma Jaya.
- Sulaiman, Abu Bakar, Gani, A., & K., Syafri. 1986. *Kosa kata bahasa Melayu Riau*. 1986. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Troeboes, Maryanto, Sandi, Gambar, Abraham, Antonius, Porat, and Hana, Alfons. 1987. *Morfologi dan sintaksis bahasa Komodo*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Verhaar, John W.M. (Ed.). 1975. *Miscellaneous Studies in Indonesian and Languages in Indonesia*, Part I. *NUSA*, Vol. 1. Jakarta: Lembaga Bahasa Atma Jaya.
- Verhaar, John W.M. (Ed.). 1978. *Miscellaneous studies in Indonesian and languages in Indonesia*, Part V. *NUSA*, Vol. 6. Jakarta: Lembaga Bahasa Atma Jaya.
- West, Michael. 1953. *A general service list of English words*. Essex: Longman.
- Winchester, Simon. 2003. *The meaning of everything: The story of the Oxford English Dictionary*. Oxford: OUP.
- Xue, Gou-yi, & Nation, Paul. 1984. A University World List. *Language Learning and Communication*, 3(2): 215-229.

Survei Program Pengolah Korpus untuk Data Bahasa Indonesia dan Bahasa Daerah di Indonesia

Prihantoro

Universitas Diponegoro
prihantoro2001@yahoo.co

Abstrak

Paper ini membandingkan beberapa program pengolah korpus (*Corpus Processing Tool*) yang cocok digunakan oleh para linguist yang tidak memiliki pengetahuan komputasi (*scripting, coding atau programming*) sama sekali. pada dasarnya, program-program tersebut bisa dikategorikan dalam tiga tipe. Pertama, program bertipe *free concordancer* yang bisa memuat teks dan melakukan penelusuran berbasiskan kata kunci. Namun program seperti ini biasanya terbatas hanya untuk korpus mentah, dan bahasa-bahasa yang aksaranya sudah diromanisasi seperti bahasa Inggris. Kedua, adalah program dengan fungsi NLP terbatas. AntConc memberikan kebebasan bagi para pengguna untuk melakukan anotasi. Namun sayang anotasi yang bisa dilakukan sangat sederhana, sebatas identifikasi lema dan kata. Untuk melakukan anotasi, teks harus dianotasi terlebih dahulu dengan program lain AnTag, lalu diproses dengan AntConc. Ini serupa dengan Xaira dan Wordsmith. Berbeda dengan program-program ini, Geuljabi memiliki lexical resource Bahasa Korea yang built in, sehingga user tidak harus berpindah program. Tapi dalam konteks bahasa Indonesia atau daerah, tetap saja program-program ini tidak cocok untuk bahasa Indonesia. Program yang cocok untuk bahasa di Indonesia adalah jenis yang ke tiga, yaitu memiliki fungsi-fungsi lengkap *Natural Language Processing* (NLP) seperti desain korpus, sistem aksara, varian ortografi, anotasi morfologi, semantik, dan sintaksis, seperti Unitex. Unitex telah teruji dalam desain serta pemrosesan korpus berbagai bahasa, termasuk bahasa-bahasa dengan aksara khusus dan tingkat kompleksitas linguistik yang beragam seperti *free order language* (Rusia), *logographic language* (Cina), *agglutinative language* (Korea), serta *Consonant Skeleton/Semitic Language* (Arab, Yahudi). Tidak seperti program lain, Unitex membebaskan *user* untuk mendesain *resource*-nya sendiri. Saat ini, sudah tersedia prototip *lexical resource* Bahasa Indonesia

(Prihantoro, 2011), digunakan pada Unitex, dan Bahasa Jawa, yang terus disempurnakan. Resource bahasa Indonesia (Kilgariff, Reddy, Pomikálek, & Avinesh, 2010), berjenis non-morfologis, digunakan oleh (Kwary, 2013) yang bisa diaplikasikan pada AntWord Profiler. Dalam program-program komputasi korpus berbasiskan pendekatan *machine learning* atau *stochastic*, pengambilan keputusan dalam kasus-kasus tertentu diambil alih oleh sistem (misal, *ambiguity resolution*) yang bergantung pada *reference corpus* atau *training corpus*. Pada Unitex/Nooj, presisi analisis bergantung penuh pada *resource* (aksara, ortografi, morfologi, semantik, sintaksis) yang didesain oleh *user*, sehingga pengambil keputusan adalah user itu sendiri, bukan sistem. Meski terkesan rumit, kompleks dan *labour-intensive*, pendekatan ini dinilai lebih cocok untuk para linguist yang memang bertujuan mendeskripsikan fenomena linguistik berorientasi pada konteks.

I. PENDAHULUAN

Dalam linguistik korpus, paling tidak ada tiga ranah pengembangan kepakaran. Ranah pertama adalah mengenai bangunan korpus itu sendiri, yang meliputi pengumpulan data dan arsitektur korpus. Ranah ke dua adalah pengolahan korpus, mulai dari *character recognition*, tokenisasi, anotasi serta penelusuran dan ekstraksi informasi linguistik. Ranah ke tiga, adalah membaca hasil pengolahan korpus, yaitu *concordance* dan statistik (Mc Enery & Hardie, 2012).

Paper ini memfokuskan pada ranah ke dua, yaitu pengolahan korpus. Ranah ini hingga saat ini didominasi oleh para ahli komputer, karena erat kaitannya dengan teknologii informasi sehingga membutuhkan kemampuan komputasi seperti *scripting*, *programming*, dan *coding*. Perhatikan beberapa program yang didedikasikan untuk BI seperti iPoS tagger (Farizki & Purwarianti, 2010), InaMorph (Pisceldo, Mahendra, Manurung, & Arka, 2008), MorphInd (Larasati, Kuboň, & Zeman, Indonesian Morphology Tool (MorphInd): Towards an Indonesian Corpus, 2011), yang penggunaanya menuntut penggunanya familier dengan metode komputasi. Peran linguist secara langsung dalam ranah ini sedikit sekali karena, 1) perangkat lunak yang digunakan membutuhkan kompetensi komputasi, 2) tidak banyak linguist yang memiliki kemampuan komputasi, dan 3) banyak linguist lebih tertarik pada bahasa pada konteks penggunaan (misal; pragmatik atau analisis wacana).

Namun penulis berpendapat, bahwa untuk bahasa-bahasa di Indonesia, linguist perlu meningkatkan kontribusinya pada ranah ke dua tersebut, paling tidak untuk beberapa alasan ini. Pertama, beberapa program-program yang didesain para ahli komputer

menggunakan metode stokastik/probabilitas, dan membutuhkan input dari linguist agar kinerja program menjadi optimal. Ke dua, linguist yang membaca hasil *concordance* atau statistic juga akan terbantu karena korpus akan dianalisis terlebih dahulu. Ke tiga, *computing resource* untuk bahasa Indonesia atau bahasa-bahasa daerah di Indonesia bisa dikatakan sangat sedikit. Khusus untuk bahasa-bahasa daerah, linguist agak sulit berkontribusi dalam ranah korpus ini, karena program-program yang sudah ada didesain khusus untuk satu bahasa.

Solusi untuk permasalahan ini adalah 1) para programmer perlu menyederhanakan desain program yang mereka buat agar cukup leluasa digunakan oleh para linguist, bahkan yang tidak memiliki kemampuan programming. 2) Program tersebut haruslah memberikan keleluasan bagi para linguist untuk berkreasi sesuai dengan fitur bahasa yang mereka analisis.

Untuk mencari solusi permasalahan tersebut ada dua pilihan; 1) membuat sendiri program yang memenuhi kriteria tersebut, atau 2) menggunakan program yang sudah ada. Solusi pertama tentu lebih baik, namun membutuhkan waktu yang lama. Program seperti ini biasanya berupa proyek tim dengan dana yang tidak sedikit. Namun kelebihanannya adalah kita bisa mendesain sendiri, fitur-fitur apa saja harus dimiliki program tersebut, yang disesuaikan dengan bahasa yang akan dianalisis. Pengguna, yang memiliki kemampuan programming, juga bisa langsung menggunakan bahasa pemrograman yang biasa digunakan untuk fungsi Natural Language Processing, seperti Perl atau Python. Pada pengolahan korpus dengan bahasa pemrograman seperti Python (Bird, Klein, & Loper, 2009), wordlist harus 'dipanggil' dengan perintah yang kompleks. Perhatikan perintah berikut dari Python yang digunakan untuk mencari wordlist (daftar token) pada korpus:

```
>>> from nltk.corpus import PlaintextCorpusReader
>>> corpus_root = '/usr/share/dict'
>>> wordlists = PlaintextCorpusReader(corpus_root, '.*')
>>> wordlists.fileids()
['README', 'connectives', 'propernames', 'web2', 'web2a', 'words']
```

Perintah di atas bisa diterjemahkan sebagai berikut. 1) gunakan modul korpus untuk mengenali format korpus *.txt, 2) temukan direktori/lokasi korpus pada hard-drive, 3) kenali semua kata pada korpus, 4) tampilkan daftar kata pada layar. Tentu saja bagi orang yang menguasai dasar-dasar pemrograman, perintah di atas cukup sederhana. Paper ini memfokuskan pada solusi yang ke-dua. Hal ini dikarenakan ternyata ada beberapa program, yang dihasilkan oleh programmer yang juga linguist. Sehingga, program-program ini mampu mengakomodasi kepentingan linguist. Pada bagian pembahasan, kami akan menjelaskan hasil survey kami mengenai kelebihan dan kekurangan program-program tersebut.

II. KERANGKA TEORI DAN METODE

Jenis-jenis tools, atau program pengolah korpus dijelaskan dengan baik dan runtut oleh Mc Enery & Hardie (2012). Di dalam penjelasan tersebut, secara historis, program pengolah korpus terbagi menjadi empat. Generasi pertama adalah concordancer yang eksklusif di satu pusat penelitian. Generasi kedua adalah concordancer yang sudah didistribusikan secara umum, namun fungsi penelusuran korpus beranotasi belumlah maksimal. Generasi ke tiga adalah concordancer yang mampu mengolah korpus beranotasi secara maksimal dan dilengkapi fungsi-fungsi statistik dan dekskriptif. Generasi ke empat secara fungsionalitas sebetulnya tidak terlalu berbeda dengan generasi ke tiga. Kemunculan concordancer generasi ke empat ini memfokuskan pada masalah 1) kompatibilitas PC, dan 2) distribusi korpus secara legal. Sehingga muncullah concordancer online yang bisa diakses tanpa batasan spesifikasi minimal apapun, dan mampu memberikan akses terhadap korpus-korpus yang datanya bebas digunakan.

Namun dalam paper ini, survei yang dilakukan memfokuskan pada tool yang berfungsi secara offline dan dapat memproses korpus-korpus personal. Perbedaan selanjutnya adalah bahwa penggolongan tipe program korpus adalah berdasarkan fungsionalitasnya. Metode penelitian yang digunakan adalah observasi partisipatoris. Penulis tidak hanya menggunakan referensi seperti kajian yang sudah dilakukan dengan tool tersebut atau manual yang ada, namun juga menggunakan tool tersebut. Dari pengalaman empiris itu, maka dirumuskanlah tiga tipe tool berdasarkan derajat fungsionalitas mereka.

III. PEMBAHASAN

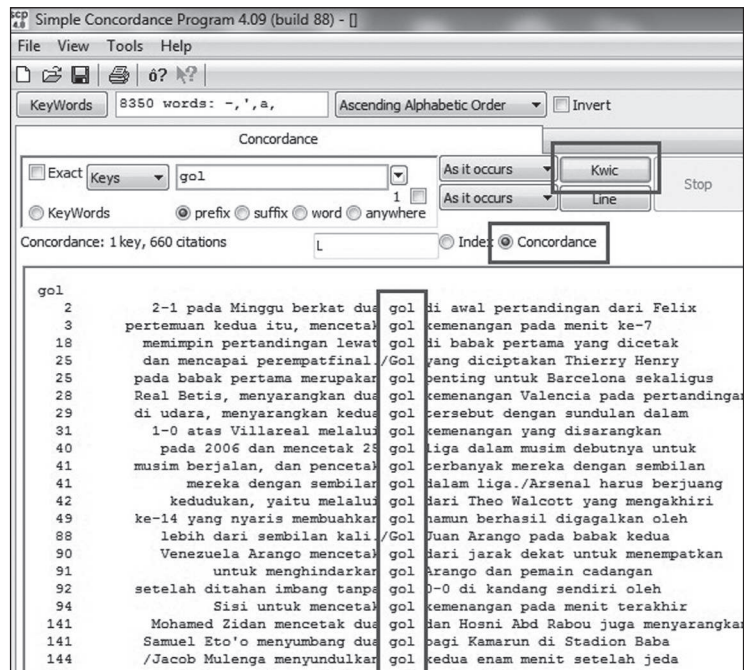
3.1 Tipe 1: Concordancer Murni

Program yang digolongkan pada tipe pertama adalah free concordancer (seperti Simple Concordancer¹ atau TextStat²), yang banyak sekali tersebar di internet dan bisa didownload secara gratis. Namun program seperti ini hanya bisa mengolah korpus mentah. Program ini sangat sederhana *interface* dan pengoperasiannya, sehingga biasa digunakan dalam kelas pengenalan linguistik korpus. User biasanya hanya harus mengupload teks berformat txt, lalu penelusuran berbasiskan ortografi bisa dilakukan.

Selama karakter dari bahasa yang dianalisis sudah bisa dikenali, concordancer ini bisa digunakan. Meski kami menggunakan program free concordancer untuk pemula, namun para linguis korpus tingkat mahir juga bisa menggunakan program ini untuk analisisnya apabila kajian korpus yang dilakukan hanya membutuhkan korpus mentah.

¹ <http://www.textworld.com/scp/>

² <http://neon.niederlandistik.fu-berlin.de/en/textstat/>



Gambar 1. Concordance dari Simple Concordance Program

3.2 Tipe 2: Korpus Prosesor dengan Fungsi Statistik, Deskriptif dan NLP Terbatas

Program tipe ke dua, adalah program-program yang sudah memiliki fungsi-fungsi NLP namun masih perlu dikembangkan lebih lanjut. Salah satu contohnya adalah AntConc³. AntConc adalah salah satu program yang ditulis oleh Dr. Lawrence Anthony. Menurut penulis, inilah program yang user-interfacenya sangat baik, dan mudah untuk dipelajari. AntConc banyak digunakan dalam pengolahan korpus mentah baik oleh para pemula maupun tingkat mahir.

Di dalam link yang bisa didownload terpisah, AntConc menyediakan lemma list, modul lematisasi (mirip dengan *stemming*, atau lebih tepatnya *lemmatizing* dalam istilah komputasi) untuk bahasa inggris yang struktur datanya sangat sederhana. Modul ini digunakan untuk mengidentifikasi bentuk infleksi dan derivasi dari satu lema dan menghitung frekwensinya, atau bentuk kontraksi. *Lemma List* pada AntConc berisi daftar lema dan daftar word form dengan notasi sebagai berikut: lema -> *wordr form1, word form2, word form3*. Tujuan dari lemma list ini sendiri adalah menunjukkan varian word form yang terdeteksi dalam teks yang kita analisis. Anda bisa mendownload lemma list di sini (<http://www.laurenceanthony.net/software/antconc/>). Jika dibuka, file berformat .txt tersebut akan nampak seperti ini:

³ <http://www.laurenceanthony.net/software/antconc/>

abandon -> abandons,abandoning,abandoned

abase -> abases,abasing,abased

abate -> abates,abating,abated

abbess -> abbesses

abbey -> abbeys

Untuk menggunakan lemma list ini, klik *tool preferences > wordlist*. Pada *lemma list*, telusuri file lemma list sesuai dengan folder dimana anda menyimpan file tersebut. Lalu klik Load. Pada *word list range*, pilih *use all words*. Pada pilihan *add words from file*, klik open lalu telusuri file *lemma list* yang sama pada folder tersebut. Jika semua sudah selesai, klik *apply*. Setelah itu, klik *start* pada kotak penelusuran.

Rank	Freq	Lemma	Lemma Word Form(s)
1	116	the	
2	71	of	
3	48	and	
4	48	to	
5	45	be	am 1 are 3 be 23 been 7 is 7 was 4
6	36	which	
7	31	in	
8	24	a	a 14 an 10
9	23	i	
10	22	my	
11	21	that	that 18 those 3
12	20	by	
13	19	have	had 1 has 5 have 12 having 1
14	17	with	

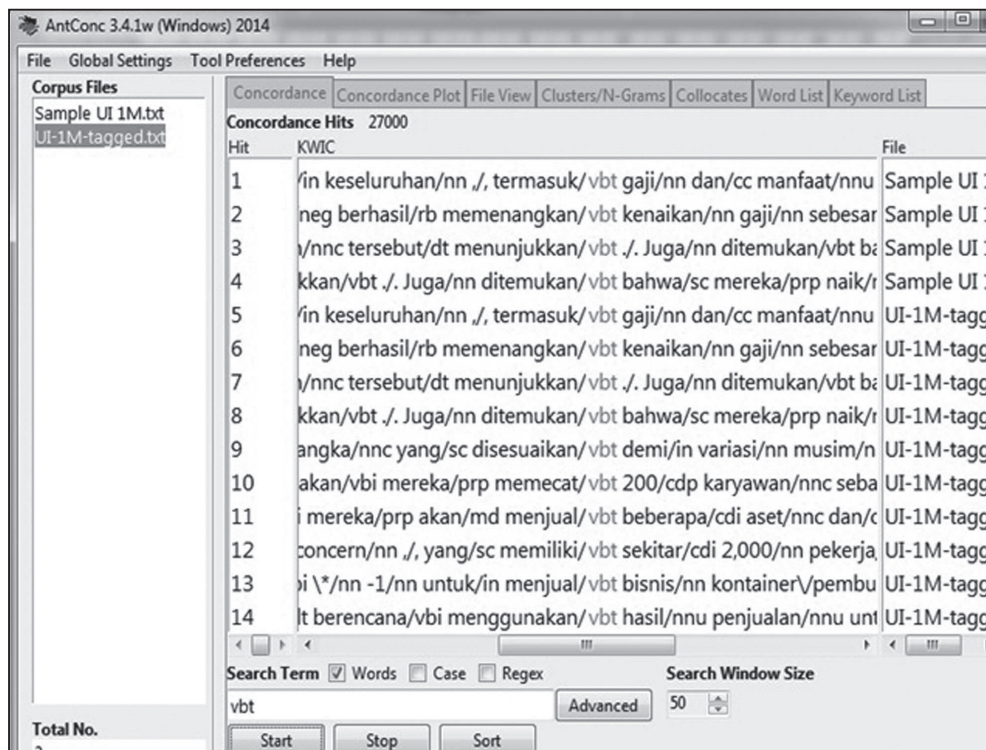
Gambar 2. Daftar Lema dan Wordform pada AntConc

Kita lihat kolom word telah berubah menjadi lemma. Sedangkan kolom lemma word forms yang tadinya kosong telah terisi sebagian. Kenapa cuma sebagian? Mari kita telusuri cara berpikir komputer. Pertama, komputer mendeteksi token yang ada dalam teks. Ke dua, komputer akan memeriksa, apakah ada dari beberapa token yang merupakan word form dari satu lema yang sama. Apabila ada, maka hasilnya akan ditampilkan. Contoh pertama, komputer mendeteksi token *a* sebanyak 14 dan token *an* sebanyak 10. Setelah berkonsultasi dengan *lemma list*, ia mengetahui bahwa dua token tersebut berasal dari satu lema, *a* (*a -> an*). Sehingga ditampilkanlah *a* pada kolom lemma, serta *a 14 an 10* pada lemma word form(s). Contoh ke dua, komputer menemukan token-token berikut *am 1 are 3 be 23 been 7 is 7 was 4*. Setelah berkonsultasi dengan lemma list, ia menemukan bahwa ke enam token tersebut berasal dari satu lema *be* (*be -> am,ar,e,is,was,were,being,been,'m,m*). Perhatikan bahwa dalam lemma list ada 9 jenis *word form*. Namun karena yang terdeteksi di teks hanya enam, maka enam token itulah yang ditampilkan. Dengan meniru format modul lemma list untuk bahasa Inggris, kita bisa membuat lemma list yang sama untuk bahasa Indonesia. Berikut contoh struktur data lemma list untuk bahasa Indonesia

[daftar lemma list untuk bahasa Indonesia
 [masih belum lengkap, bisa dikembangkan lebih jauh
 [entry line sbb
 [lemma_->_word_form1,word_form2,word_form3

makan -> makan,memakan,dimakan,termakan
 jalan -> jalan,berjalan,menjalani,dijalani,jalankan
 juang -> berjuang

Selain, AntConc, Dr. Anthony juga membuat beberapa program terpisah seperti AntTag, yang berfungsi melakukan anotasi. Ini agak berbeda dengan Xaira dan Wordsmith yang menggabungkan fungsi anotasi dalam satu program. Karena menggunakan tagger CLAW BNC, AntTag hanya bisa menganotasi teks berbahasa Inggris. Teks harus diupload dan dianalisis oleh AntTag terlebih dahulu, lalu dicopy-paste sebagai file txt, dan diupload ke AntConc untuk dianalisis. Setelah ditag dan diupload, user bisa melakukan penelusuran berdasarkan tag yang ada. Bisa juga user mengupload korpus yang sudah bertag, misalnya korpus UI-1M dari Pan Localization Project berikut:

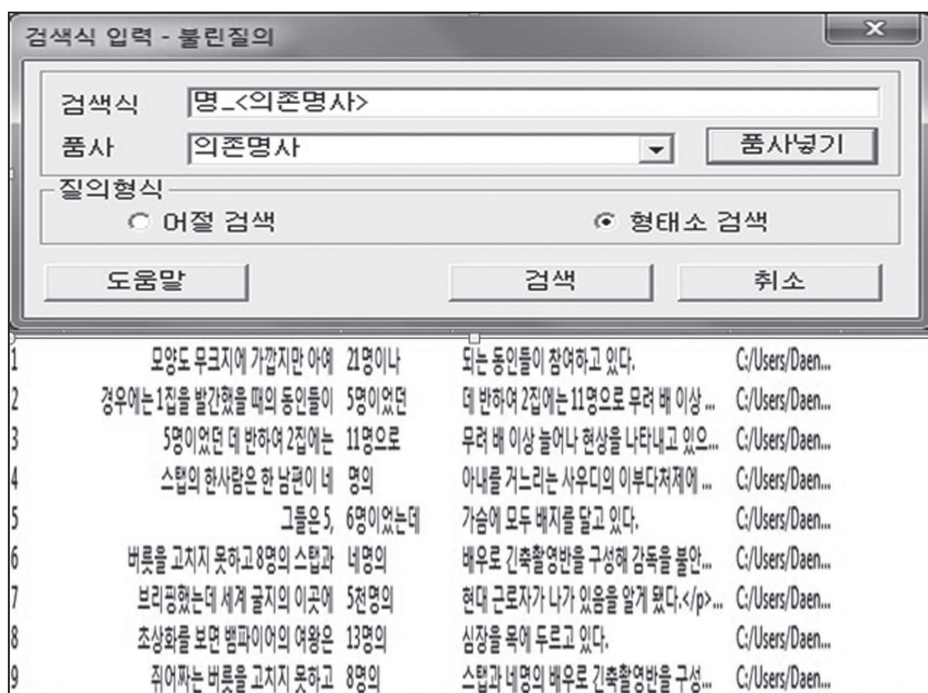


Gambar 3. Penelusuran Tag Korpus UI-1M menggunakan AntConc

Selain AntTag ada juga AntWordProfiler, program statistik yang mirip seperti program Range yang dibuat oleh Prof. Paul Nation, yang berfungsi mengukur derajat eksistensi sekelompok kata dalam satu teks. Permasalahannya adalah, struktur data lexical resource yang digunakan pada AntWordProfiler berbeda dengan lemma list pada AntConc, sehingga data harus ditulis ulang dengan format yang berbeda.

Kemudian ada WordSmith⁴ dan Xaira⁵. WordSmith dan Xaira memiliki beberapa fungsi statistik seperti T-Score, Mutual Information, atau loglikelihood yang lebih advance dari AntConc. Serupa dengan AntConc, text harus diPOStag dulu dengan program lain (atau manual, tapi tidak disarankan), sebelum diproses lebih jauh. Namun menurut hemat penulis, AntConc jauh lebih mudah digunakan dibanding Xaira dan Wordsmith.

Geuljabi, adalah sebuah program analisis korpus buatan National Korean Institute of Korean Language (Badan Bahasanya Korea). Tidak seperti AntConc, Xaira atau Wordsmith, Geuljabi memiliki beberapa keistimewaan. Pertama, Geuljabi memiliki Lexical Resource (LR) tersendiri yang sudah built-in. User tidak harus berpindah ke program lain (apalagi menganotasi secara manual) untuk melakukan tagging pada korpus. Kedua, tagging pada korpus sampai ke level morfologis.



Gambar 4. Penelusuran Classifier Menggunakan Gelujabi

⁴ <http://www.lexically.net/wordsmith/version5/index.html>

⁵ <http://xaira.sourceforge.net/>

Namun, Geuljabi memiliki beberapa kelemahan yang cukup mendasar. Penulis hanya berhasil menggunakan Geuljabi di OS berbahasa Korea (Korean Windows). Kemudian, teks yang dianalisis tidak bisa terlalu banyak; karena akan menimbulkan crash pada system.

Dalam konteks kebutuhan bahasa Indonesia atau daerah, tentu saja Geuljabi terbatas seperti tipe 1, karena LR yang ada tidak bisa digunakan untuk bahasa Indonesia. Meski demikian, bagi peneliti bahasa Korea, Geuljabi adalah program yang sangat baik dan cukup nyaman digunakan oleh para linguist Korea karena user interface yang sederhana dan mudah dipahami (oleh orang yang paham bahasa Korea tentunya).

3.3 Tipe 3: Korpus Prosesor dengan Fungsi NLP Optimal

Program tipe ke-tiga memiliki fungsi-fungsi NLP dasar yang lengkap, namun masih cukup ramah untuk digunakan tanpa kemampuan programming atau scripting. Contoh program seperti ini adalah Intex⁶, Unitex⁷, dan NooJ⁸. Pengguna dapat saja langsung mempelajari program ini, namun penulis menyarankan menguasai concordancer terlebih dahulu, lalu AntConc, dan baru menuju Unitex, Intex, atau NooJ.

Pada sebuah program kita sering melihat ada nama ekstensi seperti versi beta, versi 1.2, versi 2.2 dan sebagainya. Versi beta maksudnya adalah program tersebut masih dalam percobaan. Sedangkan penomoran digunakan untuk membedakan versi lebih baru dan yang lama. Versi yang lebih baru biasanya lebih baik dari versi yang lama, baik dalam hal penambahan fungsi, atau aspek lain seperti tampilan atau kecepatan. Khusus untuk Intex, saat ini sudah tidak dikembangkan lagi. Unitex dan Intex dikembangkan tidak hanya oleh programmernya, tapi juga dengan bantuan komunitasnya. Terkadang, komunitas pengguna inilah yang memberikan input mengenai fitur yang harus dikembangkan baik untuk pengajaran, analisis, maupun teknologi informasi.

Pada kedua program ini, sistem program dan modul linguistik adalah dua hal yang berbeda. Program hanya menjalankan fungsi-fungsi dasar NLP seperti manajemen korpus, tokenisasi, desain lexical resource, machine readable grammars, automatic retrieval, information extraction, collocation score dan sebagainya.

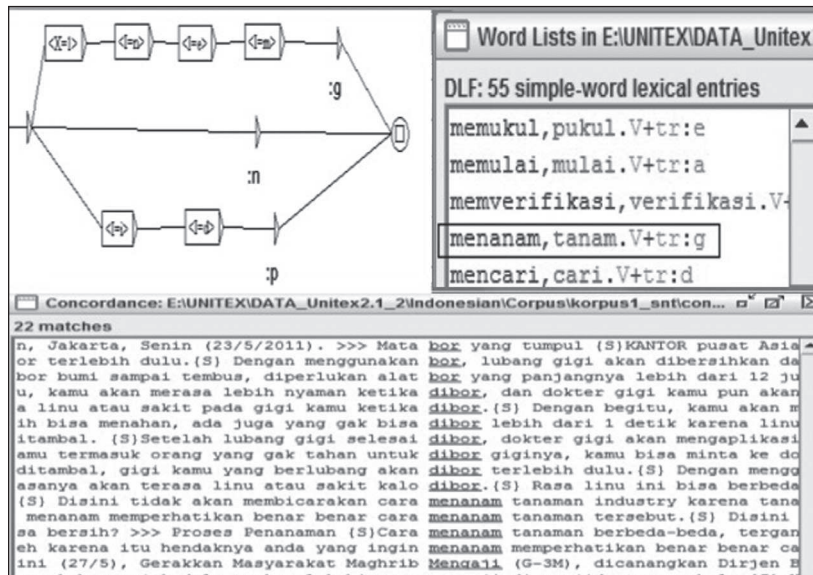
Sedangkan modul linguistik pada ke dua program ini bisa didesain secara bebas oleh para penggunanya. Dengan demikian, kita tidak perlu mengacu pada tag standar Bahasa Inggris. Unitex dan NooJ didesain dengan prinsip bahwa setiap bahasa memiliki keunikan tersendiri, sehingga modul linguistik yang digunakan bisa berbeda antara satu bahasa dan bahasa lain. Bahkan dalam bahasa yang sama, pengguna bisa memberikan kode yang berbeda sesuai dengan pendapatnya. Contohnya, prefix berkode anotasi PFX pada pengguna pertama. Pengguna lain bisa saja mendesain kode anotasi yang

⁶ <http://www.nyu.edu/pages/linguistics/intex/>

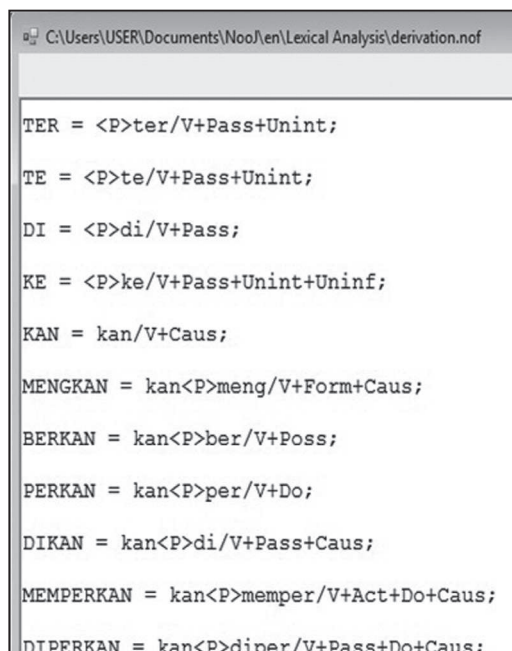
⁷ <http://www-igm.univ-mlv.fr/~unitex/>

⁸ <http://www.nooj-association.org/>

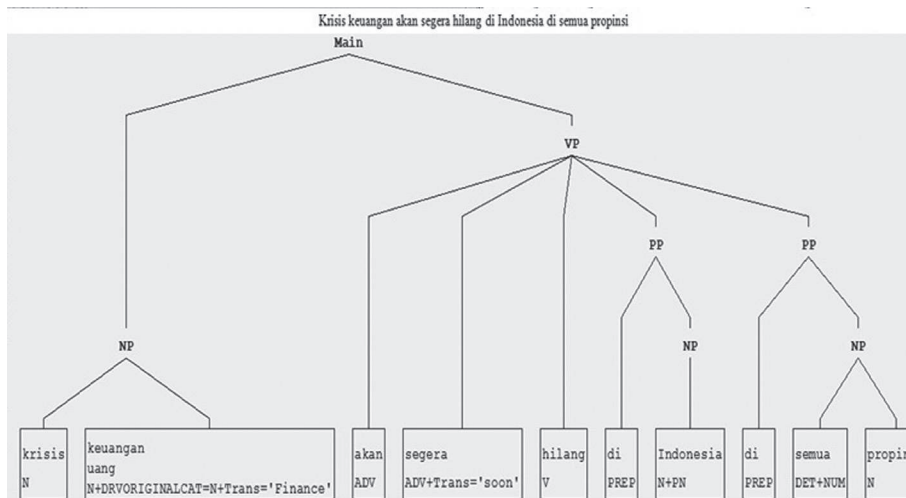
berbeda, misalnya PREF. Misalkan dalam satu bahasa daerah di Papua, seorang linguis berargumen bahwa classifier +HUMAN bisa mendeskripsikan benda mati seperti mobil atau pesawat terbang, sehingga anotasi +Hum diberikan. Namun linguis lain berpendapat, bahwa sebetulnya konsep yang tepat adalah mobilitas, sehingga seharusnya +Mob yang harus melekat pada classifier tersebut.



Gambar 5. Contoh Lexical dan Grammatical Resources untuk Prefiks Bahasa Indonesia dengan Unitex (Prihantoro 2011)



Gambar 6. Derivational Grammar Resource untuk Bahasa Indonesia dengan Nooj



Gambar 7. Parse Tree untuk Kalimat Bahasa Indonesia dengan Nooj

Sistem distribusi/sharing modul linguistik juga sangat sederhana yaitu dengan mengupload folder yang berisi modul tersebut atau mengirimkannya melalui email. Folder tersebut bisa didownload lalu dicopy ke direktori pada lokasi modul linguistik tersebut digunakan. Bahkan bisa juga hanya dicopy-paste.

Tantangan utama untuk pemrosesan bahasa daerah adalah pada bahasa-bahasa yang memiliki aksara tersendiri seperti aksara Hanacaraka (Bahasa Jawa), Aksara Kaganga (bahasa Lampung), atau aksara lain, karena sistem harus mendesain sistem font recognition. Hal ini akan menjadi lebih mudah, apabila sudah terdapat encoding font Unicode untuk bahasa tersebut. Namun hal ini sepertinya tidak menjadi masalah, karena kedua program ini sudah berhasil memproses bahasa bahasa dengan aksara tersendiri, misalnya bahasa Arab yang ditulis dari kiri ke kanan dengan atau tanpa diakritik, bahasa Korea yang menggunakan system *syllable block* dan berjenis aglutinatif, bahasa Cina yang bertipologi isolated language dengan sistem karakter logografik dan bahasa lainnya. Khusus untuk hal ini, jika tidak memiliki kemampuan programming, pengguna tidak bisa melakukannya sendiri. Namun pengguna bisa meminta bantuan programmer kedua program ini atau komunitasnya.

Berikut kami tampilkan estimasi penguasaan ke-tiga jenis program yang telah dibahas sebelumnya. Untuk Unitex/Nooj, pengguna bisa memangkas waktu kurang lebih dua jam apabila sudah menguasai AntConc, WordSmith, Xaira atau program sejenis.

Concordancer	WordSmith/Xaira/ Geuljabi	AntConc	Unitex/Intex/Nooj
30 menit	5 jam	2 jam	8 Jam
Easy	Advanced	Intermediate	Advanced
<ul style="list-style-type: none"> ✓ Download dan instalasi ✓ manajemen teks korpus ✓ penelusuran otomatis berbasis ortografi ✓ kompleksitas: regular grammar ✓ contoh analisis ✓ praktek 	<ul style="list-style-type: none"> ✓ download dan instalasi ✓ Manajemen teks korpus ✓ Penelusuran otomatis berbasis ortografi ✓ Deskripsi tag anotasi bahasa sasaran (Inggris/Korea) ✓ Penelusuran otomatis berbasis tag ✓ Statistik ✓ Visualisasi data ✓ Kompleksitas: regular grammar, context-free grammar ✓ Contoh analisis ✓ Praktek 	<ul style="list-style-type: none"> ✓ download dan instalasi ✓ Manajemen teks korpus ✓ Penelusuran otomatis berbasis ortografi ✓ Statistik ✓ Visualisasi data ✓ Kompleksitas: regular grammar ✓ Contoh analisis ✓ Praktek 	<ul style="list-style-type: none"> ✓ download dan instalasi ✓ Manajemen teks korpus ✓ Penelusuran otomatis berbasis ortografi ✓ Deskripsi tag anotasi bahasa sasaran (gramatika, semantik, morfologi) ✓ Desain Lexical Resource ✓ Penelusuran otomatis berbasis tag ✓ Statistik ✓ Visualisasi data ✓ Desain lexical resource (lexicosemantic) ✓ Machine readable grammars (infleksi/derivasi, sintaksis) ✓ Output Variable and Context ✓ Kompleksitas: regular grammar, context free grammar, context- sensitive grammar ✓ Contoh analisis ✓ Praktek

Tabel 1. Estimasi Penguasaan Fungsi Dasar Program Pengolah Korpus

IV. SIMPULAN

Meski dalam paper ini penulis membahas mengenai software atau tool pengolahan korpus, namun penulis berharap user tidak ‘silau’ dengan kecanggihan software atau tool tersebut. Pertama, dalam dunia komputer, software atau tool akan terus berkembang dan bukan tidak mungkin akan muncul tool atau software baru yang lebih baik. Kedua, ini yang paling penting, tool atau software tadi tidak bisa beroperasi maksimal tanpa resource yang memadai, khususnya leksikon. Mengapa leksikon? Desain leksikon membutuhkan, selain ketelitian, juga tenaga, karena kuantitasnya yang cukup besar, dan idealnya memang dihandle oleh tim, bukan perseorangan.

Sejauh ini, kalangan yang banyak mempublikasi *computational resource* untuk bahasa Indonesia berasal dari kalangan ahli komputer, bukan linguis. Ada beberapa linguis yang menganggap bahwa *resource* adalah satu *reference work*, dan tidak bisa dianggap sebagai sebuah karya akademis, karena tidak mendeskripsikan fenomena ‘menarik’ secara khusus seperti *voice*, *transitivity*, *morphological mirroring* dan sebagainya. Meski masih *debatable*, hal ini seharusnya tidak mematikan niat mereka yang ingin menulis *resource* ini baik *lexical/grammatical resource*. Perlu dipahami bahwa setiap area linguistik perlu mendapatkan perhatian, karena memberikan kontribusi penting bagi bahasa itu sendiri. Hal ini bisa dilakukan tanpa menegasikan kontribusi area yang lain.

Dalam analisis Dalam bidang leksikografi, terutama *electronic lexicography*, membuat *resource* adalah kontribusi yang sangat positif. Penulis berpendapat, selain paper-paper atau laporan mengenai sistem fonetik, morfologi, sintaksis suatu bahasa, perlu juga didesain *resource* komputasi yang memadai. Hal ini juga berlaku untuk bahasa Indonesia maupun daerah.

Daftar Pustaka

- Anthony, L. (2006). Concordancing with AntConc: An introduction to tools and techniques in corpus linguistics. *JACET Newsletter*, 155-185.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. Sebastopol: O'Reilly Publishing.
- Farizki, A., & Purwarianti, A. (2010). HMM Based Part-of-Speech Tagger for Bahasa Indonesia. *On Proceedings of 4th International MALINDO (Malay and Indonesian Language) Workshop, 2nd August 2010* (pp. 123-45). Jakarta: Universitas Indonesia Press.
- Gross, M. (1997). The Construction of Local Grammars. In E. Roche, & Y. Schabes, *Finite State Language Processing* (pp. 329-354). Massachusetts: MIT Press.
- Kilgarrieff, A., Reddy, S., Pomikálek, J., & Avinesh, P. (2010). A Corpus Factory for Many Languages. In LREC. LREC, (pp. 27-34).

- Kwary, D.-A. (2013). *Indonesian High Frequency Words*. Retrieved August 20, 2013, from The Indonesian High Frequency Word List (Version April 2013): <http://www.kwary.net/iwl.html>
- Larasati, S.-D. (2012). IDENTIC CORPUS: Morphologically Enriched Indonesian-English Parallel Corpus. *LREC*, 902-906.
- Larasati, S.-D., Kuboň, V., & Zeman, D. (2011). Indonesian Morphology Tool (MorphInd): Towards an Indonesian Corpus. *Systems and Frameworks for Computational Morphology* (pp. 119-129). Zurich, Switzerland: Springer Berlin Heidelberg.
- Lewis, M.-P., Simons, G.-F., & Fennig, C.-D. (2013). *Ethnologue: Languages of the World, Seventeenth edition*. Dallas, Texas: SIL International.
- Mc Enery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Paumier, S. (2008). *Unitex Manual*. Paris: Universite Paris Est Marne La Valee & LADL.
- Pisceldo, F., Mahendra, R., Manurung, R., & Arka, I.-W. (2008). A Two Level Morphological Analyser for the Indonesian Language. *Australasia Technology Association Workshop*, (pp. 142-150).
- Prihantoro. (2011). Transducer for Auto-Convert of Archaic to Present Day English: A Support for Computer Assisted Language Learning. *Proceeding for The 16th English in Southeast Asia Conference*. Yogyakarta: Sanata Dharma University Press.
- Prihantoro. (2011b). Local Grammar Based Auto Prefixing Model for Automatic Extraction in Indonesian Corpus (Focus on Prefix MeN-). *Proceedings of International Congress of Indonesian Linguists Society (KIMLI)* (pp. 32-38). Bandung: Universitas Pendidikan Indonesia Press.
- Traboulsi, H. (2009). Arabic Named Entity Extraction: A Local Grammar Based Approach. *Proceedings of the International Multiconference Science and Information Technology* (pp. 139-143). Mragowo, Poland: Polish Information Processing Society.
- Wicaksono, A.-F., & Purwarianti, A. (2010). HMM Based Part-of-Speech Tagger for Bahasa Indonesia. *Proceeding of the Fourth International MALINDO Workshop (MALINDO2010)*. Jakarta: Malindo Press.

Komputasi Linguistik untuk Penyusunan Kamus Bahasa Daerah: Studi Kasus Bahasa Sunda

Dadan Sutisna

Yayasan Kebudayaan Rancagé
dadansutisna@yahoo.com

Abstrak

Usaha memantapkan dan menyebarluaskan bahasa daérah seyogianya dilaksanakan secara terencana dan berkesinambungan, bukan hanya melalui pendidikan dan pengajaran, melainkan juga melalui penulisan pelbagai jenis buku dalam bidang tata bahasa, ejaan, peristilahan, kesusastraan, dan penyusunan serta penerbitan kamus bahasa Sunda dengan berbagai format. Dengan terbitnya pelbagai jenis kamus, khazanah dan prestise bahasa daérah juga dikukuhkan dalam bidang léksikografi, sebab dalam kebudayaan cetak modérn, kamus merupakan alat dan lambang sekaligus martabat sebuah bangsa.

Di Indonésia, penyusunan kamus menggunakan perangkat lunak yang dibuat sendiri belum pernah ada. Sementara perangkat lunak yang dibuat oleh pihak luar tidak sepenuhnya memahami karakteristik bahasa daérah di Indonésia. Kamus dapat disusun dengan cepat, tepat, dan akurat, jika menggunakan perangkat lunak berbasis komputasi linguistik. Peranan komputasi linguistik sangat penting karena berhubungan dengan éféktifitas pencarian, analisis, pengolahan, dan pemaknaan suatu bahasa. Ini adalah tuntutan nyata untuk setiap bahasa, di mana manusia tidak hanya menginginkan bantuan dalam mekanik, tetapi juga dalam upaya intelektual.

Makalah ini akan menjelaskan upaya yang pernah dilakukan dalam penyusunan Kamus Utama Bahasa Sunda (KUBS) sebanyak 150.000 éntri. Kamus ini disusun menggunakan sebuah perangkat lunak yang dibuat oleh Yayasan Kebudayaan Rancagé, dan sudah menghasilkan kamus dalam bentuk cetak setebal lebih dari 6.000 halaman.

KUBS menggunakan lebih dari lima juta halaman dokumén berbahasa Sunda sebagai sumber, terdiri dari 2.000 buku Sunda dan belasan ribu média berbahasa Sunda. Metodologi dalam penyusunan kamus ini adalah

sebagai berikut: (1) Digitalisasi untuk memindahkan sumber-sumber cetak ke dalam format digital, dan disimpan dalam sebuah database; (2) Identifikasi dan validasi sumber; (3) Otomatisasi kalimat, yaitu pemecahan téks sumber menjadi kalimat; (4) Otomatisasi korpus berdasarkan téks yang tersimpan dalam database; (5) Penyusunan définisi oléh léksikograf menggunakan referénsi database; (6) Pengalihan hasil définisi ke dalam format ebook; (7) Pencetakan kamus.

Dengan demikian, semua lema, sublema, definisi, contoh kalimat, sumber rujukan, dan data-data lainnya disimpan secara sistematis dalam pangkalan data. Hal ini memudahkan dalam pencarian dan penyajian kamus, serta secara otomatis akan terbentuk versi kamus digital. Ke depan, kamus ini dapat dikembangkan menjadi kamus elektronik berbasis internet dan telepon pinyar, juga dibuat dalam bentuk aplikasi text to speech untuk dialek serta idiom bahasa daérah.

I. PENDAHULUAN

1.1 Latar Belakang

Usaha melestarikan dan menyebarluaskan bahasa daerah seyogianya dilaksanakan secara terencana dan berkesinambungan, bukan hanya melalui pendidikan dan pengajaran, melainkan juga melalui penulisan pelbagai jenis buku dalam bidang tata bahasa, ejaan, peristilahan, kesusastraan, dan penyusunan serta penerbitan kamus bahasa daerah dalam berbagai format. Dengan terbitnya pelbagai jenis kamus, khazanah dan prestise bahasa daerah juga dikukuhkan dalam bidang leksikografi, sebab dalam kebudayaan cetak modern, kamus merupakan alat dan lambang sekaligus martabat sebuah bangsa.

Penyusunan kamus yang lengkap dan komprehensif tak hanya membutuhkan leksikograf yang handal, tetapi harus ditunjang pula oleh metodologi yang tepat. Sebelum era komputer, kamus disusun menggunakan sistem kartu. Dan saat ini, meskipun teknologi informasi telah merambah berbagai bidang, peranan komputer belum optimal dalam leksikografi. Komputer baru digunakan untuk pengurutan, pencarian, dan penyimpanan data kamus, sementara kajian analisis kebahasaannya masih terabaikan.

Selain itu, kendala dalam penyusunan kamus terdapat pada proses pemilihan kata-kata yang akan dimasukkan menjadi entri kamus. Selama ini, entri kamus seolah-olah ditentukan oleh ahli bahasa, sehingga mungkin saja ada kata-kata yang dibuang. Padahal tugas leksikograf adalah mencatat kata-kata, bukan menentukan kata yang paling tepat. Dalam penyusunan kamus bahasa asing, masalah seperti ini sudah teratasi sejak satu setengah abad silam. Simon Winchester menjelaskan bahwa kamus adalah “inventaris

bahasa” dan sama sekali bukan panduan untuk menggunakan bahasa itu dengan baik. Tugas para penyusunnya bukanlah memilah-milah kata yang akan dimasukkan berdasarkan apakah kata itu bagus atau jelek (2007:155).

1.2 Masalah

Penyusunan kamus bahasa daerah umumnya menghadapi kendala dalam pengumpulan kata-kata yang akan dijadikan lema atau entri kamus. Untuk menerbitkan sebuah kamus yang lengkap memerlukan waktu belasan sampai puluhan tahun—seperti yang terjadi pada bahasa Sunda. Akibat dari lamanya proses tersebut, kamus tidak lagi mencerminkan bahasa kekinian, karena kata-kata suatu bahasa telah berubah atau bertambah setelah kamus tersebut terbit, sehingga hanya memuat entri yang disusun bertahun-tahun silam. Masalah lain dalam pengembangan bahasa daerah—juga bahasa Indonesia, para ahli belum memanfaatkan ilmu komputer secara maksimal. Dukungan komputer dalam penyusunan kamus, hanya sebatas hal-hal dasar seperti penyimpanan dan pengurutan kata-kata. Sementara analisis linguistiknya belum berbasis komputerisasi.

Setiap bahasa memiliki keunikan tersendiri, sehingga aplikasi komputer untuk mengolah bahasa tersebut harus dibuat khusus. Saat ini memang cukup banyak aplikasi komputer yang berkaitan dengan linguistik, tetapi sebagian besar untuk bahasa asing, dan tentu tidak optimal jika digunakan untuk bahasa-bahasa daerah.

Di Indonesia, bidang komputasi linguistik belum begitu mendapat perhatian. Padahal Indonesia merupakan negara dengan bahasa terbanyak kedua setelah Papua Nugini. Selain bahasa Indonesia sebagai bahasa nasional, terdapat ratusan bahasa ibu yang tersebar dari Sabang sampai Merauke. Kendala yang paling nyata ialah karena komputasi linguistik merupakan bidang ilmu mutidisipliner, yang memerlukan pemahaman dalam bahasa dan kepiawaian dalam ilmu komputer (informatika). Tidak banyak yang menguasai kedua bidang ini secara bersamaan, sementara kita juga belum bisa menyatukan ahli bahasa dengan ahli komputer untuk bekerjasama.

1.3 Tujuan

Tujuan pengembangan komputasi linguistik untuk bahasa daerah antara lain:

1. Meningkatkan fungsi bahasa daerah dalam bidang teknologi informasi.
2. Menyediakan rujukan atau pegangan baku, terutama yang sejalan dengan bahasa daerah yang hidup (*living language*) di tengah masyarakat, dalam rangka memelihara, melestarikan dan mengembangkan bahasa daerah.
3. Menggugah dan mengembangkan kebanggaan masyarakat terhadap bahasa dan budayanya sendiri sebagai sumber identitas dan kreativitas bersama dalam upaya menghadapi berbagai perubahan dalam kehidupan bersama di kemudian hari.
4. Memelihara dan melanjutkan keragaman bahasa (*language diversity*) dan keragaman budaya (*cultural diversity*) di Indonesia, terutama sehubungan dengan

adanya gejala di tingkat global yang menunjukkan bahwa keragaman tersebut cenderung menyusut.

5. Mengukuhkan dan memantapkan bahasa daerah sebagai bahasa ilmu pengetahuan dan kebudayaan dalam makna yang seluas-luasnya.
6. Sebagai salah satu kekayaan bahasa yang dimiliki Bangsa Indonesia yang harus tetap dijaga keberadaannya sebagai bukti keberadaan Bangsa Indonesia yang besar dan beragam.

1.4 Tinjauan Pustaka

Hingga saat ini, menurut pengamatan penulis, belum ada penelitian yang secara khusus mengkaji komputasi linguistik dalam bahasa daerah sebagai rujukan awal penyusunan kamus. Sudiroatmadja (1993) dalam penelitiannya menyinggung tentang morfologi generatif dalam pembentukan kata bahasa Jawa melalui metode komputasi, tetapi penelitian tersebut sudah cukup lama serta tidak ada kaitannya dengan penyusunan kamus.

II. LANDASAN TEORI DAN METODE PENELITIAN

Sebagai menjadi mesin pencari nomor satu di Internet, Google menerapkan komputasi linguistik yang disempurnakan setiap saat. Tampilannya sederhana, tetapi di balik itu, dipenuhi algoritma kompleks berkaitan dengan bahasa, untuk memberikan hasil pencarian yang lebih mendekati keinginan pengunjung. Google menganalisa setiap teks yang diketik oleh pengunjung, mencocokkannya dengan indeks pada basis data, memperbarui kesalahan ejaan, dan berusaha memahami setiap bahasa yang digunakan oleh pengunjung. Kerumitan suatu bahasa ternyata sedikit demi sedikit dapat dipecahkan oleh komputer, dan Google melakukannya dengan membuat kelompok penerjemah terbesar di dunia.

Komputasi linguistik merupakan penggabungan antara ilmu bahasa dan ilmu komputer melalui pemodelan statistik berdasarkan aturan bahasa alami dari perspektif komputasi. Secara sederhana, komputasi linguistik dapat digambarkan sebagai upaya untuk mengajarkan komputer agar dapat berkomunikasi dengan manusia, memecah persoalan ambiguitas bahasa menjadi sesuatu yang realitis. Prinsip kerja komputer hampir mirip dengan otak manusia, maka para ahli psikolinguistik mulai mengkaji dan berusaha menciptakan sebuah sistem yang memungkinkan komputer berbahasa seperti manusia atau menerjemahkan bahasa manusia (Yendra 2016);

Menurut Hans Uszkoreit (2000), komputasi linguistik adalah disiplin antara linguistik dan ilmu komputer yang berurusan dengan pemodelan statistik berdasarkan aturan bahasa alami dari perspektif komputasi. Secara sederhana, komputasi linguistik dapat digambarkan sebagai upaya untuk mengajarkan komputer agar dapat berkomunikasi dengan manusia, memecah persoalan ambiguitas bahasa menjadi sesuatu yang realistik. Komputasi linguistik berhubungan dengan efektifitas pencarian, analisis, pengolahan,

dan pemaknaan suatu bahasa. Ini adalah tuntutan nyata untuk setiap bahasa, di mana manusia tidak hanya menginginkan bantuan dalam mekanik, tetapi juga dalam upaya intelektual. Di negara lain bidang ini telah menjadi agenda khusus dan diperbincangkan dalam suatu konferensi. Sejak tahun 2000, para ahli bahasa dan ahli komputer berkumpul dalam *Conference on Intelligent Text Processing and Computational Linguistics*.

Komputasi linguistik sangat cocok untuk membantu penyusunan kamus. Kamus merupakan sebuah buku referensi yang memuat kosa kata yang terdapat dalam sebuah bahasa, yang disusun secara alfabetis disertai keterangan bagaimana menggunakan kata itu (Keraf 2007). Buku yang dimaksud tak hanya dalam bentuk cetak, pada era teknologi kamus dapat pula disajikan secara elektronik.

Analisis komputer terhadap suatu bahasa, dapat dijadikan sebagai korpus elektronik, yaitu koleksi data kebahasaan dari berbagai teks tertulis atau transkripsi rekaman bahasa lisan yang disimpan secara elektronis dalam sebuah *database* (Kushartanti et al. 2007).

Metode dalam penelitian ini sebagai berikut:

1. Digitalisasi, yaitu pemindahan dokumen teks dan suara bahasa daerah ke dalam format digital.
2. Analisis dokumen, yaitu penelaahan secara komputerisasi terhadap dokumen yang telah disimpan dalam format digital.
3. Analisis linguistik, yaitu membuat algoritma khusus dalam bahasa komputer, agar komputer mengenal tata bahasa suatu bahasa daerah.
4. Penyusunan korpus elektronik, yaitu pengumpulan kata-kata dari hasil analisis linguistik serta dari berbagai dokumen hasil digitalisasi.
5. Penyusunan kamus, yaitu membuat definisi dari kata-kata yang telah dipilih dengan merujuk pada sumber yang tersedia.

III. ANALISIS PENELITIAN

3.1 Leksikografi Bahasa Sunda

Tradisi penyusunan kamus bahasa Sunda memang baru berlangsung sekitar satu setengah abad dan hampir semuanya ditulis dan disusun atas prakarsa pribadi. Sekurangnya ada empat macam kamus bahasa Sunda jika dilihat dari sudut bahasa pendampingnya, yaitu Kamus Sunda-Inggris, Kamus Sunda-Belanda, Kamus Sunda-Sunda, dan Kamus Sunda-Indonesia. Sedangkan ditinjau dari segi kronologi penerbitannya, maka *A Dictionary of the Sunda Language of Java* yang disusun oleh Jonathan Rigg (terbit pertama kali di Batavia tahun 1862), merupakan kamus bahasa Sunda dwibahasa yang tertua, yaitu Sunda-Inggris. S. Coolsma menerbitkan Kamus Sunda-Belanda berjudul *Soendaneesch-Hollandsch Woordenboek* (1884, 1913, 1930) dan Belanda-Sunda berjudul *Hollandsch-Soendaneesch Woordenboek* (1910), kemudian Rd. Satjadibrata menerbitkan kamus ekabahasa berjudul *Kamus Basa Sunda* (1948, 1954, 2005), dan kamus dwibahasa berjudul *Kamus Sunda-Indonesia* (1952), F.S. Eringa menyusun kamus Sunda-Belanda

berjudul *Soendaas-Nederlands Woordenboek* (1984), Raden Rabindranat Hardjadibrata menyusun kamus Sunda-Inggris berjudul *Sundanese-English Dictionary* (2003), dan selama empat puluh tahun R. A. Danadibrata menyusun *Kamus Basa Sunda* (1930-1970), dan baru bisa diterbitkan pada tahun 2006. Tampaknya, hingga hari ini hanya ada satu kamus basa Sunda yang disusun oleh sebuah tim atas prakarsa Lembaga Basa jeung Sastra Sunda, yaitu *Kamus Umum Basa Sunda* (1975, 1980, 1985).

Semua kamus bahasa Sunda yang disebutkan di atas merupakan kamus yang sudah ketinggalan zaman, banyak kata-kata baru baik yang berasal dari bahasa asing dalam pelbagai bidang keilmuan yang belum terdokumentasikan, padahal bahasa Sunda merupakan bahasa yang masih akan berkembang pesat sesuai dengan tingkat kemajuan ilmu pengetahuan dan masyarakat pemakai bahasa itu sendiri.

3.2 Komputasi Linguistik Bahasa Sunda

Pada tahun 2013, Yayasan Kebudayaan Rancage menetapkan dan mengukuhkan Tim Penyusun Kamus Utama Bahasa Sunda. Ini adalah kamus pertama yang disusun menggunakan metodologi komputasi linguistik.

Komputasi linguistik dalam bahasa Sunda, baru dibangun pada tahun 2013 dengan nama *Sundaling*. Kemampuan dari perangkat lunak ini antara lain:

1. Melakukan analisis terhadap fonologi dalam bahasa Sunda.
2. *Steeming* bahasa, yaitu otomatisasi pencarian kata dasar.
3. Otomatisasi kata turunan dari sebuah kata dasar, serta kemampuannya untuk membandingkan kata-kata tersebut dengan berbagai dokumen.
4. Media penyimpanan berbagai dokumen kebahasaan dalam bentuk teks, gambar, dan suara.
5. Otomatisasi kalimat, yaitu memecah sebuah halaman teks menjadi beberapa kalimat.
6. Otomatisasi kata, yaitu memecah kalimat menjadi daftar kata-kata.
7. Membuat korpus untuk kamus.
8. Sumber referensi dalam penyusunan kamus.

3.3 Digitalisasi

Perkembangan bahasa Sunda dapat diamati dengan adanya penambahan kata-kata yang digunakan oleh masyarakat, baik dalam percakapan lisan maupun tulis. Dengan demikian, penyusunan kamus tak hanya merujuk pada kamus yang terbit sebelumnya, tetap harus dapat merangkum kata-kata baru dari berbagai bidang keilmuan. Kegiatan penelusuran bibliografi bertujuan untuk mendapatkan sumber-sumber terpercaya sebagai rujukan penyusunan kamus, yaitu buku-buku berbahasa Sunda, media berbahasa Sunda, serta rekaman percakapan lisan. Adapun bibliografi yang dimaksud antara lain:

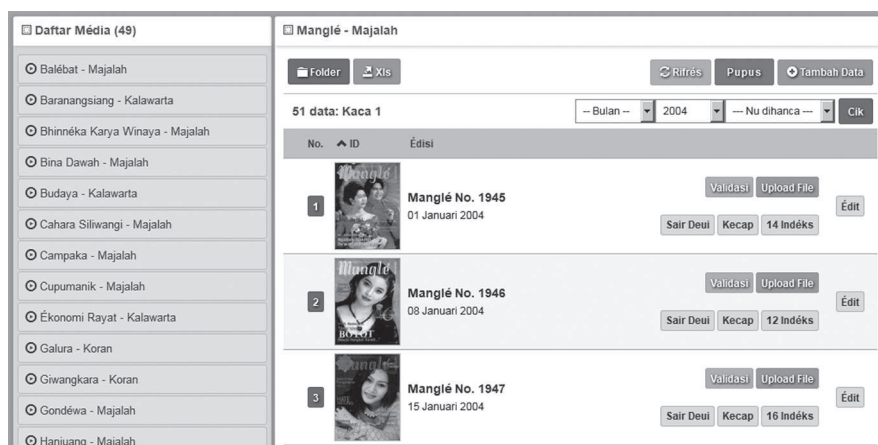
1. Buku-buku bahasa Sunda. Lebih dari 2.000 judul buku bahasa Sunda yang dipindahkan ke dalam format digital sebagai sumber kamus.

- Media cetak berbahasa Sunda, mencakup berbagai media yang terbit dalam satu abad terakhir, antara lain *Koran Sipatahoenan*, *Majalah Gondéwa*, *Majalah Hanjuang*, *Majalah Manglé*, *Majalah Cupumanik*, *Koran Galura*, *Majalah Sunda Midang*, *Tabloid Kujang*, *Koran Sunda*, *Majalah Seni Budaya*, *Majalah Suara Daérah*, *Koran Tribun Jabar* (édisi sastra Sunda), *Koran Pikiran Rakyat* (édisi sastra Sunda), *Koran Priangan* (édisi sastra Sunda), *Majalah Ujung Galuh*, *Majalah Langlang Budaya*, *Majalah Bhinneka Karya Winaya*, *Kalawarta Kudjang*, *Majalah Salaka*, *Tabloid Swara Cangkurileung*, *Koran Giwangkara*, *Tabloid Mandiri*, *Majalah Bina Dawah*, *Majalah Turus*, *Majalah Cahara Siliwangi*, *Majalah Iber*, dan *Majalah Balébat*.
- Data lisan, antarana lain dari cerita pantun yang dibacakan kemudian direkam.

Hasil dari penelusuran bibliografi masih berupa naskah dan dokumen dalam *hardcopy*. Sumber tersebut harus dipindahkan ke dalam format digital untuk memudahkan pengolahan data. Meski didukung oleh perangkat teknologi terkini, proses pemindaian sumber memerlukan waktu yang cukup lama karena harus mengalihkeun ratusan ribu halaman ke dalam format gambar. Adapun perangkat yang digunakan antara lain:

- Flatbed Scanner*, untuk memindahkan naskah buku dan majalah dengan ukuran kertas maksimal A4.
- Laser Scanner*, untuk memindahkan sumber dengan format kertas yang cukup besar, misalnya koran.
- Kamera digital*, untuk memotret dokumen lainnya.

Semua hasil pemindaian kemudan dikumpulkan dalam file format JPG, dan disusun berdasarkan halaman untuk masing-masing sumber. Karena hasil pemindaian masih dalam format gambar, langkah selanjutnya adalah mengindéntikasi teks dalam gambar telah dipindai, menggunakan teknologi *Optical Character Recognition* (OCR). Aplikasi yang digunakan antara lain ABBYY FineReader versi 11. Teks hasil OCR kemudian dialihkan ke dalam berkas format TXT.



Gambar 1. Contoh dokumen dari media massa yang sudah dipindahkan ke dalam database.

Data Buku Sunda						
240 data: Kaca 1		Filter: Kiblat Buku Utama		Taun: --- Nu dihanca: ---		
No.	ID	Judul	Penerbit	Taun		
1		A Dictionary of The Sunda Language of Java Jonathan Rigg Kamus	Kiblat Buku Utama	2009	Upload 0 Indeks	Edit +Edisi
2		Ajalna Sang Bentang Pilem ✓ Duduh Durahman Carita Pondok	Kiblat Buku Utama	2004	Validasi Upload Sair 0 Indeks	Edit +Edisi
3		Ajar Kutamanggu ✓ Yoseph Iskandar Roman/Novel	Kiblat Buku Utama	2012	Validasi Upload Sair 0 Indeks	Edit +Edisi
4		Al-Hikmah Tarjamah Al-Qur'an Basa Sunda Juz 30 M. Maryati Agama	Kiblat Buku Utama	2009	Upload 0 Indeks	Edit +Edisi
5		Al-Hikmah, Tarjamah Al-Qur'an Basa Sunda Juz 1 M. Maryati Agama	Kiblat Buku Utama	2009	Upload 0 Indeks	Edit +Edisi

Gambar 2. Contoh dokumen dalam bentuk buku yang sudah dipindahkan ke dalam database.

3.4 Fonologi Bahasa Sunda

Secara umum, komputer tidak mengenal fonem, tetapi hanya mengenali sebuah aksara berdasarkan bentuk (grafem). Saat ini bahasa Sunda menggunakan aksara Latin yang mengandung 7 vokal dan 25 konsonan. Dengan demikian, perlu dibuat sebuah sistem agar komputer mengenal tata suara bahasa Sunda.

Tabel 1 menunjukkan susunan huruf yang digunakan dalam bahasa Sunda dan tabel 2 adalah contoh penggunaannya.

Vokal	/i/, /eu/, /u/, /é/, /e/, /o/, /a/
Konsonan	/b/, /c/, /d/, /f/, /g/, /h/, /j/, /k/, /l/, /m/, /n/, /ny/, /ng/, /p/, /q/, /r/, /s/, /t/, /v/, /w/, /x/, /y/, /z/

Tabel 1. Susunan huruf dalam bahasa Sunda

Contoh Kata	<i>nyeungseung</i>
Konsonan	/n/, /y/, /e/, /u/, /n/, /g/, /s/, /e/, /u/, /n/, /g/ (11)
Pola Grafém	KKVVKKKKVVKK
Susunan Foném	/ny/, /eu/, /ng/, /s/, /eu/, /ng/ (6)
Pola Foném	KVKKVK
Lambang Foném	ñöŋsöŋ

Tabel 2. Contoh penggunaan foném jeung grafem.

3.5 Afiksasi

Komputasi linguistik dapat menyusun algoritma untuk membuat semua kata turunan dari sebuah kata dasar. Bahasa Sunda memiliki kemungkinan lebih dari 200 kata turunan, meski tidak semua kata tersebut digunakan.

Dengan mengetik sebuah kata dasar, komputer dapat mengeluarkan semua kata turunan serta menghubungkannya dengan berbagai sumber teks. Gambar 3 merupakan contoh afiksasi sebuah kata dasar disertai frekuensi penggunaannya.

Kecap: <input type="text" value="adu"/>		<input type="button" value="Pidangkeun"/>	
1. aduan ✓	-an	100+	<input type="button" value="Cek"/>
2. aduaneun ✓	-an-eun	2	<input type="button" value="Cek"/>
3. aradu ✓	-ar-	5	<input type="button" value="Cek"/>
4. adueun ✓	-eun	5	<input type="button" value="Cek"/>
5. inadu	-in-	1	<input type="button" value="Cek"/>
6. adukeun ✓	-keun	52	<input type="button" value="Cek"/>
7. aduna ✓	-na	75	<input type="button" value="Cek"/>
8. umadu	-um-	0	<input type="button" value="Cek"/>
9. baadu	ba-	0	<input type="button" value="Cek"/>
10. diadu ✓	di-	100+	<input type="button" value="Cek"/>
11. diaduan ✓	di-an	10	<input type="button" value="Cek"/>
12. diaduanan	di-an-an	0	<input type="button" value="Cek"/>
13. diaradu ✓	di-ar-	63	<input type="button" value="Cek"/>
14. diaraduan	di-ar-an	0	<input type="button" value="Cek"/>
15. diaradukeun ✓	di-ar-keun	3	<input type="button" value="Cek"/>
16. diadukeun ✓	di-keun	100+	<input type="button" value="Cek"/>

Gambar 3. Contoh afiksasi dari kata “adu”.

3.6 Penyusunan Korpus

Melalui komputasi linguistik, korpus dapat disusun secara cepat dan otomatis. Komputer dapat memilih dan memilah kata-kata dalam sebuah buku setebal 100 halaman hanya beberapa detik. Sampai saat ini, komputasi linguistik bahasa Sunda telah menyimpan lebih dari 45 juta kata dalam 10 juta kalimat. Adapun kata-kata unik mencapai 450 ribuan. Data ini dapat dijadikan sumber dalam penyusunan kamus dengan mengacu pada frekuensi penggunaannya dalam buku atau dokumen lainnya.

<div>Statistik Kalimah</div> <div><input type="button" value="Indeks Kalimah"/></div> <div>10.211.683 Kalimah Disair</div> <div>548.242Kaca</div> <div>45.230.034 Jumlah Kecap</div> <div>3.651 Kaca Can Disair</div> <div>19.163 Kaca Diindeks</div> <div>129.079 Kaca Can Diindeks</div>	<div>Pilari Kalimah</div> <div><input type="text" value="Kabéh Média..."/> <input type="button" value="Pilari"/></div> <div>Taun <input type="text"/></div>
--	---

Gambar 4. Rekapitulasi jumla kata dan kalimat yang tersimpan dalam komputer.

Seperti halnya Google, sebuah kata dapat pula dicari secara manual dengan menelusuri jutaan indeks yang sudah tersimpan dalam database.

Snippet	Source	Year
Bari ngasuh murangkalih bari ngageulis nyayagikeun ngabageakeun caroge nu mulih cape tugas.	(2395:31) <i>Manglé</i> No. 1272	1990 • kaca 31
Buktosna? Teu sakara-kara Mamah dangdos, ngageulis , ngaluis, upami Mamah teu bade angkat-angkatan.	(2592:32) <i>Katumbiri</i> No. 4	1986 • kaca 32
Dangdak dengdek, sagala diusapan, malum nu keur ngageulis .	(2504:52) <i>Manglé</i> No. 2437	2013 • kaca 52
Perjoangan neangan jodo ku jalan ngageulis , miara awak jeung sajabana geus lain mangsana deui.	(2528:31) <i>Manglé</i> No. 2460	2014 • kaca 31
Malah teu saeutik asup mangsa murohek mah, mangsa hayang ngageulis jeung ngasep, apan awak teh beuki dipiara.	(2535:15) <i>Manglé</i> No. 2468	2014 • kaca 15
Sok sanajan ngageulis sataker kebek, béang déngkak meruhkeun wedak.	(24:32) <i>Cupumanik</i> No. 4	2003 • kaca 32
Boh nu hideung boh nu bodas, boh nu geulis bawana ngajadi boh nu ukur geugeulisan, lebah ngageulis mah sigana sarua baé matak hé-meng ngabandunganana.		

Gambar 5. Contoh hasil pencarian kata “ngageulis”.

3.7 Koreksi Ejaan

Komputasi linguistik bahasa Sunda juga menyediakan fasilitas untuk mengoreksi ejaan dari sebuah sumber teks. Prinsip kerjanya adalah komputer akan membagi teks menjadi daftar kata-kata, kemudian membandingkan dengan korpus dan dokumen digital yang tersedia. Kata yang dianggap keliru atau belum terdapat dalam koleksi korpus, akan ditandai.

Heuleut sawatara jongjongan, katémbong aya nu turun, kawas timbaan nu diulur ti luhur. Manéhna sok néngo alat kawas kitu téh di gedong-gedong jangkung, paranti pagawé meresihan kaca. Jrut lalaki tiluan. Nu hiji ngora kénéh, dangdanan raphi, maké kaméja jeung didasi sagala. Nu duaan deui, dedeganaa seseg tur perenges, sidik éta mah nu méméhna datang ka balé désa. Kadés Jabrig nangtung. Nyarandé kana témbok. Buukna ngarumbay meuntas taktak.

Koreksi

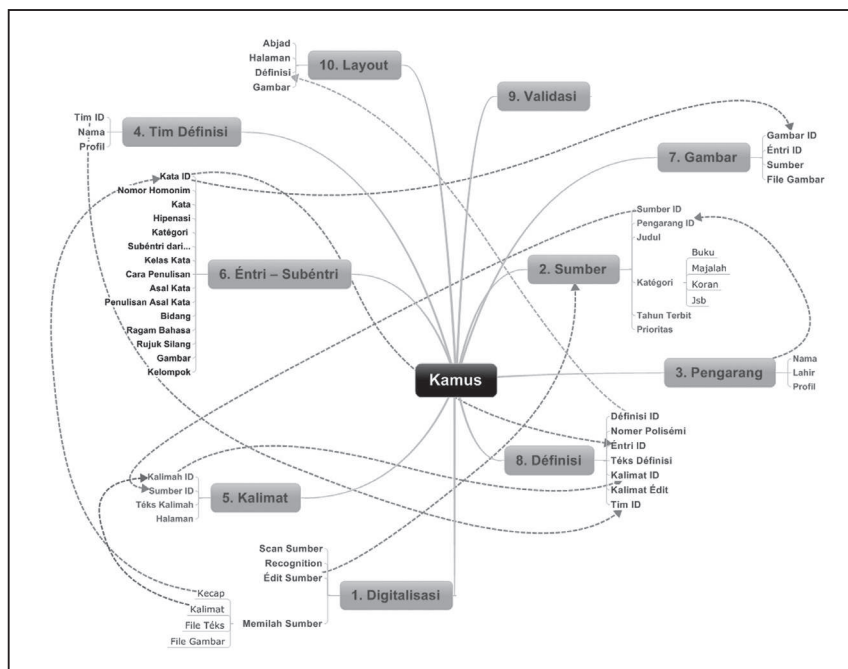
heuleut sawatara jongjongan katémbong aya nu turun kawas timbaan nu diulur ti luhur manéhna sok néngo alat kawas kitu téh di gedong-gedong jangkung paranti pagawé meresihan kaca jurut lalaki tiluan nu hiji ngora kénéh dangdanan raphi maké kaméja jeung didasi sagala nu duaan deui dedeganaa seseg tur perenges sidik éta mah nu méméhna datang ka balé désa kadés jabrig nangtung nyarandé kana témbok buukna ngarumbay meuntas taktak

Gambar 6. Contoh hasil koreksi dari sebuah teks.

3.8 Penyusunan Kamus

Secara umum, langkah-langkah dalam penyusunan kamus bahasa Sunda berbasis komputasi linguistik adalah sebagai berikut:

1. Digitalisasi, yaitu proses pemindahan dokumen teks dan suara ke dalam format digital. Dokumen tersebut berupa buku, majalah, koran, dan rekaman dalam kurun satu abad terakhir. Saat ini, dokumen yang sudah tersimpan lebih dari 2.000 buku dan puluhan ribu halaman media massa.
2. Identifikasi sumber, yaitu proses pencatatan identitas sumber, mencakup tahun penerbitan, judul dokumen, penulis, dan sebagainya. Melalui tahap ini, dapat ditelusuri sejak kapan sebuah kata mulai digunakan.



Gambar 7. Alur penyusunan Kamus Utama Bahasa Sunda

3. Pencatatan nama pengarang/penulis, untuk menelusuri siapa saja yang pernah menggunakan sebuah kata, terutama untuk kata-kata yang sudah jarang digunakan.
4. Membentuk tim definisi, yaitu leksikograf yang akan membantu membuat definisi lema dan sublema.
5. Pembuatan *database* kalimat, yaitu menyimpan kalimat dari setiap sumber. Saat ini telah terkumpul lebih dari 10 juta kalimat.
6. Penelusuran entri dan subentri, yaitu otomatisasi korpus kamus berdasarkan sumber-sumber yang tersedia. Pada tahap ini, perangkat lunak akan memecah setiap sumber menjadi daftar kata-kata, disertai dengan informasi dari mana

7. Menentukan entri yang perlu disertai dengan gambar, sebagai upaya untuk memudahkan pembaca dalam memahami arti dari sebuah kata.
8. Leksikograf menyusun definisi berdasarkan entri dan subentri yang telah ditentukan sebelumnya. Untuk memudahkan pembuatan definisi, perangkat lunak akan menampilkan puluhan contoh kalimat dari entri tersebut, serta rujukan ke kamus-kamus lain termasuk kamus bahasa asing.
9. Tahap validasi, yaitu mengaji ulang setiap entri dan definisi.
10. Pembuatan tata letak kamus, yaitu otomatisasi hasil definisi ke dalam format buku sehingga menghasilkan kamus dalam bentuk elektronik yang siap dicetak.

Gambar 8 menunjukkan hasil akhir dari kamus yang telah diberi définisi, siap dicetak, dijadikan *e-book* atau dibuat kamus versi elektronik.

é.cék¹ 299 é.cék²

é.cag-é.cag *kp* kakara écag: Régeyeng, jongjon ngahenen anu ngajarni, écag-écag geus anjog ka rambu leuweung (A. Bakri, Sanghiyang Lutung Kasarung, 1976, k. 9);

é.cag.keun *kp* turunkeun; eureunkeun: *Tau nanaon Kresna milu perang, asal écagkeun heula gelar kabataaran* (Antologi, Kanagan 2, 2008, k. 170);

mang.é.cag.keun *kp* mantuan ngécagkeun ka nu lian: *Si Ayi mangécagkeun Si Nyai ti nu ngagandogna* (Godi Suwarna, *Sandékala*, k. 142);

nga.ré.cag.keun *kp* papada ngécagkeun: *Keur di cai kénéh daratang baturna tiluan, areureun handapeun linus, ngaréccagkeun tanggungan* (Moh. Ambri, *Burak Siluman*, k. 19);

ngé.cag.keun *kp* 1 neundeun; nunda: *Ké, ké, ké, cék sémah téh bari ngécagkeun piring* (A. Bakri, *Kabandang ku Kuda Lumping*, 1969, k. 53); 2 sina eureun; ninggalkeun: *Kuring ngécagkeun pagawéan* (Sjarif Amin, *Nyi Haji Saonah*, k. 30);

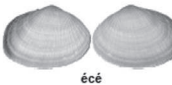
é.cék¹, ngé.cé *kb* sindir sampir bari ngahina: *Karasa aya nada ngécé tina sorana* (Antologi, Kanagan, 2003, k. 343);

di.é.cé *kp* disindirani; dihina: *Nu kaopat henteu sahijueun da sahenteuna singa mah sato galak, tapi kalah ka diécé ku dulur-dulurna* (Ajjp Rosidi, *Gerakan Kasundam*, k. 61);

pang.é.cé *kb* panghina: *Wiyahna gelar dipépéndé pangécé* (Étti RS, *Maug Bayangan*, k. 19);

pang.é.cé.an *kb* panghina: *Na haté patéép nanté-ranté pangécéan* (Ajjp Rosidi, *Sajak Sunda*, k. 545);

é.cék² *kb* *Asaphis deflorata* sabangsa haremis, hirupna di walungan, ngasakanana ditumis at. digoréng;



écé

Gambar 8. Contoh hasil definisi.

IV. SIMPULAN

Melalui komputasi linguistik, penyusunan kamus dapat lebih lengkap, cepat, dan akurat. Hal ini terbukti dalam penyusunan Kamus Utama Bahasa Sunda yang dilakukan kurang dari tiga tahun, dan bisa menghasilkan lebih dari 150.000 entri. Contoh kalimat pada kamus juga akan merujuk pada sumber-sumber yang tersedia, sehingga dapat memperkaya keragaman gaya bahasa. Penyebarannya kamus pun menjadi lebih luas

karena secara otomatis akan tercipta sebuah kamus digital, yang dapat dibuka melalui berbagai perangkat teknologi informasi.

Perangkat lunak yang dibuat untuk penyusunan kamus bahasa Sunda, dapat digunakan pula untuk bahasa daerah lainnya, meski tidak setiap bahasa daerah memiliki sumber-sumber rujukan yang lengkap. Tetapi setidaknya, ini adalah upaya untuk mengatasi berbagai masalah leksikografi bahasa daerah, melalui metodologi yang lebih canggih dan profesional.

Daftar Pustaka

- Clark, Alexander et al. 2010. *The Handbook of Computational Linguistics and Natural Language Processing*. Blackwell Publishing. Singapore: 2010.
- Coolsma, S. 1904. *Soendaneesche Spraakkunst*. Leiden: A.W. Sijthoff.
- Coolsma, S. 1910. *Hollandsc –Soendaneesch Woordenboek*. Leiden: A.W. Sijthoff's Uitgevers – Maatschappij.
- Coolsma, S. 1913. *Soendaneesch – Hollandsch Woordenboek*. Leiden: A.W. Sijthoff's Uitgevers – Maatschappij.
- Hardjadibrata, R.R. 2003. *Sundanese English Dictionary*. Jakarta: Dunia Pustaka Jaya.
- Kerap, Gorys. 2007. *Diksi dan Gaya Bahasa*. Jakarta: Gramedia Pustaka Utama.
- Kushartanti et al. 2007. *Pesona Bahasa: Langkah Awal Memahami Linguistik*. Jakarta: Gramedia Pustaka Utama.
- Rigg, Jonathan. 1862. *A Dictionary of the Sunda Language of Java*. Batavia: Lange & Co.
- Satjadibrata, R. 2008. *Kamus Basa Sunda*. Bandung: Kiblat Buku Utama.
- Sudiroatmadja, M.H. "Morfologi Generatif Komputasi Pembentukan Kata Bahasa Jawa". dalam PELLBA 7 halaman 213-254. 1993.
- Sutisna, Dadan. "Komputasi Kamus Sunda", *Cupumanik* No. 106, Taun IX, No.10, Mei 2012.
- Uszkoriet, Hans. 2000. "What Is Computational Linguistics?". http://www.coli.uni-saarland.de/~hansu/what_is_cl.html diakses 6 Mei 2016.
- Yendra. 2016. *Mengenal Ilmu Bahasa (Linguistik)*. Yogyakarta: Deepublish.
- Winchester, Simon. 1998. *The Professor dan the Madman*. Jakarta: Serambi Ilmu Semesta.

Penelitian Korpus Leksikografi: Aplikasi Fonetik Akustik Berkomputer

Hajah Norati Bakar

Dewan Bahasa dan Pustaka, Brunei Darussalam
norati_bakar@yahoo.com.sg

Abstrak

Informasi fonetik merupakan antara komponen yang signifikan dalam kajian leksikografi dan dipaparkan dalam struktur mikro kamus (microstructure). Tradisi dan konvensi yang berlaku zaman-berzaman dalam kajian leksikografi mengaplikasikan metode auditori pada deskripsi fonetik terhadap data yang hendak dikumpulkan. Jika metode ini sahaja masih mahu diteruskan, hasil kajiannya sangatlah subjektif dan kebenarannya tidak akan dapat dibuktikan secara objektif. Oleh hal yang demikian, hasil catatan auditori tersebut perlu diuji secara saintifik, dengan menggunakan kaedah fonetik akustik berkomputer. Analisis akustik berkomputer yang dibentangkan dalam kajian ini menggunakan dua perisian komputer, iaitu (i) Sound Filing System (SFS) + WASP for Windows OS dan (ii) Wavesurfer (WS) for Windows OS. Dalam analisis akustik berkomputer, analisis bentuk gelombang bunyi (wave form), pola formant (formant pattern) dalam spektrogram, surihan tekanan bunyi (energy trace), dan surihan gegaran bunyi (noise trace), misalnya, berguna untuk memastikan kehadiran bunyi- bunyi bahasa yang dikaji. Hasilnya sekali gus membolehkan pengkaji menentukan ketepatan sesuatu lambang fonetik IPA sebagai pegangan. Bagi keperluan kajian leksikografi, kertas kerja ini mengemukakan beberapa contoh fonetik (dialek Tutong dan dialek Belait) yang diperoleh hasil eksperimen fonetik akustik berkomputer. Pertama, mendeskripsikan kehadiran bunyi aspirat, bunyi hentian glotis, bunyi pembibir (labialization), bunyi geluncuran (gliding sound), atau kehadiran penglelangit-kerasan (palatalization) dapat dibuktikan oleh aplikasi WS secara kebalikan. Kedua, memperkenalkan kaedah fonetik akustik berkomputer sebagai salah satu kaedah fonetik eksperimental yang mendedahkan fakta bahasa tanpa terikat kepada teori-teori fonetik yang tertentu. Penelitian semula terhadap kajian-kajian lepas (yang tidak menggunakan kaedah fonetik akustik berkomputer) belum

dikerjakan secara meluas oleh pengkaji- pengkaji generasi baru. Kritikan terhadap kajian-kajian lepas akan ditekankan untuk memahami kelemahan dan kesalahfahaman yang masih berterusan dalam kalangan akademik dan juga untuk membetulkannya untuk kepentingan kita bersama.

Kata Kunci: bentuk gelombang bunyi, bunyi aspirat, bunyi hentian glotis, bunyi pembibiran, bunyi penglelangit-kerasan, fonetik akustik berkomputer, pola formant, spektrogram, surihan gegaran bunyi, surihan tekanan bunyi.

RESEARCH ON LEXICOGRAPHY: An Application of Computer-aided Acoustical Phonetics

Abstract

Phonetic information is among the significant components in the study of lexicography and indicated in the microstructure of the dictionary. Traditions and conventions employed for ages in the study of lexicography have been applying the auditory method on the description of the data to be collected. If only this method is still insisted, the study results are very subjective and the truth will not be able to be proved objectively. As such, the results of the auditory records need to be tested scientifically, using a method of computer-aided acoustical phonetics. The analyses of phonetic data presented in this research is to use two computer software, namely (i) Sound Filing System (SFS) + WASP for Windows OS and (ii) Wavesurfer (WS) for Windows OS. In the computer-aided acoustical analyses, the analyses of sound waves, formant patterns in the spectrogram, energy trace, and noise trace, for example, are useful to ensure the presence of the sounds of language studied. At the same time, the results allow researchers to determine the accuracy of the IPA phonetic symbols as an authoritative. For the purposes of lexicography study, this paper suggests some phonetic examples (Tutong and Belait dialects) obtained through the experiments of acoustical phonetics. First, to describe the presence of aspirate sound, glottal stop, labialization, gliding sound, or the presence of palatalization that can be proved by the WS in its reverse function. Secondly, to introduce the computer-aided acoustical phonetics as a method of experimental phonetics that reveals the linguistic facts without being bound to a certain phonetic theory. Reviews of previous studies (that did not use a method of computer-aided acoustical phonetics) have not been widely carried out by our new

generation of researchers. Criticism of these studies will be emphasized in order to understand the weakness and the misunderstanding that still persist in our academic circle and also to correct them for our mutual benefit.

Keywords: aspirate, computer-aided acoustical phonetics, energy trace, formant patterns, glottal stop, labialization, noise trace, palatalization, sound waves, spectrogram.

I. PENDAHULUAN

Korpus fonetik projek leksikografi dialek-dialek tempatan secara tradisinya bersumberkan kajian lapangan berasaskan kajian auditori. Hasil kajian auditori sangatlah subjektif dan tidak dibuktikan secara objektif. Oleh hal yang demikian, hasil catatan auditori tersebut perlu diuji secara saintifik, dengan menggunakan kaedah fonetik akustik berkomputer. Kajian ini mengaplikasi gabungan perisian komputer, iaitu Sound Filing System (SFS) + WASP for Windows OS dan Wavesurfer (WS) for Windows OS . Aplikasi perisian ini dikesan sangat bermanfaat dalam penelitian korpus fonetik yang autoritatif dalam leksikografi. Secara ringkasnya kedua-dua perisian ini merupakan kajian fonetik akustik, iaitu mengenai transmisi bunyi bahasa melalui udara daripada penutur kepada pendengar yang melibatkan analisis gelombang tuturan dan spektrogram.

II. MASALAH KAJIAN

Bagi keperluan kajian leksikografi, kertas kerja ini mengemukakan beberapa contoh fonetik (dialek Tutong dan dialek Belait) yang diperoleh hasil eksperimen fonetik akustik berkomputer. Kajian ini memfokuskan penelitian Pertama, mendeskripsikan kehadiran bunyi aspirat, bunyi hentian glotis, bunyi pembibiran (labialization), bunyi geluncuran, atau bunyi penglelangit-kerasan (palatalization) yang dapat dibuktikan oleh aplikasi WS secara kebalikan. Kedua, memperkenalkan kaedah fonetik akustik berkomputer sebagai salah satu kaedah fonetik eksperimental yang mendedahkan fakta bahasa tanpa terikat kepada teori-teori fonetik yang tertentu.

III. KAJIAN LEPAS

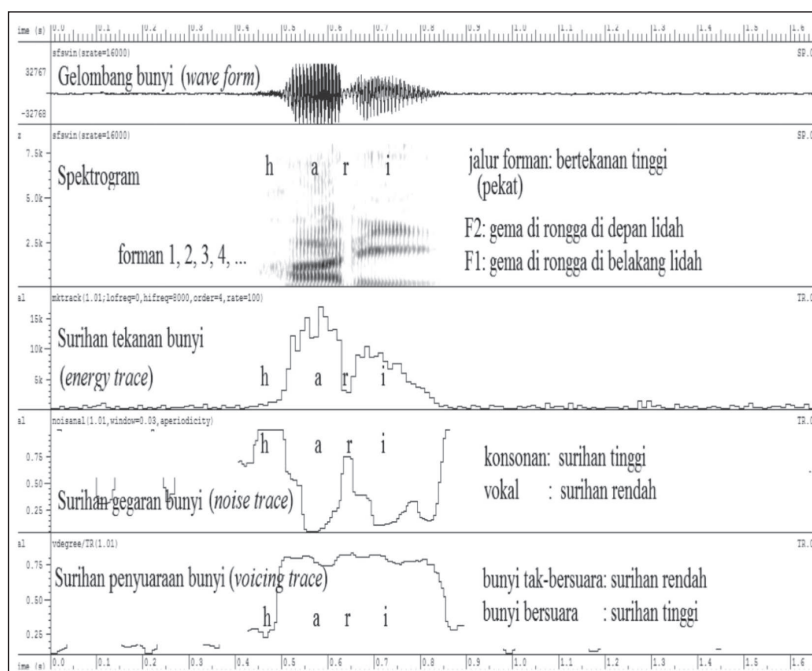
Dalam dialek Tutong dan dialek Belait, persebaran bunyi-bunyi tersebut sering ditampilkan dalam kajian-kajian lepas oleh para pengkaji dialek Tutong dan dialek Belait yang terdiri daripada ahli leksikografi tempatan dan pengkaji individu. Penelitian

bunyi aspirat dan hentian glotis dalam dialek Tutong dan dialek Belait sangat penting kerana kedua-dua bunyi tersebut merupakan pasangan minimal dalam sepasang kata yang tertentu dan dikesan sebagai unsur yang mempunyai sifat distingtif. Antara kajian lepas yang mendeskripsikan fonetik bahasa-bahasa daerah di negara ini ialah dialek Tutong ialah Kamus Melayu–Tutong - Tutong–Melayu (DBP, 1991), Struktur Bahasa Tutong (Ramlee

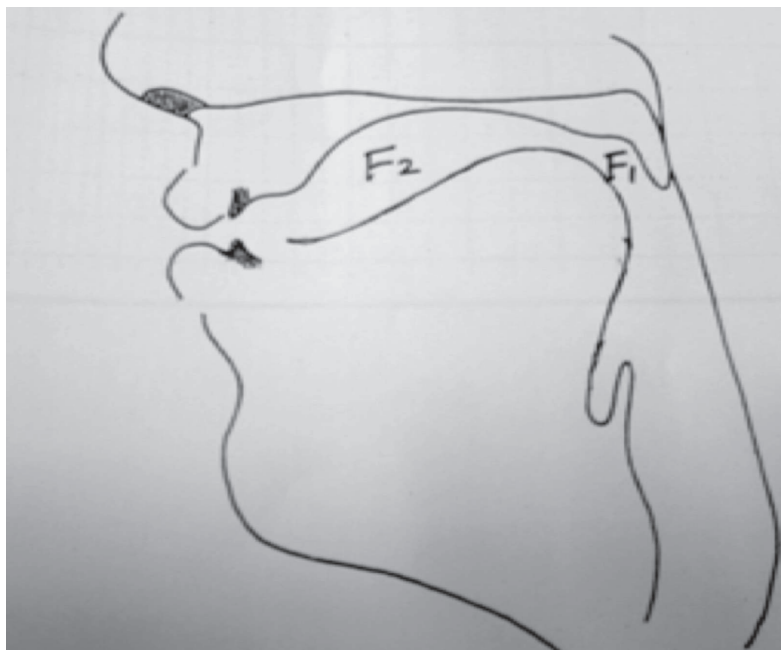
Tunggal, 2005) dan Struktur Bahasa Belait (Noor Alifah, 2004). Kajian-kajian lepas ini menampilkan aspek fonetik dialek Tutong dan dialek Belait berasaskan kajian auditori. Dapatan pengkaji ini masih bersifat subjektif dan perlu dikaji secara saintifik.

IV. METODOLOGI KAJIAN

Seperti yang dinyatakan di atas bahawa analisis akustik berkomputer ini akan mendedahkan kesahihan fonetik sesuatu kata. Sound Filing System (SFS) + WASP for Windows OS menghuraikan bunyi bahasa yang dikaji dalam bentuk: 1. Gelombang bunyi (wave form); 2. Spektrogram dengan pola forman (formant pattern); 3. Surihan tekanan bunyi (energy trace); dan 4. Surihan gegaran bunyi (noice trace). Seterusnya Wavesurfer (WS) for Windows OS pula bermanfaat menghuraikan bunyi ujaran dalam bentuk gelombang bunyi. Dalam kajian ini aplikasi WS akan memanfaatkan kaedah kebalikan (reverse) untuk memastikan kehadiran bunyi-bunyi bahasa yang dikaji secara saintifik, seperti gambaran dan keterangan yang berikut:



Gambar 1. Contoh Hasil Kajian Fonetik Akustik Berkomputer



Gambar 2. Pergerakan Belakang Lidah dan Depan Lidah

Gambar 1 menjelaskan contoh visual hasil kajian fonetik akustik, seperti yang berikut:

Pertama, Gelombang bunyi (bunyi ujaran) dicirikan oleh pola bentuk (waveform) berfungsi memperlihatkan perubahan pita suara yang disebabkan oleh arus udara yang keluar masuk dari atau ke paru-paru; Sementara Spektrogram (spectrogram) dan forman (F) (lihat spektrogram Gambar 2.1) menggambarkan gema, dengan kata lain forman berfungsi menggambarkan pergerakan lidah, iaitu jalur forman (F) yang menerima tekanan bunyi yang kuat akan ditampilkan dengan warna pekat. Dalam hal ini F1 menggambarkan gema di belakang lidah (rongga farinks), sementara F2 pula menggambarkan gema di depan lidah (rongga mulut) (lihat Gambar 4.2). Yang demikian hasil analisis spektrogram menjelaskan bahawa pergerakan lidah dan perubahan bentuknya mempengaruhi nilai forman (F1) dan (F2).

Kedua, Surihan tekanan bunyi (energy trace) menggambarkan turun naik tekanan bunyi suatu ujaran. Bunyi aspirasi dalam ujaran natural memperlihatkan penurunan tekanan bunyi dalam masa yang singkat.

Ketiga, Surihan gegaran bunyi (noise trace) menggambarkan pola sifat vokal dan konsonan. Konsonan biasanya dicirikan oleh pola frekuensi yang tidak berkala (aperiodic), tetapi vokal pula dicirikan oleh pola frekuensi yang agak tetap yang berkala (periodic). Manakala konsonan nasal memperlihatkan ciri vokal dari segi gelombang bunyi.

Keempat, Surihan penyuaran bunyi (voicing trace) berfungsi untuk mengesan perbezaan penyuaran bunyi. Hasil kajian dapat menggambarkan turun naik kadar

getaran pita suara, makin tinggi kadar getaran suara, makin naik surihan penyuaran bunyi. Dengan yang demikian surihan penyuaran rendah menjelaskan bunyi tak-bersuara, manakala surihan penyuaran tinggi membuktikan bunyi bersuara.¹

V. OBJEKTIF KAJIAN

Objektif kajian ini terbahagi kepada dua, iaitu:

- i. Meneliti korpus leksikografi (bertujuan menghasilkan fonetik yang autoritatif) dengan mengapalikasi fonetik akustik berkomputer dengan meneliti beberapa contoh dialek Tutong dan dialek Belait sebagai eksperimen;
- ii. Memperkenalkan kaedah fonetik akustik berkomputer, khususnya kepada ahli leksikografi, sebagai salah satu kaedah fonetik eksperimental yang mendedahkan fakta bahasa tanpa terikat kepada teori-teori fonetik yang tertentu.

VI. DAPATAN KAJIAN

Yang berikut ialah hasil kajian fonetik auditori dan analisis fonetik akustik:

6.1 Kajian Fonetik Auditori

Seperti yang dinyatakan bahawa tradisi penjanaan korpus fonetik leksikografi dilakukan melalui kajian fonetik auditori. Antara contohnya dipaparkan sebagai yang berikut:

Bil.	Kata	Sebutan	Makna	Kata	Sebutan	Makna
1	pati	[patʔ]	patah	pati	[pati ^h]	kematian
2	jelu	[dʒəluʔ]	semat	jelu	[dʒəlu ^h]	sejenis ikan yu
3	puti	[putiʔ]	putih	puti	[puti ^h]	pisang
4	tuju	[tuɖʒuʔ]	tujuh	tuju	[tuɖʒu ^h]	tuju; hala
5	bulu	[buluʔ]	bulu	bulu	[bulu ^h]	buluh
6	bua	[buaʔ]	buah	bua	[bua ^h]	buah (penjodoh bilangan)

¹ Keterangan yang lebih terperinci (lihat Sato dalam Jurnal Bahasa Januari–April 2011).

7	mulo	[muloʔ]	kiasan 1. gatal tangan 2. gatal mulut	mulo	[mulo ^h]	menanam (sesuatu benih dan lain-lain)
8	tayi	[tajiʔ]	tahi	tayi	[taji ^h]	hilang;l enyap
9	kali	[kaliʔ]	mungkin	kali	[kali ^h]	kali (darap)
10	ilu	[iluʔ]	lampu (pelita)	ilu	[ilu ^h]	kuku
11	ala	[alaʔ]	hasil	ala	[ala ^h]	kalah

Jadual 1. Dapatan Kajian Lapangan Auditori

Hasil kajian lapangan (kajian fonetik auditori) di atas mendedahkan kewujudan bunyi hentian glotis [ʔ] dan aspirat [h] seperti contoh-contoh yang dipaparkan dalam Jadual 6.1, yang berjalur biru merupakan leksikal-leksikal yang menggambarkan bunyi-bunyi hentian glotis[ʔ]. Yang berjalur hijau pula ialah hasil lapangan auditori yang menggambarkan bunyi aspirat [h]:

- 1) a. Pati [pati ʔ] bermaksud ‘Apatah’ dalam ayat: [poo^h ni^h patiʔ](kakinya patah)
b. pati [pati^h] bermaksud ‘kematian’ Contoh ayat :
[kudo^h pati^h ni^h] (bilahkah dia meninggal dunia ?)
- 2) a. Jelu [ɟəluʔ] bermaksud ‘semat’ Contoh ayat :
[ɟəluʔ butang baɟuʔ anaʔ inaʔ] (sematkan butang baju budak itu)
b. Jelu [ɟə^h] bermaksud sejenis ikan laut, berbadan panjang, seperti ular berukuran 2– 3 kaki, tidak bersisik, dan kepalanya pipih.
Contoh ayat: [ndo^h ɟiʔ sajiʔ kəgitan putaʔ jəlu^h] (saya tidak pernahmelihatikan jelu).
- 3) a. puti [putiʔ] bermaksud warna ‘putih’
Contoh ayat: [kənai^h ni^h kətabiʔ putiʔ](adiknya semua putih)
b. puti [puti^h] bermaksud ‘pisang’ Contoh ayat:
[Mijan buaʔ puti^h] (makan buah pisang)
- 4) a. tuju [tuɟuʔ] bermaksud tujuh (7)
Contoh ayat [tujuʔ tuliʔ buaʔ puti^h] (tujuh biji buah pisang)
b. tuju bermaksud pergi
[dəmbo^h dai tuɟu^h juʔ kənalom ni^h](kemanakah awak pergi semalam?)

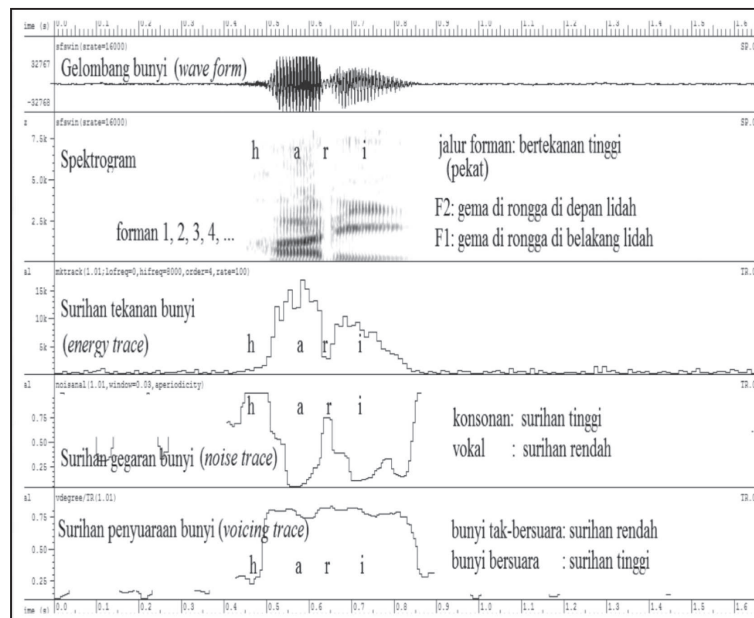
Contoh leksikal di atas memperlihatkan adanya kelainan makna setiap kata berdasarkan sebutan dan konteks yang diutarakan. Dengan yang demikian kehadiran hentian glotis dan aspirat setiap kata perlu ditampilkan dengan fonetik yang betul.

Menurut Nortofer bahawa bahasa Tutong ada konsonan geseran yang lemah sekali, iaitu [h] dan [ʔ] yang tidak ditulis dalam ejaan, misalnya [buŋa^h] dan [tana^ʔ](DBP, 1991).

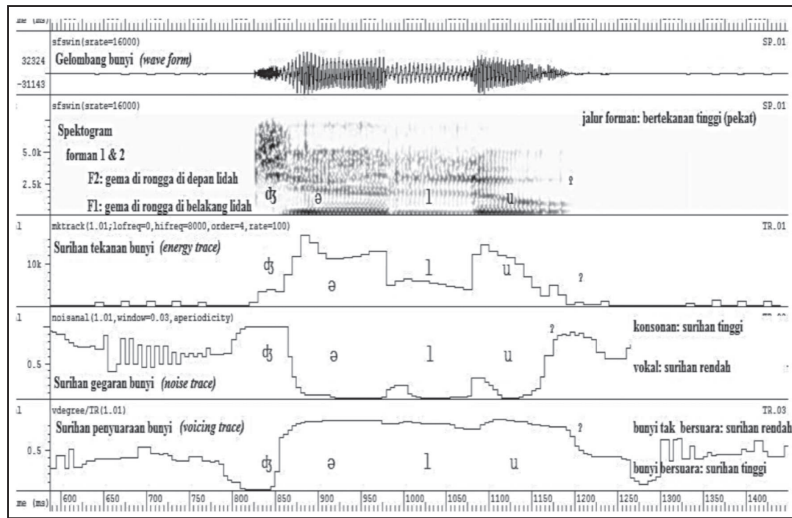
Sehubungan dengan kajian-kajian lepas, hasil kajian fonetik akustik berkomputer yang dikemukakan ini merupakan kajian awal. Contoh yang dibentangkan dalam kertas kerja ini hanya sebagai perintis kepada ahli leksikografi tempatan. Namun, kajian ini membuktikan bahawa kesahihan korpus fonetik untuk tujuan leksikografi dapat dimantapkan dengan memanfaatkan eksperimen fonetik akustik bagi mendeskripsikan fonologi bahasa yang belum mantap sistem ejaan dan sebutannya, misalnya dialek Tutong dan dialek Belait.

6.2 Analisis Fonetik Akustik Bunyi Hentian Glotis

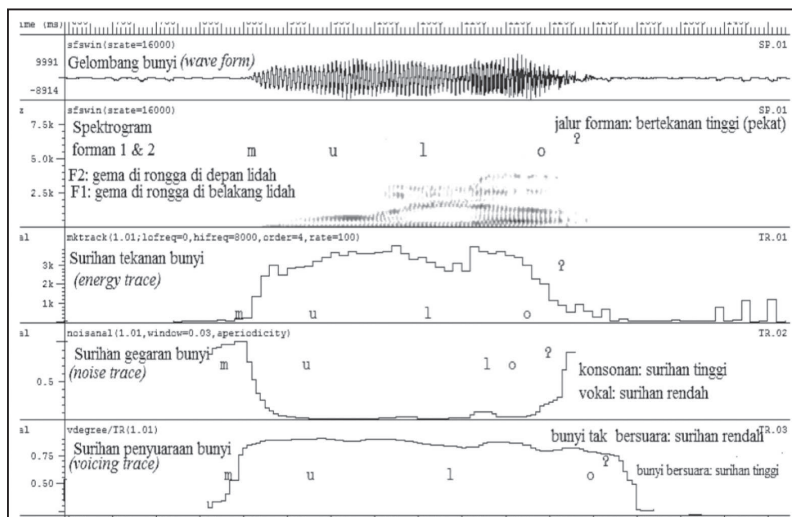
Rajah yang berikut merupakan antara hasil eksperimen yang menggambarkan hasil kajian fonetik akustik, iaitu bunyi hentian glotis yang hadir di akhir kata dalam dialek Tutong:



Gambar 6.2.1. Hasil Kajian Fonetik Akustik Bunyi Hentian Glotis [ɟʔəluʔ]



Gambar 6.2.2. Hasil Kajian Fonetik Akustik Bunyi Hentian Glotis [iluʔ]



Gambar 6.2.3. Hasil Kajian Fonetik Akustik Bunyi Hentian Glotis [muloʔ]

Gambar analisis akustik di atas merekodkan hasil penelitian perkataan “jelu”, “ilu” dan “mulo”. Hasil penelitian gelombang bunyi yang dicirikan oleh pola bentuk (waveform) memperlihatkan perubahan tekanan arus udara amplitud tidak tersekat dengan sempurna yang menandakan celah glotis masih terbuka sedikit (tidak disekat dengan sempurna). Ini membuktikan kehadiran hentian glotis [ʔ] di akhir ketiga-tiga perkataan, iaitu [ɔʔəluʔ] (lihat Gambar 6.2.1), [iluʔ] (lihat Gambar 6.2.2) dan [muloʔ] (lihat Gambar 6.2.3).

Seterusnya gambar di atas merekodkan hasil analisis spektrogram (spectrogram). Hasilnya jalur F1 dan F2 menggambarkan gema bertekanan tinggi, namun kepekatan

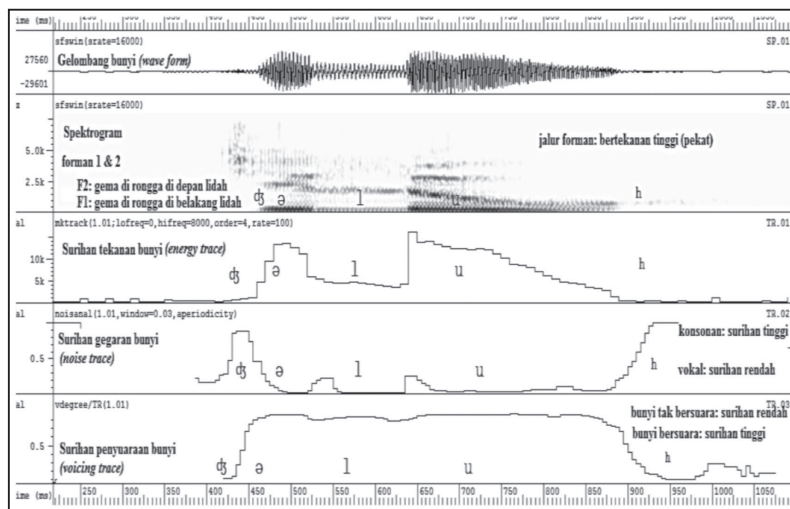
jalur F1 dan F2 kelihatan berkurangan dibandingkan dengan jalur F1 dan F2 pada bunyi sebelumnya yang juga menerima tekanan tinggi, iaitu bunyi vokal [u] [ɔʝəluʔ] (lihat Gambar 6.2.1) dan [u] [iluʔ] (lihat Gambar 6.2.2) dan [o] [muloʔ] (lihat Gambar 6.2.3).

Hasil penelitian surihan gegaran gelombang bunyi (noise trace) kata “jelu”, “ilu” dan “mulo” di atas merekodkan sifat tak berkala (aperiodic). Surihan gegaran bunyi di akhir kata (selepas vokal [u]) ketiga-tiga kata yang dikaji direkodkan semakin menaik. Hal ini mencirikan frekuensi konsonan hadir di akhir kata (selepas vokal [u]). Dengan yang demikian, membuktikan kehadiran hentian glotis [ʔ] di akhir kata ketiga-tiga kata tersebut, iaitu [ɔʝəluʔ] (lihat Gambar 6.2.1), [iluʔ] (lihat Gambar 6.2.2) dan [muloʔ] (lihat Gambar 6.2.3).

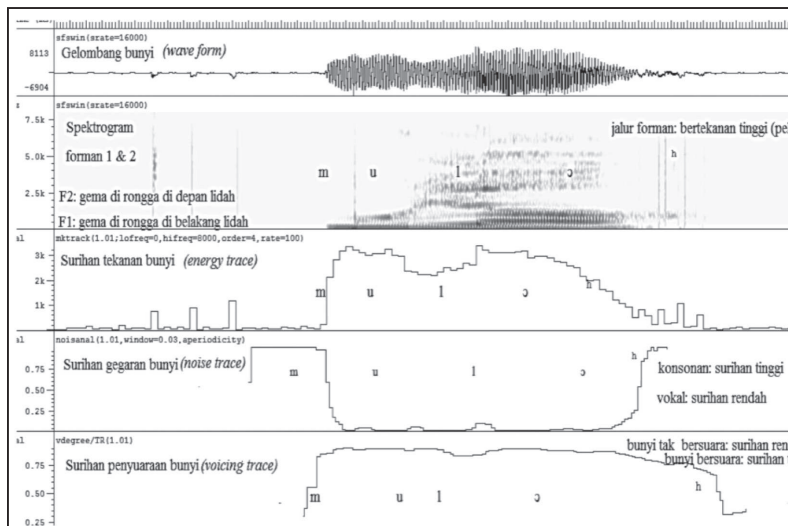
Terakhir hasil SFS memperlihatkan analisis surihan penyuaran bunyi (voicing trace) “jelu”, “ilu” dan “mulo” menggambarkan surihan penyuaran kadar getaran suara kelihatan mula menurun atau rendah. Menurut jadual konsonan IPA bahawa bunyi hentian glotis [ʔ] merupakan bunyi tak bersuara. Dengan yang demikian hasil surihan penyuaran pada akhir ketiga-tiga kata “jelu”, “ilu”, dan “mulo” yang semakin rendah ini jelas membuktikan kewujudan hentian glotis [ʔ], iaitu [ɔʝəluʔ] (lihat Gambar 6.2.1), [iluʔ] (lihat Gambar 6.2.3) dan [muloʔ] (lihat Gambar 6.2.4).

6.3 Analisis Akustik Bunyi Aspirat [h]

Rajah yang berikut merupakan antara hasil yang menggambarkan hasil kajian akustik, iaitu persebaran bunyi aspirat dalam dialek Tutong: [ɔʝəlu^h], [ilu^h], dan [mulo^h]:



Gambar 6.3.1. Hasil Kajian Fonetik Akustik Bunyi Aspirat[ɔʝəlu^h]



Gambar 6.3.3. Hasil Kajian Fonetik Akustik Bunyi Aspirat [muloh]

Hasil penelitian gelombang bunyi fonologi pekataan “jelu”, “ilu” dan “mulo” dicirikan oleh pola bentuk (waveform) memperlihatkan perubahan tekanan arus udara amplitud tidak tersekat yang menandakan celah glotis terbuka. Ini membuktikan kehadiran konsonan. Pola bentuk ini sekali gus membuktikan kehadiran bunyi aspirat [h] di akhir perkataan [ɟəlu^h], [ilu^h] dan [mulo^h].

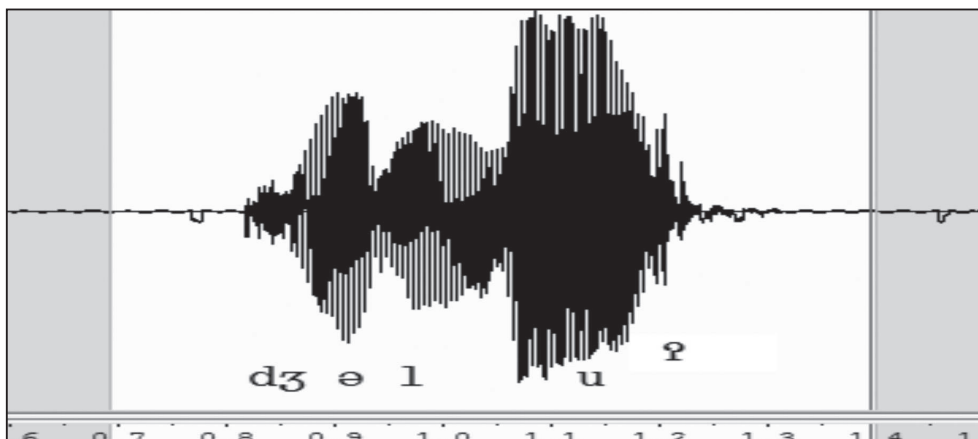
Seterusnya analisis spektrogram (spectrogram) menggambarkan gema, iaitu jalur formant (F) dikesan menerima tekanan bunyi yang lemah. Ini dibuktikan oleh corak F1 dan F2 berwarna cair, iaitu menggambarkan ciri aspirat [h]. Manakala hasil penelitian surihan gegaran gelombang bunyi (noise trace) kata “jelu”, “ilu” dan “mulo” di atas bersifat tak berkala (aperiodic) yang mencirikan frekuensi konsonan di akhir kata. Hasil penelitian ini juga membuktikan kehadiran aspirat [h], iaitu [ɟəlu^h] (lihat Gambar 6.3.1), [ilu^h] (lihat Gambar 6.3.2) dan [mulo^h] (lihat Gambar 6.3.3).

Terakhir hasil analisis surihan penyuaran bunyi (voicing trace) perkataan “jelu”, “ilu” dan “mulo” menggambarkan surihan penyuaran bunyi semakin rendah atau menurun sehingga lenyap. Sebagaimana yang dinyatakan bahawa surihan penyuaran bunyi yang rendah menggambarkan bunyi tak-bersuara. Hal ini membuktikan bahawa akhir kata [ɟəlu^h] (lihat Gambar 6.3.1), [ilu^h] (lihat Gambar 6.3.2) dan [mulo^h] (lihat Gambar 6.3.3) merekodkan kehadiran aspirat. Menurut IPA bahawa bunyi [h] merupakan bunyi konsonan tak bersuara.

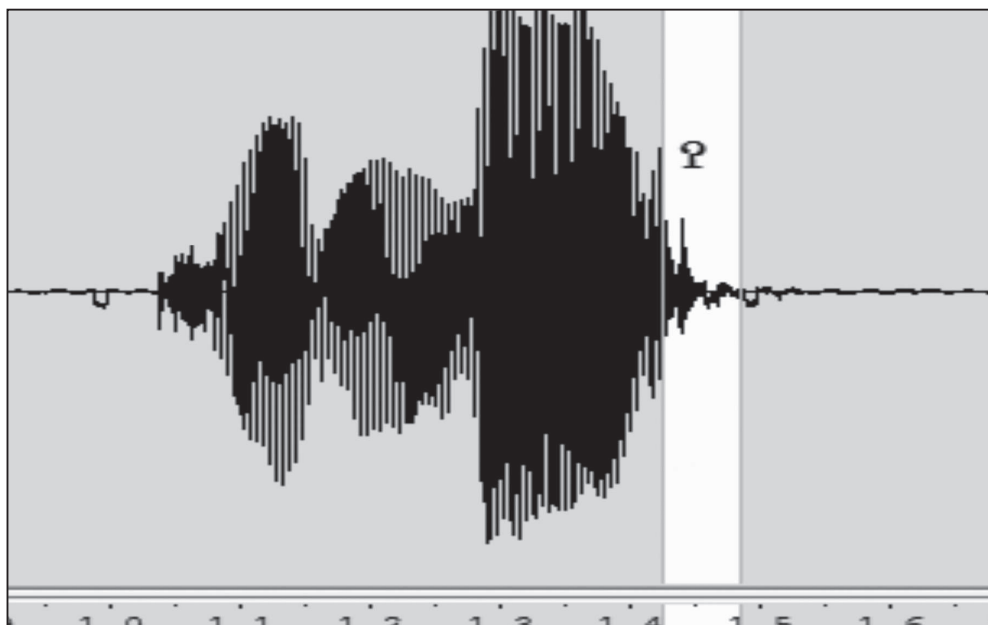
6.4 Hasil Analisis Aplikasi Wavesurfer (WS) for Windows OS

Seperti yang dinyatakan sebelumnya bahasa analisis wavesurfer atau (WS) dapat membuktikan kesahihan fonetik dengan cara menterbalikkan pola bentuk gelombang bunyi asal kata yang dikaji (sama ada kebalikan keseluruhan pola gelombang atau

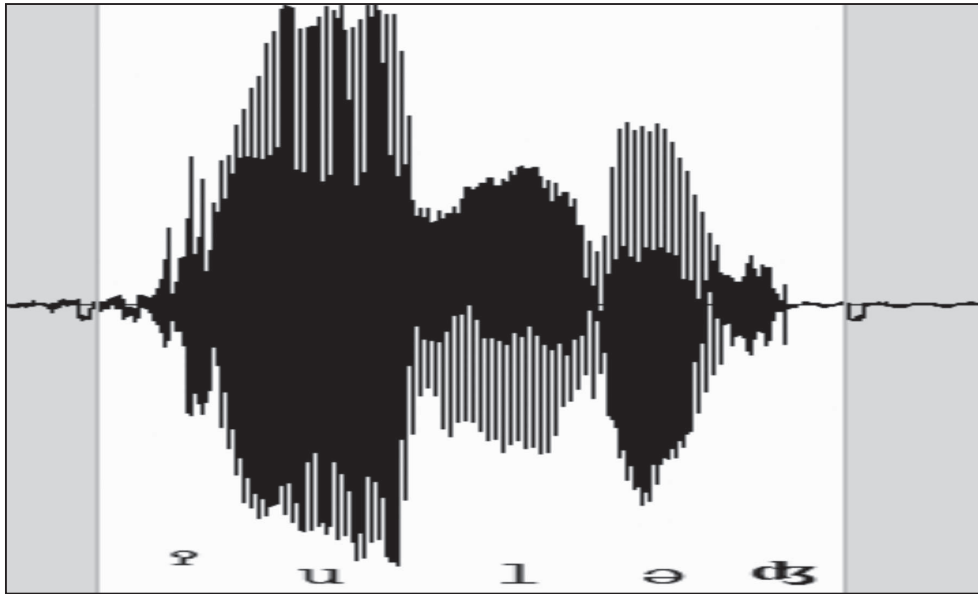
sebahagian yang relevan daripadanya). Antara contoh analisis adalah seperti penjelasan yang berikut:



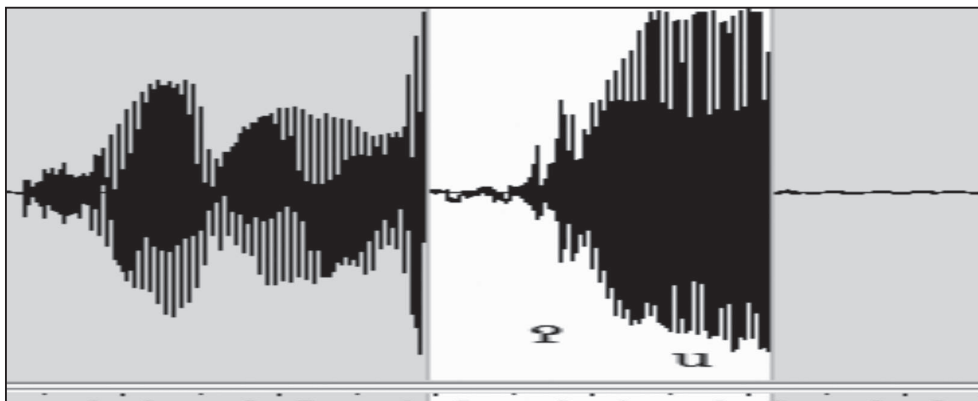
6.4.1. Analisis WS Bunyi Hentian Glotis Gambar 6.4.1.1 Sebutan Asal Jelu (semat) [dʒəluʔ]



Gambar 6.4.1.2. Bunyi Hentian Glotis [ʔ]



Gambar 6.4.1.3. Bunyi kebalikan [ɔ̌ɐluɔ]-> [ɔ̌ɐɫɔ̌]

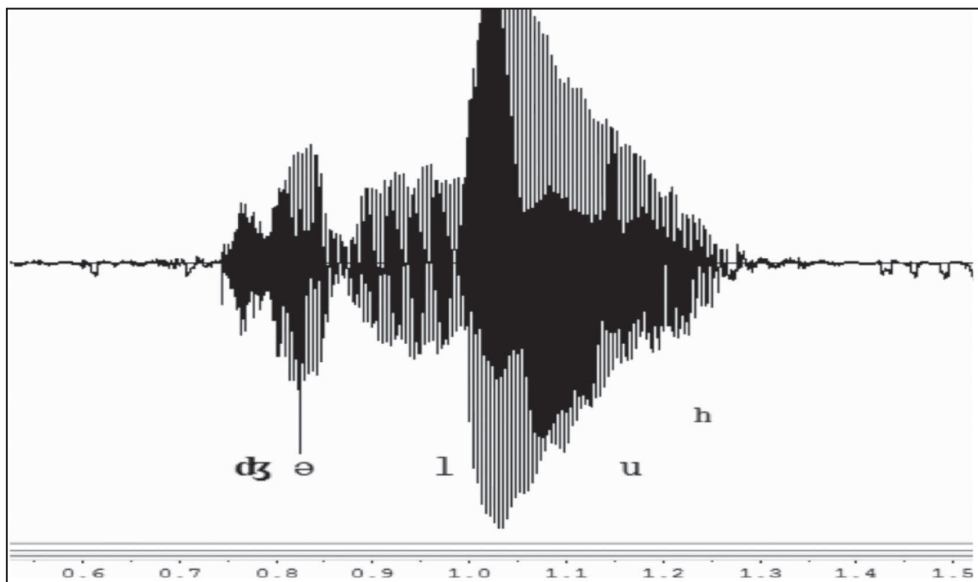


Gambar 6.4.1.4. bunyi kebalikan [luɔ] suku kata akhir

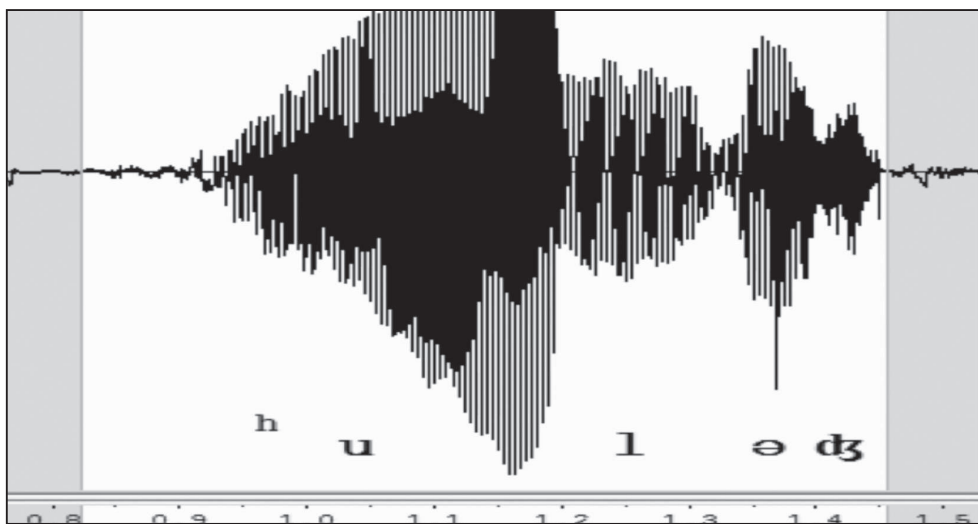
Gambar 6.4.1.1 merekodkan bunyi sebutan asal perkataan “jelu”. Hasil analisis pola gelombang akhir suku kata, pola gelombang menggambarkan arus udara tersekat namun tidak sempurna. Sementara Gambar 6.4.1.2 merekodkan kehadiran hentian glotis [ɔ̌] yang tidak sempurna kerana pola gelombang menandakan arus udara tidak tersekat dengan sempurna. Apabila bunyi kata ini diterbalikkan (reverse) secara keseluruhannya, bunyi kebalikan kata “jelu” melahirkan bunyi [ɔ̌ɐɫɔ̌]. Gambar 6.4.1.3 bunyi kebalikan merekodkan kata ini dimulai dengan kehadiran hentian glotis. Apabila suku kata akhir ([luɔ]) diterbalikkan, sekali lagi dapat merekodkan kehadiran hentian glotis yang tidak sempurna. (lihat Gambar 6.4.1.4)

6.4.2 Analisis WS Bunyi Aspirat [h]

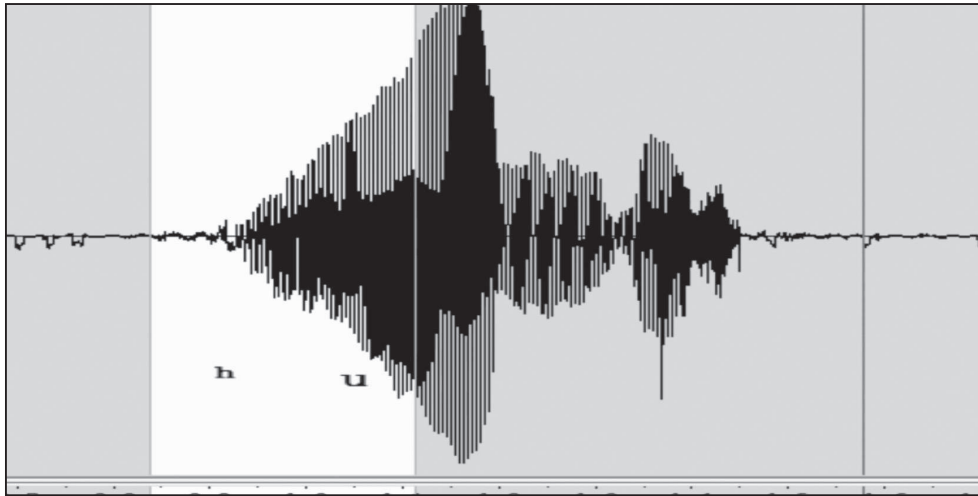
Yang berikut merupakan hasil eksperimen aplikasi WS bagi fonetik aspirat [h]:



Gambar 6.4.2.1. Sebutan Asal Kata Jelu (sj ikan laut) [dʒəlu^h]



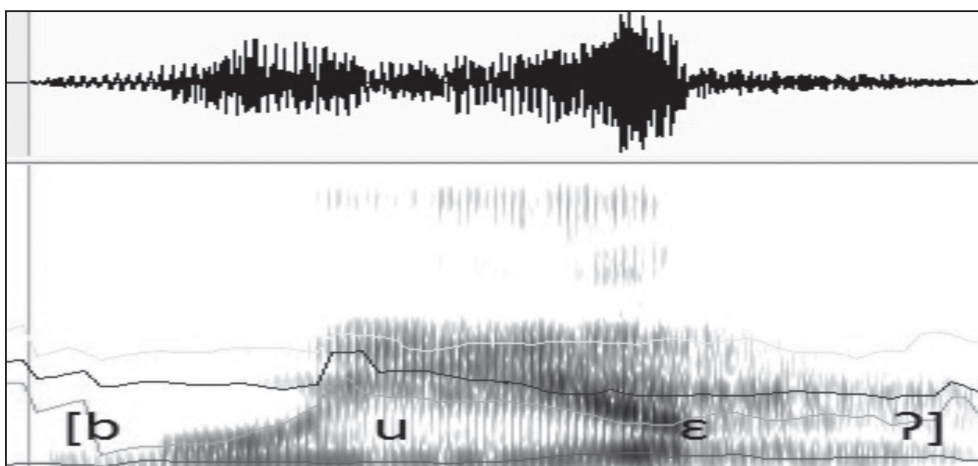
Gambar 6.4.2.2. Kebalikan Kata Jelu[dʒəlu^h]



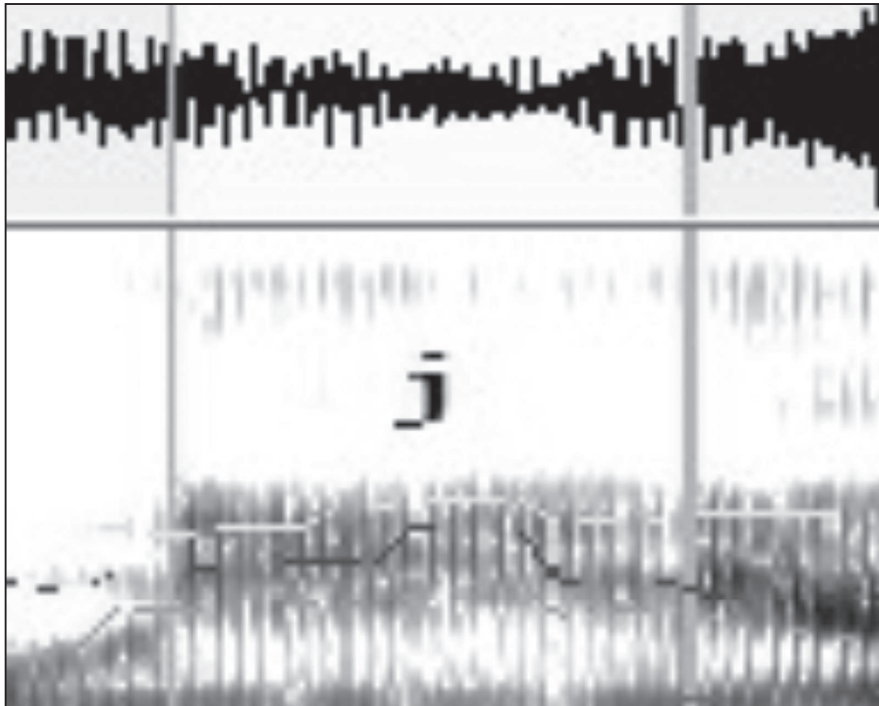
Gambar 6.4.2.3. Kebalikan Bunyi [u] Pada Suku Kata Akhir

Gambar 6.4.2.1 merekodkan sebutan asal perkataan yang menggambarkan kehadiran aspirat ([ɔʝəlu^h]). Sementara Gambar 6.4.2.2 merupakan hasil analisis kebalikan kata “jelu” secara keseluruhannya. Hasil analisis kebalikan ini merekodkan kata “jelu” sebagai [ʰuləɔʝ], iaitu menunjukkan hadirnya bunyi aspirat yang lemah. Begitu juga apabila bunyi [u] pada suku kata akhir diterbalikkan, sekali lagi bunyi aspirat [ʰ] kedengaran dan melahirkan bunyi [ʰu]. Dengan yang demikian kaedah WS ini dapat memantapkan lagi analisis fonetik akustik dengan mendeskripsikan bunyi-bunyi ujaran secara terbalik.

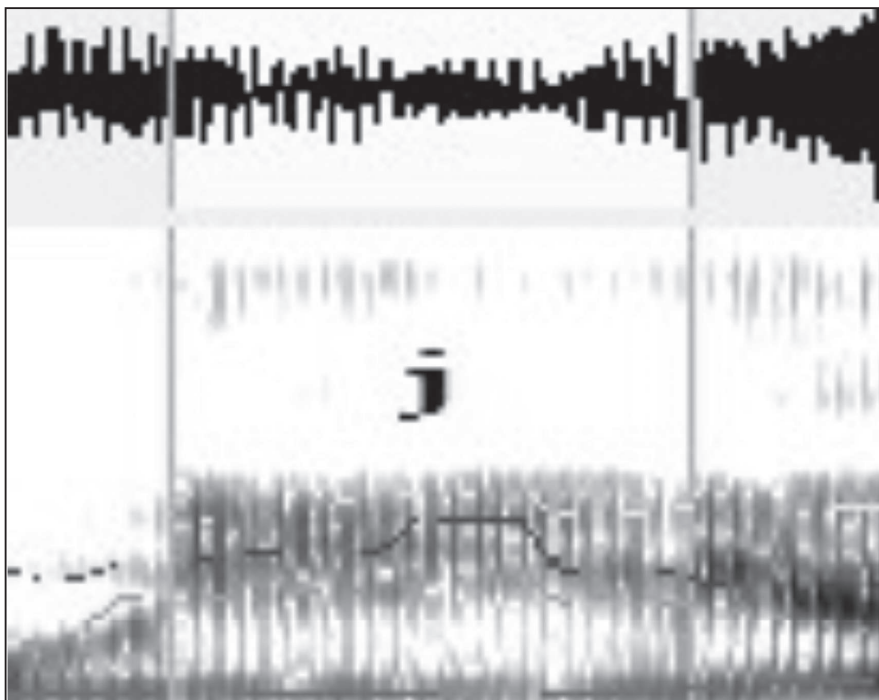
6.4.3 Analisis WS Bunyi Geluncurn (gliding sound) atau Penglenglangit-Kerasan (Palatalization)



Gambar 6.4.3.1 Gambar Penuh Bunyi Buek [buɛʔ]

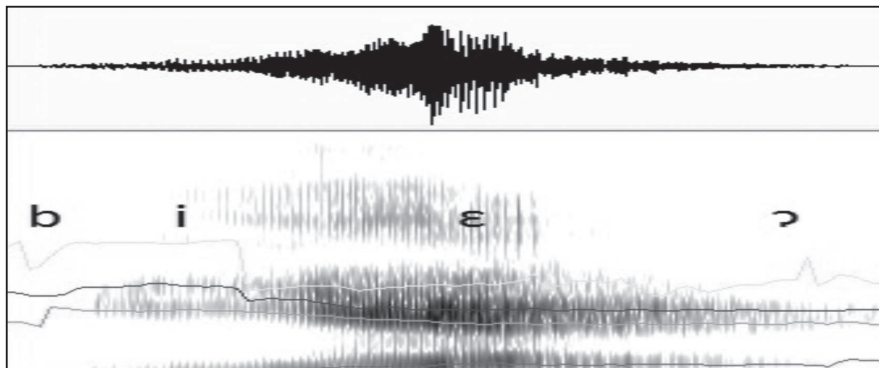


Gambar 6.4.3.2. Bunyi Asal Bahagian [j]

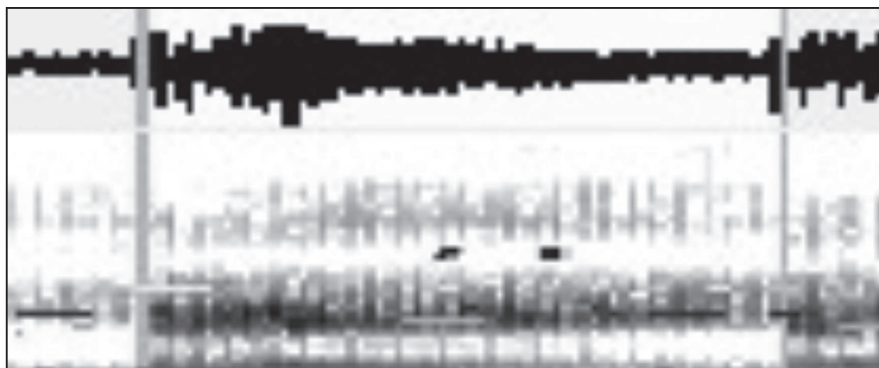


Gambar: 6.4.3.3. Setelah Didekod Terbalik

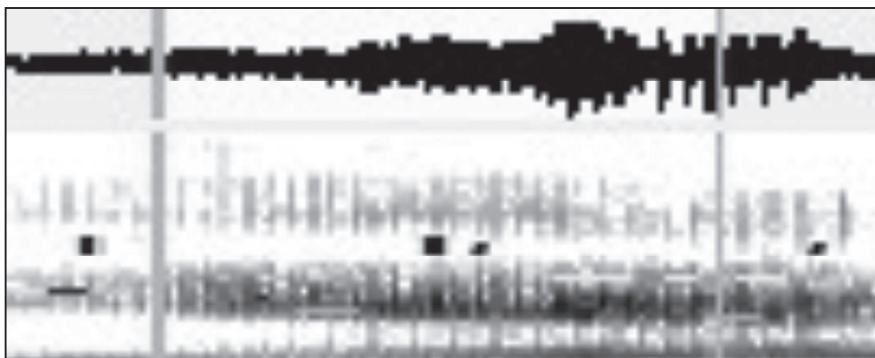
Gambar: 6.4.3.1 merupakan contoh hasil analisis sama ada hadirnya bunyi pengel langit-kerasan (palatalization) atau cuma bunyi geluncuran dalam kata (buek), yang bermakna bunga dalam bahasa Belait. Hasil analisis fonetik auditori kedengaran bunyi [j] di antara bunyi [u] dengan [ɛ] kata tersebut. Seterusnya hasil analisis fonetik akustik merekodkan kehadiran bunyi [j] (lihat Gambar 6.4.3.2). Untuk pembuktian kehadiran bunyi [j], bahagian yang sama didekod secara terbalik. Hasilnya merekodkan bunyi geluncuran [j] kekal seperti hasil fonetik auditori (lihat gambar 6.4.3.3). Dengan yang demikian, fakta ini mengesahkan kehadiran bunyi geluncuran [j]). Hal ini dibandingkan dengan hasil analisis bunyi geluncuran pada kata (biek), seperti keterangan yang berikut:



Gambar 6.4.3.4. Gambar Asal Kata [biɛʔ]



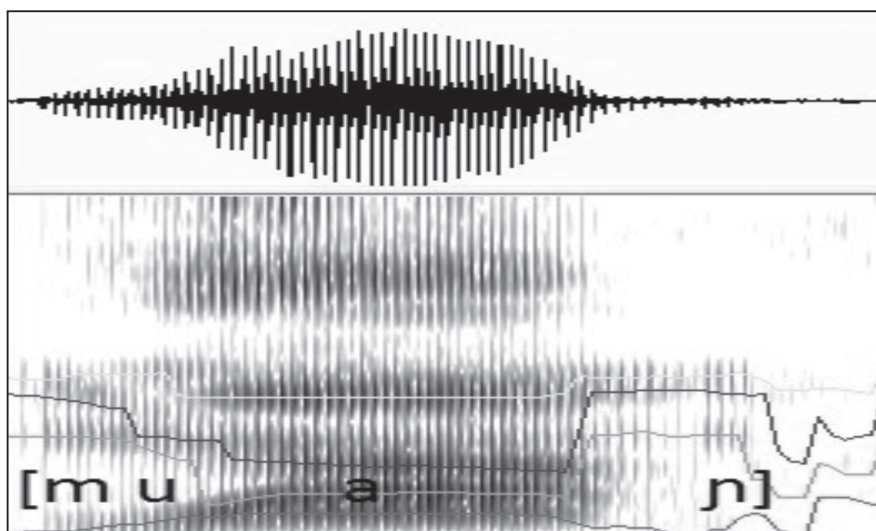
Gambar 6.4.3.5. Hasil Bunyi Didekod Bunyi [iɛ]



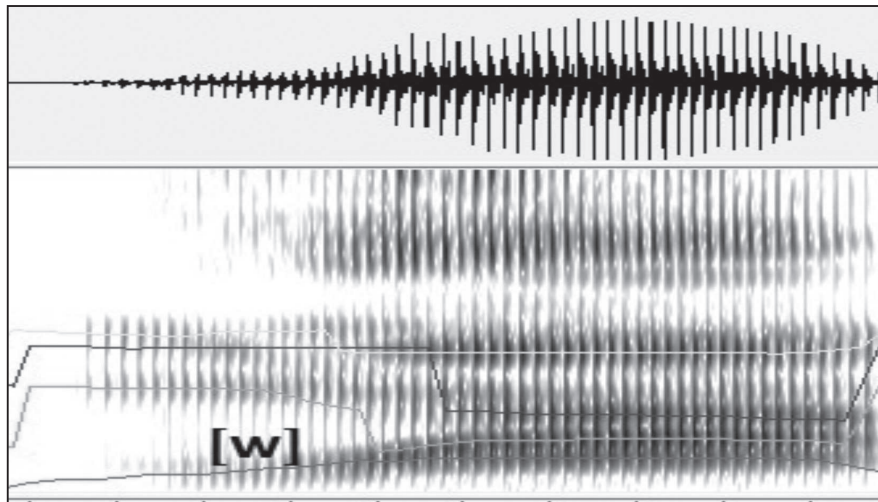
Gambar 6.4.3.6. Hasil Didekod Kebalikan Membuktikan [ɛi] Bukan [j]

Gambar 6.4.3.4 memperlihatkan kata (biek) yang bermakna basah dalam dialek belait. Berdasarkan analisis fonetik auditori kedengaran bunyi [j] yang sama seperti dapatkan pada kata buyek [bujɛʔ]. Namun setelah dianalisis secara fonetik akustik (WS) ternyata hanya hadir buyi peralihan [i], [iɛ], [ɛ] yang tidak dapat dipisahkan (lihat Gambar 6.4.3.5). Untuk kepastian lebih lanjut, bunyi ini didekodkan secara terbalik menggunakan aplikasi WS. Dapatkan metode ini merekodkan bunyi [ɛi] bukan [j].

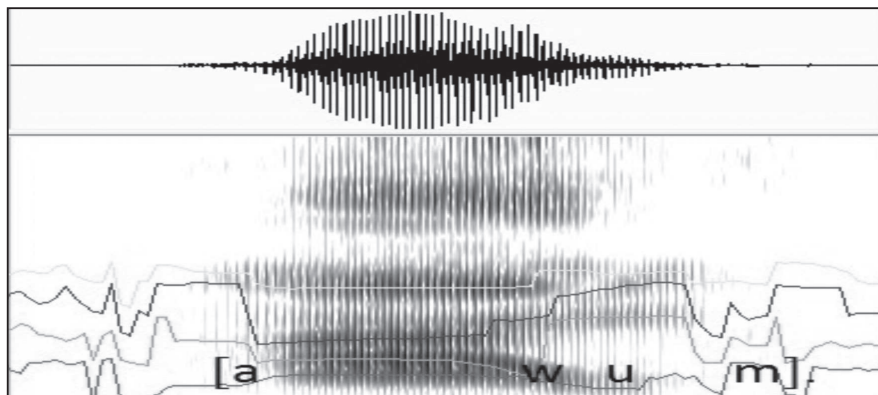
6.4.4 Analisis WS Bbunyi Pembibiran (labialization)



Gambar 6.4.4.1. Kata Muang



Gambar 6.4.4.2. [w] dalam Kata Muang



Gambar 6.4.4.3. Bunyi Didekod Kebalikan

Bunyi yang dikaji seterusnya ialah bunyi kata (muang) yang bermaksud luar (lihat Gambar

6.4.4.1 Kata Muang). Berdasarkan kajian fonetik auditori bahagian kata ini juga kedengaran bunyi [w]. Apabila dianalisis menggunakan WS, hasilnya merekodkan bunyi [w]. (lihat Gambar 6.4.4.2). Seterusnya bunyi di bahagian ini didekodkan secara terbalik, hasilnya merekodkan bunyi [awum], seperti paparan Gambar: 6.4.4.3).

VII. KESIMPULAN

Makalah ini meneliti korpus leksikografi aplikasi fonetik akustik berkomputer. Melalui eksperimen terhadap contoh-contoh korpus leksikografi, kajian ini berhasil mendeskripsikan fonetik kehadiran bunyi aspirat, bunyi hentian glotis menggunakan SFS. Manakala bunyi penglelangit-kerasan (palatalization) dan bunyi pembibir (labialization)

merupakan antara contoh yang dapat dibuktikan oleh aplikasi WS secara kebalikan. Melalui penelitian rintis ini dikesan metodologi kajian fonetik akustik berkomputer sangat bermanfaat dalam penelitian korpus leksikografi. Metode ini juga dikesan meningkatkan ketelitian deskripsi fonetik yang dibuat secara fonetik auditori. Dengan yang demikian kaedah baru ini bukan sahaja membetulkan kesalahafaman pengkaji secara berterusan, malah dengan metodologi ini dapat meningkatkan jumlah korpus leksikografi, sekali gus korpus fonologi dan fonetik leksikografi dapat disahihkan secara objektif.

Daftar Pustaka

- International Phonetic Association, 1999. *Handbook of International Phonetic Association: A Guide to the use of the International Phonetic Alphabet*. Cambridge University Press.
- Nothofer 1991. "Panduan Kepada Pengguna" dlm Kamus Tutong–Melayu Melayu–Tutong
- Ramlee Tunggal, 2005. *Struktur Bahasa Tutong*. Dewan Bahasa dan Pustaka Brunei Sato,
- Hirobumi @ Rahmat, 2011. "Analisis Akustik Konsonan/-h/ Dalam Kata Terbitan Berimbuhan {-an} /-an/" dlm Jurnal Bahasa Januari-April, hlm.110-126. Dewan Bahasa dan Pustaka Brunei
- Rujukan Internet: www.speech.kth.se/wavesurfer www.phon.ucl.ac.uk/resource/sfs/download.php

Perancangan Aplikasi Kamus *Mobile* Bahasa Madura sebagai Upaya Pelestarian Bahasa Daerah

Salim Anshori & Iqbal Nurul Azhar

Prodi Sastra Inggris, Universitas Trunojoyo Madura

salimku@gmail.com

iqbalnurulazhar@yahoo.com

Abstrak

Saat ini, upaya pengembangan bahasa daerah (Madura) mengalami banyak kendala. Salah satunya adalah masalah ketertarikan generasi muda dalam menggunakan dan melestarikan bahasa Madura. Bahkan, ada anggapan dari sebagian generasi muda, bahwa bahasa Madura itu sudah kuno, penggunaanya ketinggalan jaman dan akhirnya penggunaan bahasa Madura cenderung menjadi bahan olok-olokan. Akibatnya, generasi muda yang seharusnya menjadi pewaris bahasa Madura justru semakin enggan untuk menggunakan bahasa Madura dalam kegiatan sehari-hari.

Untuk menjawab tantangan ini, maka kami memanfaatkan platform teknologi “mobile” berbasis Android (salah satu teknologi yang populer dikalangan generasi muda saat ini) untuk merancang kamus bahasa Madura. Metode yang diterapkan adalah Hashmap (metode yang digunakan untuk memudahkan pencarian data/kata). Tujuan penelitian ini adalah untuk merancang aplikasi kamus mobile bahasa Madura. Dari sini diharapkan, bahasa Madura semakin dekat dengan generasi muda penutur aslinya sehingga bahasa Madura bisa semakin lestari, adanya anggapan bahwa bahasa Madura itu kuno, tertinggal jaman akan terbantahkan, dan untuk memudahkan penutur bahasa lain dalam mempelajari bahasa Madura.

Key words: bahasa Madura, kamus android, pelestarian bahasa daerah

I. PENDAHULUAN

1.1 Latar Belakang

Kondisi bahasa daerah di Indonesia dewasa ini semakin mengkhawatirkan. Sudah banyak penelitian dilakukan untuk merevitalisasi bahasa daerah. Pun banyak yang menemukan fakta bahwa keberadaan bahasa daerah di Indonesia semakin terpuruk (Santoso, Budiwiyanto, Azhar. 2008, Darwis. 2011). Penyebab semakin terpinggirkannya bahasa daerah di Indonesia ini bermacam-macam. Namun dari berbagai penelitian, hampir semuanya menyebutkan bahwa salah satu faktor utama terpinggirkannya bahasa daerah pada saat ini adalah karena adanya gempuran atau tantangan globalisasi yang luar biasa. Tantangan globalisasi ini utamanya terkait dengan generasi muda kita. Hal ini sangat penting untuk diperhatikan, karena mau tidak mau, generasi muda kita inilah yang akan menjadi pewaris bahasa – bahasa daerah tersebut. Tanpa mereka, bahasa – bahasa daerah manapun akan punah dengan sendirinya.

Budiyanto, misalnya, menyebutkan bahwa para ahli bahasa meyakini akan ada sekitar setengah dari bahasa – bahasa di dunia yang akan punah. Sedangkan khusus di Indonesia, Moseley (2010) menyebutkan bahwa ada 146 bahasa yang terancam punah dan 12 bahasa yang telah punah. Bahasa-bahasa daerah itu umumnya berada di Indonesia bagian timur. Bahasa-bahasa daerah yang teridentifikasi telah punah misalnya adalah bahasa Hukumina, Kayeli, Liliali, Moksela, Naka'ela, Nila, Palumata, Piru, dan Te'un di Maluku, Mapia dan Tandia di Papua, serta Tobada' di Sulawesi. Banyak dari bahasa-bahasa daerah yang punah itu disebabkan karena jumlah penuturnya yang sudah habis atau tinggal sedikit. Namun pada kenyataannya, bahkan bahasa- bahasa daerah yang penuturnya masih sangat banyak (jutaan bahkan puluhan juta penutur) sekalipun, saat ini juga bisa terancam punah karena arus globalisasi. Misalnya adalah bahasa Madura. Sekalipun bahasa Madura masih mempunyai penutur bahasa yang sangat banyak, bahkan puluhan juta penutur¹, namun tantangan dan upaya untuk menyelamatkan bahasa ini dari kepunahan juga tidak mudah. Azhar (2008) mencermati bahwa generasi muda Madura saat ini semakin tidak 'akrab' untuk menggunakan bahasa daerahnya untuk berkomunikasi dalam konteks bahasa tulis, baik dalam bentuk surat maupun SMS. Terlepas dari adanya beberapa hambatan secara internal bahasa, adanya tren globalisasi

¹ Data dari Biro Pusat Statistik yang diungkapkan oleh Kepala balai Bahasa Propinsi Jawa Timur, Drs. Amir Mahmud, M. Pd. dalam wawancaranya dengan reporter Lontar Madura (*El Iemawati*) menunjukkan bahwa jumlah penutur bahasa Madura ada jutaan, walaupun ada kecenderungan selalu menurun dari waktu ke waktu. (Pada tahun 1980 jumlah penutur bahasa Madura berjumlah 6.913.977 orang, pada tahun 1990 berjumlah 6.792.447 penutur). Sedangkan Situs www.ethnologue.com menyebutkan, berdasar sensus tahun 2000, penutur bahasa Madura ada 6.770.000 penutur. Walaupun dalam situs tersebut status bahasa Madura adalah "berkembang" namun jelas pada kenyataannya, jumlah penuturnya semakin berkurang walaupun secara total, jumlahnya masih jutaan.

yang ditunjukkan dengan perkembangan teknologi, tidak bisa diabaikan perannya dalam permasalahan ini. Adanya perkembangan teknologi dan penggunaannya oleh generasi muda ternyata ikut mempengaruhi pola pikir mereka. Yang paling utama, perubahan terhadap sikap generasi muda pada bahasa- bahasa daerahnya.

Hal senada juga disampaikan oleh Darwis (2011) bahwa ternyata tidak semua generasi muda kita masih memperoleh dan mempelajari bahasa daerahnya masing-masing. Kejadian ini terjadi tidak hanya di daerah perkotaan saja, namun juga terjadi di daerah pedesaan. Artinya, bahasa- bahasa daerah sudah (dan masih terus berlanjut) terdesak perkembangannya. Dalam kasus bahasa Makasar dan bahasa Bugis (yang menjadi kajiannya Darwis, 2011) disebutkan paling tidak ada 3 faktor yang menyebabkan bahasa daerah tersebut tergeser dari posisinya sebagai bahasa pertama untuk generasi muda di daerah perkotaan. Yang pertama adalah lingkungan pergaulan yang majemuk bahasa (suku). Kedua, medan tugas yang relatif tidak tetap. Ketiga, orang tua berlainan suku. Sedangkan untuk daerah pedesaan, faktor utama yang berpengaruh adalah faktor kehadiran lembaga pendidikan Taman Kanak-Kanak (TK) di pedesaan. Di sekolah TK ini biasanya guru-guru menggunakan bahasa Indonesia sebagai bahasa pengantarnya, akibatnya, orang tua anak-anak juga ikut terpengaruh. Mereka mempersiapkan anak-anaknya untuk menguasai bahasa Indonesia (bukan bahasa daerahnya) agar tidak tertinggal dalam pelajaran di sekolahnya.

Setali tiga uang, hal serupa juga terjadi di Propinsi Aceh. Tim pemetaan bahasa Balai Bahasa Banda Aceh melaporkan bahasa Devayan yang merupakan bahasa daerah di Kepulauan Banyak, mulai diindikasikan akan punah. Lebih lanjut, dikhawatirkan hal serupa juga akan terjadi pada bahasa daerah lainnya, misalnya bahasa Aceh dan Gayo. Sekalipun saat ini bahasa Aceh masih digunakan hampir setiap wilayah di Provinsi Aceh. Sebagai catatan tambahan, bahasa Aceh merupakan bahasa daerah dengan jumlah penutur terbanyak di propinsi ini, dan disusul oleh bahasa Gayo (Santoso. TT). Masih menurut Santoso, kasus bahasa daerah di Aceh juga menarik untuk dikaji. Menurutnya, ada hal yang ironis terjadi di Aceh, dimana banyak orang-orang yang menyadari pentingnya melestarikan bahasa daerah, namun pada kenyataannya, dia sendiri sering menggunakan istilah-istilah asing yang notabene sudah ada padanannya dalam bahasa daerah (dan Indonesia).

Lebih khusus, Santoso menyoroti generasi muda Aceh yang mulai meninggalkan bahasa daerahnya. Mereka cenderung menggunakan bahasa gaul. Fenomena ini lebih marak dijumpai di daerah perkotaan. Hal ini bisa dipahami karena generasi muda ini masih labil dan mudah terpengaruh. Mereka sedang dalam masa pencarian jati dirinya. Namun menurut Santoso (TT), seharusnya akan lebih baik jika mereka lebih dulu menguasai dan mengutamakan bahasa daerahnya sendiri sebagai pondasinya. Namun yang terjadi saat ini adalah sikap berbahasa penutur bahasa Aceh, utamanya generasi mudanya, yang cenderung negatif. Mereka lebih suka menggunakan istilah-istilah asing agar terlihat lebih modern dan intelek. Padahal sebenarnya mereka juga belum tentu

benar-benar paham dengan istilah-istilah asing yang mereka gunakan tersebut. Bukan hanya generasi mudanya, namun juga penggunaan papan-papan nama di pinggir jalan, mereka lebih suka menggunakan istilah asing. Inilah yang disebut oleh Santoso (TT) sebagai fenomena *budaya nginggris* yang diyakini tidak hanya terjadi di Aceh, tapi juga diberbagai daerah Indonesia lainnya.

Tekanan terhadap bahasa daerah juga datang dari para penguasa pembuat kebijakan yang belum benar-benar berpihak kepada kelestarian bahasa Daerah. Sebutlah apa yang terjadi di Provinsi Aceh. Santoso merasakan adanya diskriminasi perlakuan pemerintah terhadap bahasa daerah. Perlakuan itu salah satunya mengakibatkan tertundanya kegiatan Kongres Bahasa Aceh yang sebelumnya (akan) rutin diagendakan. Memang, saat ini telah ada produk hukum berupa Undang-Undang No.24 Tahun 2009 tentang Bendera, Bahasa, Lambang Negara, dan Lagu Kebangsaan. Khusus tentang bahasa, di dalam undang-undang tersebut telah diatur bagaimana kedudukan dan fungsi bahasa daerah, bahasa Indonesia, dan bahasa asing. Namun seiring dengan adanya kebijakan otonomi daerah, implementasi dari Undang-Undang ini menjadi tumpang tindih dan menjadi kabur. Utamanya untuk bahasa daerah, ternyata tidak semua daerah benar-benar memahami pentingnya kelestarian bahasa daerahnya masing-masing, seperti halnya yang terjadi di Aceh. Pada akhirnya, Santoso berharap, sekalipun kita menjadi bagian pemain di dalam modernitas itu, tetapi tetap kukuh dengan nilai-nilai kedaerahan (aceh). Dengan demikian, identitas kedaerahan (aceh) itu tetap dapat dipertahankan.

Dari uraian diatas, cukup jelas terasa betapa beratnya tantangan bahasa daerah untuk tetap bertahan di era globalisasi ini. Mulai dari hambatan internal, yakni adanya perbedaan sistem tata bunyi, perbendaharaan kata dan sisten tata kalimatnya dengan bahasa lain, sampai dengan hambatan eksternalnya. Hambatan eksternal ini cukup kompleks dan melibatkan banyak unsur. Baik masyarakat penuturnya (utamanya generasi muda) yang bersikap negatif terhadap bahasa daerahnya, tekanan dari bahasa lainnya (bahasa Indonesia dan bahasa Inggris) sampai dengan minimnya dukungan dari pembuat kebijakan.

Makalah ini tidak hendak menyelesaikan semua permasalahan diatas. Namun penulis melihat ada benang merah antara permasalahan keengganan generasi muda menggunakan bahasa daerahnya dengan tren globalisasi. Di uraian diatas disinggung bahwa banyak ahli bahasa yang beranggapan bahwa globalisasi dan modernisasi dengan berbagai produknya adalah salah satu faktor penting yang membuat generasi muda semakin menjauh dengan bahasa daerahnya. Namun penulis justru kebalikannya, berkeyakinan bahwa globalisasi dan modernisasi bisa menjadi salah satu alat untuk mendekatkan generasi muda dengan bahasa daerahnya. Globalisasi dan modernisasi sudah kita sadari bersama, tidak mungkin dihindari. Oleh karenanya, tidak perlu dimusuhi. Justru sebisa mungkin kitalah yang harus memanfaatkannya untuk kepentingan kita bersama.

Dalam hal ini, penulis melihat saat ini generasi muda kita sedang mengalami tren menggunakan *smartphone* sebagai salah satu media komunikasi. Semakin hari semakin terasa betapa intensnya mereka menggunakan perangkat ini. Hal ini tidak bisa dielakkan.

Oleh karenanya penulis menawarkan bagaimana caranya mendekatkan masyarakat (utamanya generasi muda) kepada bahasa daerahnya dengan memanfaatkan produk modern ini. Bagaimana generasi muda sebagai pengguna *smartphone* bisa tetap dekat dengan bahasa daerahnya? Salah satu caranya adalah dengan menyediakan ‘bahasa daerah’ tersebut di perangkat yang mereka gunakan. Dalam hal ini, penulis menawarkan perancangan kamus mobile bahasa daerah (Madura) sebagai salah satu upaya untuk mendekatkan kembali generasi muda dengan bahasa daerahnya. Harapannya, dengan semakin akrabnya mereka dengan bahasa daerahnya, maka bahasa daerah tersebut juga akan semakin lestari.

1.2 Masalah

Permasalahan yang diangkat dalam makalah ini adalah bagaimanakah perancangan kamus *mobile* bahasa Madura berbasis Android untuk mendukung upaya pelestarian bahasa daerah?

1.3 Tujuan

Tujuan yang hendak dicapai dalam makalah ini adalah merancang kamus *mobile* bahasa Madura berbasis Android untuk mendukung upaya pelestarian bahasa daerah.

1.4 Tinjauan Pustaka

Upaya memanfaatkan teknologi untuk pembelajaran bahasa daerah di Indonesia sudah dilakukan oleh banyak *programmer*. Beberapa diantaranya adalah Farida (2015), Lamusu, Koniyo dan Rahandi (2013), Ely, Simon dan Irwansyah (2013), Sembiring (2013), dan Priharyanto (2012).

Farida (2015) menulis artikel “*Aplikasi Kamus Indonesia-Jawa Berbasis Android*”. Karya Farida (2015) ini dilatarbelakangi keprihatinannya karena generasi muda yang semakin tidak mengenal bahasa daerahnya (Jawa) dengan baik dan benar. Dia menghasilkan aplikasi kamus yang menerjemahkan kata bahasa Indonesia ke

Bahasa Jawa dengan tiga tingkatannya - ngoko, krama madya dan krama inggil. Sedangkan metode yang digunakan dalam program ini adalah metode *Sequential Search* dan database *JSON (Java Script Object Notation)*.

Lamusu, Konito dan Rahandi (2013), membuat aplikasi kamus 3 bahasa – bahasa Gorontalo, bahasa Indonesia dan bahasa Inggris. pembuatan aplikasi ini selain dilatarbelakangi oleh kenyataan bahwa bahasa daerah (Gorontalo) mulai ditinggalkan oleh masyarakat penuturnya, juga karena kesadaran penulisnya akan kekayaan adat dan budaya Gorontalo. Penulis ingin memudahkan para pendatang/wisatawan, baik yang berasal dari daerah lain maupun dari luar negeri untuk mengenal dan berinteraksi dengan masyarakat asli Gorontalo. Aplikasi ini menggunakan metode Hashmap yang digunakan untuk efisiensi pencarian kata.

Ely, Simon dan Irwansyah (2013) menghasilkan aplikasi kamus lima bahasa daerah di Indonesia yaitu bahasa Melayu, Minang, Jawa, Sunda dan Batak. Sedangkan Sembiring (2013) menghasilkan Aplikasi Kamus Bahasa Indonesia – Karo Online Berbasis Web. Sebelumnya, pada tahun 2012, Priharyanto juga sudah membuat aplikasi kamus dwi bahasa, yaitu kamus bahasa Indonesia-Jawa. Berbeda dengan karya Farida (2015) yang menghasilkan terjemahan kata bahasa Indonesia dalam bahasa Jawa dengan 3 tingkatan sosialnya, karya Priharyanto (2012) menampilkan arti kata bahasa Indonesia dalam bahasa Jawa dengan aksara Jawa.

Sejauh ini, menurut pengamatan penulis, semua karya tersebut dihasilkan oleh para ahli IT, yang nota bene tidak memiliki latar belakang keilmuan di bidang bahasa. Hal ini tentu saja suatu hal yang positif dan menggembirakan sehingga harus juga ditanggapi dengan positif dalam upaya pelesatarian bahasa daerah. Lebih lanjut, banyak dari karya tersebut yang ditulis ‘hanya’ dalam rangka untuk memenuhi tugas akademis saja. Harapan penulis, kedepannya, karya-karya ini bisa dipublikasikan secara luas dan semakin lengkap dan komprehensif. Tempat publikasi aplikasi berbasis android yang paling efektif tentu saja di android market atau yang sekarang dikenal dengan Play Store. Namun sampai saat ini penulis belum menemukan aplikasi kamus bahasa Madura di Play Store.

II. KERANGKA TEORI DAN METODE PENELITIAN

2.1 Kepunahan Bahasa (daerah)

Grimes (2000) menjelaskan ada setidaknya enam gejala yang bisa dijadikan petunjuk bahwa suatu bahasa akan punah. Yang pertama adalah penurunan jumlah penutur aktif bahasa tersebut secara drastic. Yang kedua, ranah penggunaan bahasa tersebut semakin berkurang. Yang ketiga, saat bahasa ibu sudah mulai diabaikan oleh generasi mudanya. Yang keempat, ketika usaha-usaha untuk merawat identitas etnis tidak lagi menggunakan bahasa ibu. Yang kelima, ketika penutur generasi terakhir sudah tidak lagi cakap dan terampil dalam menggunakan bahasa tersebut (*understanding without speaking*) dan yang keenam, contoh-contoh mengenai semakin punahnya variasi-variasi (dialek) satu bahasa, keterancaman bahasa Kreol dan bahasa sandi.

Sedangkan *Summer Insitute of Linguistics* (SIL) (2008) menyebutkan ada dua belas faktor yang berhubungan dengan kepunahan bahasa,

- (1) kecilnya jumlah penutur,
- (2) usia penutur,
- (3) penggunaan bahasa ibu oleh anak-anak,
- (4) penggunaan bahasa lain secara reguler dalam aneka latar budaya.
- (5) perasaan identitas etnik dan sikap terhadap bahasanya secara umum,
- (6) urbanisasi kaum muda,
- (7) kebijakan pemerintah,

- (8) penggunaan bahasa dalam pendidikan,
- (9) intrusi dan eksploitasi ekonomi,
- (10) keberaksaraan,
- (11) kebersastraan, dan
- (12) kedinamisan para penutur membaca dan menulis sastra.

Menurut Darwis (2011), selain faktor-faktor tersebut di atas, ada satu tambahan faktor penting lainnya terkait dengan kepunahan bahasa daerah, yaitu tekanan dominasi bahasa lain (bahasa Indonesia dan bahasa Asing – Inggris).

Apabila kita mencermati faktor-faktor tersebut di atas, kita akan menyadari bahwa hampir semua bahasa daerah di Indonesia mengalami gejala kepunahan bahasa, termasuk bahasa Madura.

2.2 Upaya Pelestarian Bahasa Daerah dalam Perencanaan Bahasa

Pembuatan kamus adalah salah satu langkah penting dalam upaya pelestarian suatu bahasa, dalam hal ini adalah bahasa daerah (Madura). Sekalipun begitu, pembuatan kamus hanyalah sebagian dari langkah (kodifikasi) perencanaan bahasa (*language planning*). Haugen (1966) menawarkan empat langkah dalam perencanaan bahasa yang meliputi:

- a. Pemilihan (*selection*), yaitu penentuan bahasa/variasi bahasa mana yang akan dijadikan obyek.
- b. Kodifikasi (codification), yaitu pengembangan dari bahasa atau variasi bahasa yang sudah dipilih. Termasuk dalam kodifikasi adalah penentuan ragam bahasa standard dan ragam bahasa yang lainnya, pembuatan kamus, grammar dan sebagainya.
- c. Penerapan (*implementation*), yaitu usaha untuk menerapkan dan menyebarluaskan bahasa yang sudah dipilih.
- d. Pengembangan (*elaboration*), merupakan usaha terus menerus untuk memperbaharui bahasa terpilih agar selalu sesuai dengan jamannya dan selalu bisa memenuhi semua kebutuhan penggunaanya.

Fishman (1977) menambahkan bahwa perencanaan bahasa akan berhasil apabila didukung oleh semua pihak, khususnya: pemerintah atau menteri terkait, pendidik, ahli bahasa, hakim, kalangan swasta, dan rakyat sendiri. Jadi perencanaan bahasa yang baik, memerlukan partisipasi aktif dari semua pihak.

Lebih lanjut, Ferguson menyatakan bahwa ciri-ciri bahasa yang memudahkan masyarakat awam menerima perencanaan bahasa adalah sebagai berikut.

- a. Bahasa itu adalah bahasa pribumi (penduduk asli) atau bahasa ibu negara itu.
- b. Bahasa itu pernah menjadi *lingua franca* dalam negara itu dan antarnegara tetangga.

- c. Bahasa itu berpotensi (kreatif dan fleksibel) untuk perkembangan pendidikan, agama, sastra, iptek (ilmu pengetahuan dan teknologi), dan media massa.
- d. Bahasa itu mempunyai budaya yang mantap dan agung.
- e. Sejarah bahasa itu mantap dan sah.
- f. Bahasa itu mempunyai banyak bahan dokumentasi untuk dikaji.
- g. Bahasa itu mempunyai pakar tradisional dan modern.
- h. Bahasa itu mempunyai kebijakan (*policy*) perencanaan yang dilakukan oleh individu atau kelompok yang cinta bahasanya.
- i. Bahasa itu dihormati oleh pemakainya dan masyarakat pemakai kelompok lain.
- j. Bahasa itu mempunyai ciri kebangsaan atau nasional.
- k. Bahasa itu mempunyai daya tarik yang memudahkan pemakainya taat dan setia kepadanya.
- l. Bahasa itu mudah memupuk persatuan bangsa dan negara.

Pelaksanaan perencanaan bahasa ini kemungkinan besar akan mengalami hambatan yang mungkin akibat dari perencanaannya yang kurang tepat; bisa juga dari para pemegang tampuk kebijakan, dari kelompok sosial tertentu, dari sikap bahasa para penutur, maupun dari dana dan ketenagaan. Perencanaan yang kurang tepat bisa bersumber dari pengambilan kebijakan yang tidak tepat atau keliru, karena salah *mengestimasi* masalah kebahasaan yang harus diteliti.

2.3 Bahasa Madura

Bahasa Madura masuk dalam kerabat bahasa Austronesia dan digunakan oleh suku Madura. Bahasa daerah ini tidak hanya digunakan di pulau Madura saja, karena suku Madura sudah tersebar ke berbagai wilayah Indonesia. Namun yang paling utama, selain di pulau Madura sendiri, bahasa ini juga digunakan penduduk di pulau-pulau kecil di sekitar pulau Madura. Bahasa Madura juga digunakan oleh penduduk di beberapa daerah di propinsi Jawa Timur, di bagian pesisir utara (Surabaya, Pasuruan, Probolinggo, Situbondo, Bondowoso, Jember and Lumajang) dan di bagian barat dan selatan pulau Kalimantan.

Tidak seperti halnya bahasa Indonesia yang tidak mengenal adanya tingkat tutur, bahasa Madura mengenal adanya tingkatan sosial tuturan. Secara umum, tingkat tutur dalam bahasa Madura ada tiga: *enja-ia*, *engghi enten* dan *engghi bhunten*. Sedangkan jumlah bunyi vokal bahasa Madura sama dengan bahasa Indonesia, yaitu ada enam bunyi vokal. Namun bunyi vokal bahasa Madura tidak sama dengan bahasa Indonesia. Vokal dalam bahasa Madura adalah /a/, /i/, /u/, /□/, /ə/, dan /□/. Untuk bunyi konsonan dalam bahasa Madura, ada 31 bunyi. (Sofyan, 2008). Untuk sistem ejaan bahasa Madura, sayangnya sampai saat ini masih belum benar-benar seragam. Masih ada perdebatan diantara sesama ahli bahasa Madura yang sampai sekarang belum mencapai titik temu yang memuaskan. Misalnya bunyi *hamzah* harus dituliskan dengan lambang tersendiri

atau cukup menggunakan apostroph (Sofyan, 2008). Sistem ejaan yang belum standar ini menjadi salah satu kendala dalam penyusunan kamus.

2.4 Kamus

Kamus merupakan buku rujukan yang berisi kata-kata dan keterangannya. Buku ini berguna untuk membantu seseorang mengenal kata-kata baru. Selain menerangkan arti kata, kamus tertentu juga mungkin mempunyai pedoman sebutan, asal-usul (etimologi) suatu kata dan juga contoh penggunaannya dalam sebuah kalimat. Untuk memperjelas kadang kala terdapat juga ilustrasi di dalam kamus. Fungsi lain kamus adalah juga sebagai kebanggaan identitas suatu bahasa dan bangsa (Dep. Pendidikan dan Kebudayaan, 2002:499).

Saat ini kita mengenal ada berbagai macam jenis kamus. Dari kamus berbentuk buku, aplikasi kamus yang terpasang di komputer atau laptop, kamus online yang selalu tersedia di web-web tertentu sampai kamus digital yang mobile yang terpasang di *smartphone* kita.

Kamus digital mempunyai keunggulan disbanding kamus-kamus tradisional yang berbentuk buku. Dari sisi ukuran, jelas kamus digital jauh lebih praktis, karena sudah terpasang di perangkat *mobile* kita. Dari sisi berat, menambahkan aplikasi kamus dalam perangkat *mobile* tidak akan menambah berat perangkat tersebut. Sedangkan dari sisi kepraktisan, kamus digital jauh lebih unggul karena untuk mencari arti kata, kita tidak perlu lagi membuka buku kamus berlembar-lembar, cukup ketikkan kata yang akan kita cari, aplikasi akan bekerja secara otomatis untuk menemukan arti kata yang kita ketikkan.

Salah satu *operating system* paling populer untuk perangkat mobile saat ini adalah OS Android yang diproduksi oleh perusahaan Google.

2.5. Android

Android adalah sistem operasi berbasis Linux yang saat telah menjadi sistem operasi yang paling populer di telepon selular, smart phone, dan komputer tablet. Android adalah platform terbuka (*open source*) bagi para pengembang (*developer*). Sehingga para pengembang bisa menciptakan aplikasi mereka sendiri yang bisa di pasang di perangkat *mobile*.

Karena Android bersifat *open source* maka banyak pihak yang tertarik dan menggunakannya untuk perangkat mobilnya. Akibatnya, android berkembang dengan sangat cepat. Sampai kini Google telah mengeluarkan android dengan berbagai versi. Android pertama kali dikeluarkan pada tahun 2007 pada perangkat Nexus One. Kemudian android semakin dikembangkan dengan menambahkan fitur-fitur di versi-versi Android yang lebih baru; mulai dari Android 1.1 kemudian dikembangkan di versi 1.5 (*Cupcake*), 1.6 (*Donut*), Android versi 2.0/2.1 (*Eclair*), Android 2.2 Froyo (*Frozen Yoghurt*), Android

versi 2.3 (*Gingerbread*), Android versi 3.0/3.1 (*Honeycomb*), Android 4.0 *Ice Cream*, 4.1 *Jellybeans*, dan sampai yang terakhir tahun 2016 ini akan dikelarkan android terbaru, android *Marshmallow* (Safaat, 2012).

III. Analisis Penelitian

3.1 Analisis sistem, kebutuhan input, proses, output dan sistem

Mengikuti perkembangan jaman, kita tidak boleh kalah oleh arus globalisasi. Justru seharusnya, globalisasi dan modernisasi ini kita manfaatkan untuk melestarikan bahasa-bahasa daerah di Indonesia. Saat ini banyak generasi muda yang mulai meninggalkan bahasa-bahasa daerahnya masing-masing. Pada sisi lain, hampir semua generasi muda kita saat ini memegang perangkat *mobile* atau *smartphone*. Dengan teknologi, *smartphone* saat ini tidak hanya berguna sebagai alat komunikasi saja. Dengan berbagai macam aplikasi yang ada, perangkat *mobile* ini bisa diberdayakan sebagai alat untuk belajar bahasa daerah. Salah satu aplikasi yang penting untuk belajar bahasa ini adalah aplikasi kamus *mobile* di Android.

Aplikasi kamus di perangkat *mobile* ini dapat membantu pembelajaran bahasa secara praktis dan membantu melestarikan bahasa daerah (Madura). Pencarian atau terjemahan kata pada aplikasi kamus *mobile* Android dapat menjadi lebih mudah dan cepat, dikarenakan oleh algoritma dan metode yang digunakan oleh aplikasi kamus tersebut. Penelitian ini akan menerapkan *AutoCompleteTextView* dan metode *HashMap* untuk pencarian kata dalam database *SQLite*.

Kata yang diinputkan dalam aplikasi ini berupa teks dalam bahasa Indonesia atau Madura. Kemudian *AutoCompleteTextView* akan berfungsi untuk menampilkan daftar saran kata yang dicari sesuai awalan huruf yang diketikan oleh pengguna.

Ketika pengguna mengetikkan kata yang akan dicari pada field pencarian, sistem akan langsung meresponse dengan mengurutkan data dalam database sesuai dengan huruf yang diketikkan oleh pengguna. Dalam proses ini, sekaligus, sistem akan menampilkan hasil pengurutan data tersebut untuk memudahkan pengguna memilih kata yang dikehendaknya (*AutoCompleteTextView*) agar proses lebih cepat. Namun pengguna tetap bisa mengetikkan kata yang akan dicari, tanpa memilih kata-kata yang sudah ditampilkan. Proses ini berjalan terus selama pengguna mengetikkan kata yang akan dicari.

Output yang akan ditampilkan oleh aplikasi ini adalah padanan kata yang dicari dalam bahasa Indonesia dan Madura. Jika kata yang dimaksud sudah ada dalam database, maka hasil pencarian akan ditampilkan pada bidang terpisah sedangkan jika tidak ditemukan dalam database, maka akan muncul pesan yang memberitahukan bahwa kata yang dicari tidak ada.

Untuk analisa kebutuhan sistem aplikasi ini adalah sebagai berikut:

- a. Aplikasi menyediakan menu untuk pencarian kata dalam bahasa Indonesi dan Madura.
- b. Aplikasi mampu mecari padanan kata - kata Bahasa Indonesia dan Madura atau sebaliknya.
- c. Aplikasi bisa menampilkan daftar saran kata sesuai dengan huruf yang diketikan pengguna pada kolom pencarian.
- d. Aplikasi bisa menampilkan padanan kata - kata Bahasa Indonesia dan Madura atau sebaliknya jika data ditemukan dalam database. Jika tidak ditemukan dalam database, aplikasi akan menampilkan pesan pemberitahuan.

3.2 Perancangan Aplikasi Kamus

3.2.1 Use case diagram

Use case diagram ini berfungsi untuk memberikan gambara penggunaan fungsi sistem sehingga pengguna memahami fungsi sistem aplikasi yang dirancang.

3.2.2 Activity Diagram

Activity diagram memberikan gambaran langkah-langkah dalam aliran kerja sitem. Diagram ini akan menjelaskan siapa yang melakukan apa dan sampai tahap apa dan akan dilanjutkan oleh siapa.

3.2.3 Sequence diagram

Sequence diagram adalah diagram untuk menggambarkan bagaimana suatu operasi dilakukan, pesan apa yang dikirimkan dan waktu pelaksanaannya. Diagram ini bermanfaat untuk menggambarkan rangkaian langkah-langkah yang dilakukan sebagai jawaban (*response*) dari suatu event untuk menghasilkan sebuah output tertentu serta menjelaskan perubahan inernal apa saja yang terjadi selama proses tersebut.

3.2.4 Story board Kamus

Pada saat aplikasi dibuka, akan langsung muncul tampilan kamus. Pada bagian awal ini ada label judul aplikasi, yaitu Kamus bahasa Madura-Indonesia. dibawahnya ada radio button berisi pilihan kamus yang aan digunakan, Madura-Indonesia atau sebaliknya Indonesia-Madura. Diikuti dengan field kolom pencarian. Dibawah kolom tersebut disediakan tombol “cari” untuk memulai pencarian. Hasil pencarian akan langsung ditampilkan dibawah tombol tersebut. Namun jika data tidak ditemukan, maka yang muncul dibawah tombol adalah pemberitahuan. Tombol keluar untk menutup aplikasi diletakkan di bagian paling bawah.

IV. SIMPULAN

Globalisasi dan modernisasi bisa berakibat negatif pada perkembangan bahasa-bahasa daerah (Madura). namun pada sisi yang lain, globalisasi dan modernisasi bisa dimanfaatkan untuk melestarikan dan mengembangkan bahasa- bahasa daerah di Indonesia. *Smartphone* sebagai salah satu produk modernisasi bisa dimanfaatkan untuk mendekatkan kembali bahasa-bahasa daerah dengan generasi muda penuturnya. Salah satu caranya adalah dengan menghadirkan kamus bahasa daerah untuk perangkat *mobile* tersebut. Penggunaan kamus *mobile* selain lebih cepat menemukan kata yang dicari juga jauh lebih praktis dibanding penggunaan kamus konvensional.

Diharapkan dengan semakin dekatnya generasi muda dengan bahasa daerahnya, bahasa-bahasa daerah di Indonesia semakin terjaga kelestariannya.

Daftar Pustaka

- Azhar, Iqbal Nurul, 2008. Ketika bahasa madura tidak lagi bersahabat dengan kertas dan tinta (Sebuah Kajian Ethnolinguistics Ditinjau Dari Sudut Pandang Sociolinguistics) dalam *Bahasa dan Sastra dalam Berbagai Perspektif*. the Faculty of Language and Art, Yogyakarta State University dan Tiara Wacana publisher.
- Budiwiyanto, Adi. (TT). Pendokumentasian bahasa dalam upaya revitalisasi bahasa daerah yang terancam punah di Indonesia. Diambil dari laman <http://badanbahasa.kemdikbud.go.id/lamanbahasa/artikel/1823>
- Darwis, Muhammad, M.S. 2011. Nasib bahasa daerah di era globalisasi: Peluang dan tantangan, Makalah pada Workshop Pelestarian Bahasa Daerah Bugis Makassar, Balitbang Agama. Makassar.
- Departemen Pendidikan dan Kebudayaan. 2002. Kamus besar bahasa indonesia edisi ketiga. Jakarta: Balai Pustaka
- Ely, Kurniawan Dwi, dkk. 2013. Aplikasi kamus aneka bahasa daerah berbasis smartphone Android. Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 5, Pekanbaru, 2 Oktober 2013,
- Farida, Intan Nur. 2015. Aplikasi kamus indonesia-jawa berbasis android. Nusantara of Engineering. Vol. 2. No. 1 diambil dari laman <http://efektor.unpkediri.ac.id/index.php/ti/article/viewFile/90/60>
- Fishman. Joshua, A. 1977. Advances in language planning. De Gruyter
- Grimes, B. F. Ed. 1988. Ethnologue: languages of the world. Dallas, Texas: Summer Institute of Linguistics, Inc

- Lamusu, M. N. F, Koniyo, M. H., dan Rahandi, M (2013), Pembuatan aplikasi kamus 3 bahasa pada smartphone android. Jurnal Karya Ilmiah Mahasiswa Fakultas Teknik, Vol. 01 No. 01
- Lontar Madura. 2012. Marginalisasi bahasa Madura dan usaha-usaha pengembangannya. Diambil dari laman <http://www.lontarmadura.com/marginalisasi-bahasa-madura-usaha-pengembangannya/>
- Misnadin. 2011. An observation towards some phonological rules and processes in Madurese, Prosodi Volume V, Nomor 2, Juli
- Moseley, Christopher. 2010. Atlas of the world's languages in danger. UNESCO Publishing
- Mubah, A. S. 2011. Strategi meningkatkan daya tahan budaya lokal dalam menghadapi arus globalisasi. Jurnal Unair, Vol. 24 No. 4
- Priharyanto, Ibnu. 2012. Aplikasi kamus bahasa indonesia – jawa – jawa krama berbasis android. Diambil dari laman <http://repository.amikom.ac.id/files/Publikasi2010.21.0489.pdf>
- Safaat, H Nazaruddin. 2012. Android : Pemrograman aplikasi mobile smartphone dan tablet pc berbasis Android. Bandung : Informatika..
- Santoso, Teguh. (TT). Menatap masa depan bahasa daerah: studi kasus di aceh. Diambil dari laman <http://badanbahasa.kemdikbud.go.id/lamanbahasa/artikel/1291>
- Sembiring, Jhoni Pranata. 2013. Perancangan aplikasi kamus bahasa indonesia - karo online berbasis web dengan Metode Sequential Search. Pelita Informatika Budi Darma, Volume : IV, Nomor: 2
- Sofyan, Akhmad. 2008. Variasi, keunikan dan penggunaan bahasa madura, Departemen Pendidikan Nasional, Pusat Bahasa, Balai Bahasa Surabaya
- www.ethnologue.com. 2016. Madura. Diambil dari laman www.ethnologue.com/language/mad

Pengembangan WordNet Bahasa-Bahasa Daerah di Indonesia: Kasus WordNet Bahasa Jawa

Totok Suhardijanto

Universitas Indonesia
suhardiyanto@gmail.com

Abstrak

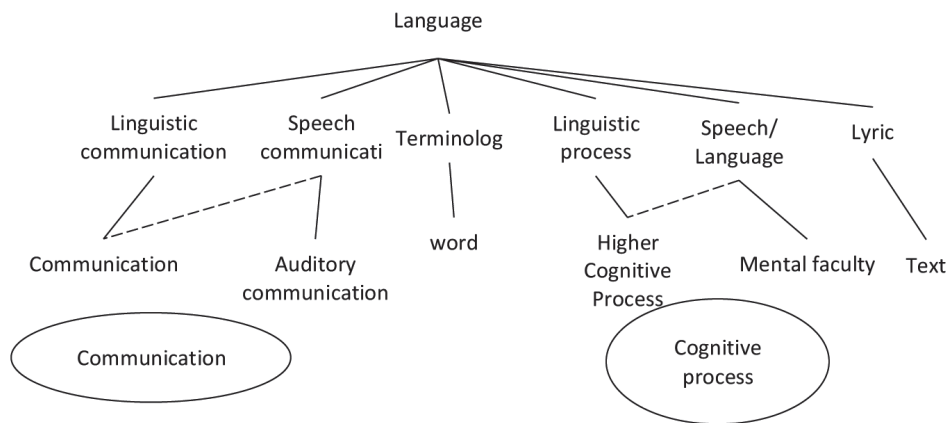
Salah satu tantangan masa kini yang dihadapi para peneliti dan pegiat kebahasaan di dunia adalah upaya penyusunan sumber daya bahasa dalam bentuk digital. Pada masa mendatang, semua bahasa, termasuk bahasa daerah di Indonesia, sebaiknya harus mempunyai pangkalan data elektronik. Dalam kerangka itulah, makalah ini menyoal tantangan dan peluang dalam pembentukan WordNet Bahasa Jawa. Wordnet merupakan pangkalan data leksikal dalam bentuk jaringan semantik. Wordnet paling luas cakupan dan penerapannya dalam pemrosesan bahasa alami, misalnya analisis sentimen, mesin terjemahan, dan aplikasi pembedaan kata.

Ada dua pendekatan yang lazim digunakan dalam membangun wordnet untuk bahasa yang baru: pendekatan penggabungan (*merge approach*) dan peluasan (*extend approach*). Pendekatan penggabungan mengambil sumberdaya ekabahasa yang ada dan memetakannya ke dalam struktur wordnet. Pendekatan penambahan menggunakan set sinonimi (sinset) dari WordNet Princeton dan menambahkan lema dari bahasa sasaran. Makalah ini membahas pendekatan mana yang digunakan dalam pengembangan WordNet bahasa Jawa dan apa kelebihan dan kekurangannya. Selain itu, dalam makalah ini dibahas aspek teknis dan leksikografis yang terkait dengan pengembangan wordnet untuk bahasa Jawa.

Kata Kunci: wordnet, sinset, metode peluasan, metode penggabungan, bahasa Jawa

I. PENDAHULUAN

Wordnet yang pertama kali diciptakan di Laboratorium Ilmu Kognitif Princeton University merupakan salah satu pangkalan data leksikal elektronik yang paling banyak dan paling luas digunakan. Data leksikal dalam pangkalan data ini disusun menurut jaringan semantiknya (Fellbaum 1998). Di dalam wordnet, kelas terbuka seperti nomina, verba, adjektiva dan adverbial dikelompokkan ke dalam set sinonim kognitif (*cognitive synonym set*) yang disingkat *synset* (dalam makalah ini selanjutnya disebut *sinset*). Sinset ini ditautkan satu sama lain dengan relasi semantis seperti sinonimi, antonimi, atau hiponimi. Wordnet versi 3.0 yang paling mutakhir saat ini memiliki lebih dari 117.000 sinset (<http://wordnet.princeton.edu/>).



Gambar 1. Subset Wordnet Princeton University

Menurut Fellbaum, Wordnet tidak hanya berkembang dari segi ukuran, tetapi juga dalam jumlah bahasa yang dimasukkan ke dalam pangkalan data wordnet di seluruh dunia. Wordnet telah mengilhami banyak peneliti untuk mengembangkannya ke dalam bahasa lain, khususnya bahasa-bahasa di Asia (lihat Bond dkk 2014; Tan dan Lim 2010; Thoongsup 2009; Isahara 2008). Banyaknya penerapan wordnet pada bahasa-bahasa di Asia menjadi salah satu alasan penyusunan wordnet bahasa Jawa ini.

Makalah ini menyajikan pengembangan Wordnet Bahasa Jawa (selanjutnya disingkat WBJ) yang dilakukan secara semi-otomatis. Hingga saat ini, leksikon semantis bahasa Jawa yang terhubung dan dapat diakses secara luas belum ada. Mengapa bahasa Jawa menjadi penting untuk diterapkan ke dalam sistem pangkalan data leksikal elektronis WordNet? Paling tidak ada beberapa alasan. Pertama, Bahasa Jawa merupakan salah satu bahasa utama di wilayah Asia Tenggara dengan total penutur berjumlah 84 juta orang penutur yang wilayah tradisionalnya meliputi Propinsi Jawa Tengah, Jawa Timur, dan Daerah Istimewa Yogyakarta. Kedua, bahasa Jawa juga memiliki tiga dialek utama, yakni timur, tengah, dan barat, serta tiga jenis ragam, yakni *ngoko*, *madya*, dan *krama*

yang tentunya cukup menantang dari segi teknologi dalam penanganan aspek dialektal dan ragam tersebut. Oleh sebab itu, dalam pengembangan WBJ ini, dipilih dialek tengah sebagai dasar karena dialek ini dianggap dialek standar bahasa Jawa (lihat Subroto dkk 1991). Pada proyek ini digunakan 5000 kosakata atau sinset inti dari Wordnet Bahasa Inggris Princeton (selanjutnya disingkat WBIP) (<http://wordnet.princeton.edu/>). Namun, untuk tahap pertama, pengembangan WBJ baru dilakukan dengan 2000 sinset. Dalam penelitian ini, WBJ disusun dengan menggunakan kamus dwibahasa dengan berdasarkan WBIP. Kamus dwibahasa ini disusun berdasarkan kamus bahasa Jawa-Inggris (Robson & Wibisono 2002). Sinset dalam WBJ disejajarkan dengan sinset WBIP melalui bahasa Inggris sebagai perantara. Dalam proses penjajaran, dilakukan pula penerjemahan manual.

II. POKOK BAHASAN DAN TUJUAN PENULISAN

Dalam makalah ilmiah, fokus pembicaraan berkenaan dengan penyusunan pangkalan data leksikal elektronik WBJ. Lebih lanjut, secara lebih detail, pada makalah ini perhatian dititikberatkan pada tiga permasalahan, yakni metode mana yang paling sesuai digunakan, aspek-aspek bahasa Jawa apa saja yang menjadi kendala dan tantangan dalam penyusunan WBJ, dan bagaimana mengatasi kendala dan tantangan tersebut.

Oleh sebab itu, makalah ilmiah ini bertujuan untuk mengungkapkan metode yang tepat dalam penyusunan WBJ. Selain itu, dalam makalah ini, juga diidentifikasi beberapa aspek bahasa Jawa yang potensial menjadi kendala dan tantangan dalam penyusunan WBJ secara teknis. Kemudian, pada makalah ini, juga dieksplorasi cara dan teknik untuk mengatasi kendala-kendala yang muncul terkait dengan bahasa Jawa.

III. PENDEKATAN DAN METODE

Pada bagian ini dibahas pendekatan dan metode yang digunakan dalam penyusunan WBJ. Bagian Pendekatan dan metode ini dipilah ke dalam dua subbagian agar pembicaraan mengenai keduanya berjalan dengan lebih sistematis, lengkap, dan mendalam.

3.1 Pendekatan

Seperti yang dikemukakan sebelumnya, dalam pengembangan wordnet untuk bahasa baru, lazim ditempuh dengan dua pendekatan: metode penggabungan dan metode perluasan. Kedua pendekatan ini mempunyai kekurangan dan kelebihan.

Dalam metode penggabungan, taksonomi bahasa baru (sinset dan relasi) dibuat dan kemudian diperakan ke struktur WBIP. Dengan pendekatan ini, diciptakan struktur wordnet baru karena pangkalan datanya disusun berdasarkan taksonomi bahasa tertentu. Meskipun demikian, bagi bahasa minor atau bahasa yang jarang diteliti, pendekatan

inilah yang paling sesuai karena sulit menemukan kamus dwibahasa dalam bahasa-bahasa seperti demikian.

Pada pendekatan peluasan, bahasa baru diterjemahkan langsung ke WBIP dengan menambahkan kosakata bahasa baru ke WBIP. Pendekatan ini mempertahankan struktur asli BWIP karena bahasa baru diproyeksikan ke WBIP dengan menggunakan struktur aslinya. Pendekatan ini sederhana dan hasilnya dapat disejajarkan secara otomatis dengan wordnet bahasa-bahasa lainnya (Clark dkk 2008a dan 2008b). Meskipun demikian, dalam hal bahasa Jawa, beberapa konsep bahasa Jawa yang tidak ada di dalam WBIP tidak akan ada di dalam pangkalan data leksikon yang disusun.

3.2 Metode

Seperti disebutkan pada bagian sebelumnya, pada makalah ini, digunakan pendekatan peluasan untuk membangun WBJ dengan menerjemahkan sinset WBIP ke dalam bahasa Jawa. Pada tahap pertama, dikembangkan WBJ dengan 2000 sinset yang berasal dari 5000 makna dasar kata di dalam WBIP. Kata yang berjumlah 5000 tersebut merupakan makna kata yang paling sering digunakan dalam bahasa Inggris.

a [distant%3:00:01::] [distant] separated in space or time
a [distinct%3:00:00::] [distinct] easy to perceive
a [domestic%3:00:00::] [domestic] concerning the internal affairs of a nation
a [domestic%3:01:00::] [domestic] relating to the home
n [figure%1:06:01::] [figure] design, pattern
n [figure%1:21:00::] [figure] amount of money expressed numerically
n [file%1:06:00::] [file] steel hand tool
n [file%1:06:01::] [file] file cabinet, filing cabinet
n [file%1:10:00::] [file] data file
n [file%1:14:00::] [file] single file, Indian file
v [amuse%2:32:00::] [amuse] make laugh
v [answer%2:32:00::] [answer] reply, respond
v [apologise%2:32:00::] [apologize]
v [appeal%2:41:04::] [appeal] challenge a decision
v [appear%2:30:00::] [appear] come into view

Gambar 2 Contoh Daftar Kosakata Dasar WordNet Princeton University

Penerjemahan makna dasar kata ke dalam bahasa Jawa dilakukan secara manual dengan menggunakan aplikasi berbasis web. Penerjemahan dilakukan oleh penutur asli bahasa Jawa yang terdiri atas peneliti dan mahasiswa jurusan Jawa di lingkungan FIB Universitas Indonesia.

IV. PEMBAHASAN

Pada bagian ini, akan dibahas isu-isu yang muncul baik terkait teknis penyusunan maupun terkait sifat dan karakter kedua bahasa yang terkait dengan proyek penyusunan wordnet ini, yakni bahasa Jawa dan bahasa Inggris. Oleh karena itu, pada bagian ini, akan dibahas dua hal, yakni penerjemahan sinset dan isu-isu kebahasaan.

4.1 Penerjemahan Sinset

Pada bagian ini dibahas permasalahan yang muncul dalam proses pengembangan WBJ. Ada dua isu penting terkait dengan proses pengembangan WBJ, yakni yang terkait dengan struktur WBIP dan yang terkait dengan karakteristik bahasa Jawa. Meskipun demikian, dalam makalah ringkas ini hanya akan dibahas isu pertama. Pada makalah versi lengkap, isu terkait karakteristik bahasa Jawa akan dibahas secara panjang lebar. Selanjutnya, pada bagian ini membahas bagaimana menangani konsep yang tidak mempunyai padanan dalam leksikon bahasa Jawa. Untuk mengatasi hal ini, apakah mungkin menggunakan kata majemuk atau frasa? Sebagai contoh,

colt -- (a young male horse under the age of four)

mare -- (female equine animal)

stallion -- (uncastrated adult male horse)

Dalam hal ini, frasa atau kata majemuk akan digunakan sebagai padanan unsur leksikal. Misalnya, untuk menerjemahkan *mare* digunakan frasa *jaran wedok* yang secara harfiah bermakna ‘kuda betina dewasa’).

Isu lain yang tidak kalah penting adalah mana yang lebih baik: penerjemah sinset dengan sinset atau penerjemahan kata? Dalam bahasa Inggris, misalnya, konsep goat terdiri atas tiga kata: *goat*, *caprine*, *animal*.

WBIP:

- *goat*, *caprine animal* -- (any of numerous agile ruminants related to sheep but having a beard and straight horns)

Dalam bahasa Jawa, padanan sinset *goat* adalah

WBJ:

- a) *wedhus* ‘goat’ (plain word)
- b) *menda* ‘goat’ (high polite word)
- c) *kambing* ‘caprine animal’

Jadi, pertanyaannya adalah bagaimana memilih kosakata Jawa yang tepat untuk padanan bahasa Inggrisnya? Hal ini penting karena tiga kosakata bahasa Inggris digunakan dalam konteks yang berbeda dalam bahasa Jawa. Untuk alasan ini, dalam hal ini, penerjemahan kata dengan kata merupakan solusi terbaik.

- 1) goat → a) *wedhus* (plain word)
→ b) *menda* (high polite word)
- 2) *caprine animal* → c) *kambing* (generic term)

Dengan demikian, prinsip penerjemahan yang digunakan dalam penyusunan WBJ adalah sebagai berikut.

- Kata tunggal hanya digunakan untuk konsep yang terleksikalisasi dalam bahasa Jawa: *milk* > *susu*
- Kata majemuk/frasa digunakan untuk beberapa konsep yang tidak ditemukan padanan katanya dalam bahasa Jawa: *stallion* > *jean lanang*
- Pemadanan sinset ke sinset digunakan untuk sinset bahasa Jawa bersesuaian dengan sinset dalam sistem WBIP.
- Pemadanan kata dengan kata digunakan untuk kosakata bahasa Jawa yang bersesuaian dengan kata-kata yang ada pada sinset di dalam WBIP.

4.2 Isu-Isu Kebahasaan

Paling tidak ada tiga isu terkait bahasa yang perlu ditangani dalam penyusunan WBJ. Isu-isu tersebut adalah bagaimana menangani masalah dialek atau subdialek, ragam, dan tidak adanya konsep ekuivalen di antara sinset WBJ dan WBIP.

Sebagaimana diketahui, bahasa Jawa memiliki tiga dialek utama, yakni dialek TIMUR, TENGAH, dan BARAT. Terkait dengan dialek atau subdialek, seperti disebutkan pada bagian sebelumnya, fokus pengembangan WBJ pada tahap awal akan difokuskan pada salah satu dialek saja, yakni dialek TENGAH. Meskipun demikian, untuk pengembangan selanjutnya, masalah dialek dapat ditangani dengan penggunaan fitur *tag* yang disediakan oleh wordnet untuk menganotasi kata lokal atau dialektal. Sebagai contoh,

koen -- (you, 2SG) + (EAST)
kowe -- (you, 2SG) + (CENTRAL)
rika -- (you, 2SG) + (WEST)

Seperti yang telah disebutkan sebelumnya ragam atau gaya bahasa Jawa terdiri atas tiga, yakni *ngoko*, *madya*, dan *krama* (Subroto et al. 1991). Sejalan dengan dialek, terkait dengan ragam atau gaya, dapat pula diterapkan penggunaan tag untuk menganotasi kata berdasarkan gaya. Sebagai contoh,

kowe	-- (you, 2SG) (CENTRAL) + (LOW)
sampeyan	-- (you, 2SG) (CENTRAL) + (HIGH)
panjenengan	-- (you, 2SG) (CENTRAL) + (POLITE HIGH)

Pada tahap pertama penyusunan digunakan 5000 makna kata dasar dalam pengembangan WBJ. Meskipun tidak banyak, dijumpai beberapa konsep yang tidak

muncul pada salah satu bahasa, yakni bahasa Jawa atau bahasa Inggris. Untuk konsep-konsep yang tidak muncul baik dalam WBIP, akan ditambahkan sinset baru. Sebagai contoh,

WBIP

rice -- (grains used as food either unpolished or more often polished)

bahasa Jawa

beras	'rice grain' (plain word)
wos	'rice grain' (polite word)
sega	'cooked rice' (plain word)
sekul	'cooked rice' (polite word)
pari	'rice plant' (plain word)
pantun	'rice plant' (polite word)
upa	'a single cooked rice seed'
menir	'broken rice'

V. SIMPULAN

Pada proses penyusunan WBJ, digunakan pendekatan peluasan karena faktor kemudahan dan kemungkinannya untuk dihubungkan dengan wordnet bahasa lain di dunia. Pada tahap pertama, disusun 2000 sinset bahasa Jawa berdasarkan 5000 makna dasar yang diperoleh dari laman web Wordnet Lexical Database for English (<http://wordnet.princeton.edu/>).

Dalam proses penyusunan, pertama kali dilakukan pengumpulan data dari kamus bahasa Jawa seperti *Bausastra Jawa* (Tim Penulis Balai Bahasa Yogyakarta 2011) dan *Javanese-English Dictionary* (Robson dan Wibisono 2002). Dengan data kamus tersebut, disusunlah kamus terjemahan yang berisi kosakata bahasa Inggris dan padanan katanya dalam bahasa Jawa. Kemudian, pada tahap berikutnya, dilakukan penerjemahan antar-sinset dengan menggunakan aplikasi web.

Penyusunan WBJ merupakan salah satu upaya untuk mempertautkan pangkalan data bahasa dengan bahasa-bahasa lain di dunia. Selain itu, penciptaan WBJ merupakan salah satu upaya melestarikan bahasa Jawa dengan menyimpan salah aspeknya ke dalam bentuk elektronik atau digital. Dengan demikian, sumber daya elektronik bahasa Jawa akan terwujud dan makin kaya dengan keberadaan WBJ.

Daftar Pustaka

- Bond, F., Lim, L.T., Tang, E.K. and Riza, H. (2014). "The combined Wordnet Bahasa." In Siaw-Fong Chung and Hiroki Nomoto, eds. *Current Trends in Malay Linguistics, NUSA* 57:83-100, 2014.
- Budanitsky, A., & Hirst, G. (2006). Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics*, 32 (1), 13-47.
- Clark, P., Fellbaum, C., Hobbs, J., Harrison, P., Murray, W. and J. Thompson. (2008). Augmenting WordNet for Deep Understanding of Text. *ACL-SigSem*, Venice, Italy.
- Clark, P., Fellbaum, C., and Hobbs, J. (2008). Using and Extending WordNet to Support Question Answering. Dalam *Proceedings of the Fourth Global WordNet Conference*, eds. A. Tanacs, D. Csendes, V. Vincze, C. Fellbaum and P. Vossen. Hungary: University of Szeged, pp. 111-119.
- Fellbaum, C. (1998). *Wordnet: An electronic lexical database*. MIT Press, Massachusetts, 1998.
- Isahara, H., Bond, F., Uchimoto, K., Utiyama, M., & Kanzaki, K. (2008). "Development of the Japanese WordNet". In *LREC*, 2008.
- Morato, J., Marzal, M. A., Lloréns, J., & Moreiro, J. (2004). Wordnet applications. In *Global Wordnet Conference* (Vol. 2, pp. 270-278).
- Robson, S. And Wibisono, S. 2002. *Javanese-English Dictionary*. Periplus, Singapore.
- Subroto, D.E., Soenardji, dan Sugiri. (1991). *Tata Bahasa Deskriptif Bahasa Jawa*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Tan, P.P.P.L. and Lim, N.R.T. (2010). "Toward a Filipino Wordnet." Diambil dari <http://www.dlsu.edu.ph/research/centers/adric/nlp/downloads/paper7.pdf>
- Thoongsup, S., Robkop, K., Mokarat, C., Sinthurahat, T., Charoenporn, T., Sornlertlamvanich, V., and Isahara, H. (2009). "Thai Wordnet Construction." *Proceedings of the 7th Workshop on Asian Language Resources, ACL-IJCNLP 2009*, pages 139-144, Suntec, Singapore, 6-7 August 2009.
- Tim Penulis Balai Bahasa Yogyakarta. (2011). *Kamus Basa Jawa (Bausastra Jawa). Edisi Kedua*. Yogyakarta: Kanisius.
- Wordnet Lexical Database for English. (2016). <http://wordnet.princeton.edu/>. Diakses pada 21 Juni 2016.

Beberapa Kendala Pengayaan Kosakata Bahasa Indonesia dari Bahasa Daerah

Asep Rahmat Hidayat

Balai bahasa Jawa Barat
kang.abu2@gmail.com

Abstrak

Indonesia memiliki 719 bahasa daerah (<http://www.ethnologue.com/country/ID>). Jumlah itu tentu saja merupakan sumber daya yang sangat besar untuk mengembangkan kosa kata bahasa Indonesia. Bahasa Indonesia yang berasal dari bahasa Melayu Riau berkembang pesat dan semakin jauh dari akar bahasanya itu. Jika melihat perkembangan jumlah lema dalam KBBI, peran bahasa daerah cukup signifikan dalam pengayaan KBBI.

Jika kita berandai-andai ada satu kata yang bersinonim dalam semua bahasa daerah, satu kata itu dapat menambah 719 lema dalam KBBI. Hal itu tentu tidak perlu dilakukan karena hanya akan menambah sinonim tetapi tidak menambah konsep. Akan tetapi, itu menggambarkan secara konkret potensi bahasa daerah dalam memperkaya kosa kata bahasa Indonesia.

Di samping potensi besar itu tersimpan beragam masalah dalam proses penyerapan bahasa daerah menjadi bahasa Indonesia. Beberapa kendala berkaitan dengan hal itu antara lain, perbedaan sistem fonologi dan morfologi. Dalam bahasa Sunda sebagai contoh terdapat gabungan konsonan eu. Sementara, Pedoman Ejaan Bahasa Indonesia menyatakan hanya ada empat gabungan vokal yang disebut diftong, yaitu ai, au, ei, oi. Pedoman itu tentu mengakibatkan kata-kata yang penggunaannya tinggi, seperti keukeuh, sulit untuk menjadi kosa kata bahasa Indonesia.

Tulisan ini berupaya untuk menguraikan kendala-kendala yang ada ketika kosa kata bahasa daerah berpotensi menjadi kosa kata bahasa Indonesia. Pembahasan berfokus pada bahasa Sunda.

I. PENDAHULUAN

Indonesia memiliki 719 bahasa Daerah (<http://www.ethnologue.com/country/ID>). Jumlah itu tentu saja merupakan sumber daya yang sangat besar untuk mengembangkan kosa kata bahasa Indonesia. Bahasa Indonesia yang berasal dari bahasa Melayu Riau berkembang pesat dan semakin jauh dari akar bahasanya itu. Jika melihat perkembangan jumlah lema dalam KBBI, seharusnya peran bahasa Daerah cukup signifikan dalam pengayaan KBBI. Akan tetapi ternyata dari 90.049 lema dalam KBBI Edisi Keempat (2008) “hanya” 3.592 lema yang berasal dari bahasa Daerah atau lebih kurang 3,99% (Budiwiyono, 2012).

Jika kita berandai-andai ada satu kata yang bersinonim dalam semua bahasa Daerah, satu kata itu dapat menambah 719 lema dalam KBBI. Hal itu tentu tidak perlu dilakukan karena hanya akan menambah sinonim tetapi tidak menambah konsep. Akan tetapi, itu menggambarkan secara konkret potensi bahasa Daerah dalam memperkaya kosa kata bahasa Indonesia.

Di samping potensi besar itu mungkin tersimpan beragam masalah dalam proses penyerapan bahasa Daerah menjadi bahasa Indonesia. Hal itu mengakibatkan kontribusi bahasa Daerah dalam memperkaya bahasa Indonesia masih kecil.

II. BAHASAN

Peraturan Pemerintah Republik Indonesia Nomor 57 Tahun 2014 tentang Pengembangan, Pembinaan, dan Pelindungan Bahasa dan Sastra, serta Peningkatan Fungsi Bahasa Indonesia, Pasal 12 ayat 1 mengamanatkan pengembangan bahasa Daerah untuk meningkatkan fungsinya antara lain sebagai sarana pendukung Bahasa Indonesia dan sumber pengembangan Bahasa Indonesia.

Beberapa kriteria penyerapan bahasa Daerah adalah kekerapan penggunaan kosakata bahasa daerah oleh wartawan di media massa, kekerapan penggunaan kosakata bahasa daerah oleh penulis atau sastrawan dalam karangannya, kekerapan penggunaan kosakata bahasa daerah oleh tokoh publik, dan ketersediaan konsep baru pada kosakata bahasa daerah yang tidak dimiliki oleh bahasa Indonesia (Budiwiyono, 2012). Sementara itu, bahasa Daerah untuk memadankan istilah asing, harus memenuhi syarat paling tepat dengan konsep termaksud, singkat, berkonotasi baik, eufonik, dan seturut kaidah (Pusat Bahasa, 2008).

Beberapa kendala pengayaan kosakata bahasa Indonesia dari bahasa Daerah tentu berkaitan dengan semua faktor itu. Selain itu, kendala berupa perbedaan sistem linguistik antara bahasa Indonesia dan bahasa Daerah.

2.1 Kendala Fonologis

Bahasa Sunda, sebagai contoh, memiliki konsonan: *h, n, c, r, k, d, t, s, w, l, p, j, y, ny, m, g, b, ng* dan vokal *a, i, u, é, o, e, eu*. Sementara itu, ada empat bunyi sengau yang dimiliki, yaitu, *n, ny, m, dan ng/nga*.

Bahasa Sunda memiliki vokal *eu* (Ö). Sementara, Pedoman Ejaan Bahasa Indonesia menyatakan hanya ada empat gabungan vokal yang disebut diftong, yaitu *ai, au, ei, oi*. Pedoman itu tentu mengakibatkan kata-kata yang penggunaannya cukup tinggi, seperti *kadeukeuh, keukeuh* sulit untuk menjadi kosa kata bahasa Indonesia. Penyesuaian bisa dilakukan secara fonetis dengan mengubah kadedeh [kad@d@h], kekeh [k@k@h], hanya tentu penutur jati akan mempermasalahkan hal tersebut.

2.3 Kendala Morfologis

Beberapa bahasa Daerah memiliki proses nasalisasi dalam pembentukan kata. Pembentukan kata kerja yang paling penting dan paling biasa dalam bahasa Sunda adalah dengan cara nasalisasi (Coolsma, 1983). Proses morfologis itu memberikan aktivitas tertinggi pada sebuah kata yang biasa didapat oleh kata kerja.

Pembentukan kata kerja dalam bahasa Sunda dilakukan melalui beberapa pola berikut:

1. nasalisasi, seperti *sangu* → *nyangu* ‘menanak nasi’;
2. nasalisasi dan pengimbuhan akhiran *-an*, seperti *baju* → *ngabajuan* ‘memakaikan baju’ ;
3. nasalisasi dan pengimbuhan akhiran *-keun*, seperti *denge* → *ngadengekeun* ‘mendengarkan’ ;
4. pengimbuhan *mang-* dan *-keun meuli* → *mangmeulikeun* ‘membelikan’ ;
5. pengimbuhan *nyang-*, seperti *hareup* → *nyanghareup* ‘menghadap’;
6. pengimbuhan *nyang* dan *-an/-keun*; seperti *hareup* → *nyanghareupan* ‘menghadapi’;
7. pengimbuhan *barang-*, seperti *gawe* → *baranggawe* ‘bekerja’;
8. pengimbuhan *-an*, seperti *anak* → *anakan* ‘beranak’;
9. pengimbuhan *-eun*, seperti *sidu* → *sisidueun* ‘cegukan’;
10. pengimbuhan *pi-* dan *-eun*, seperti *untung* → *piuntungeun* ‘mendatangkan keuntungan’;
11. pengulangan konsonan awal dan vokalnya, seperti *beresih* → *beberesih* ‘bersih-bersih’;
12. pengulangan konsonan awal dan vokalnya, dengan *-an* atau pengulangan *-an*, seperti *salam* → *sasalaman* ‘bersalaman’;
13. pengimbuhan *pa-*, seperti *pisah* → *papisah* ‘berpisah’;
14. pengimbuhan *pa-* dgn *-an*, seperti *cengkad* → *pacengkadan* ‘berselisih, perselisihan’;
15. pengimbuhan *silih-*, seperti *banting* → *silihbanting* ‘saling banting’;

16. pengimbuhan *si-*, seperti *deang* → *sideang* ‘berdiang’;
17. pengimbuhan *ti-*, seperti *tajong* → *titajong* ‘tersandung’;
18. pengimbuhan *-um-*, seperti *rasa* → *rumasa* ‘merasa’;
19. pengimbuhan *-in-*, seperti *tangtu* → *tinangtu* ‘tentu, pasti’;

Proses nasalisasi memiliki pola sebagai berikut. Kata-kata yang berhuruf awal konsonan *t* akan mengalami proses nasalisasi menjadi *n*, seperti *tenjo* → *nenjo* ‘melihat’. Kata-kata yang berhuruf awal konsonan *c*, *j*, dan *s* akan mengalami proses nasalisasi menjadi *ny*, seperti *celuk* → *nyeluk* ‘memanggil’, *jieun* → *nyieun* ‘membuat’, dan *sawah* → *nyawah* ‘menggarap sawah’. Kata-kata yang berhuruf awal konsonan *p* akan mengalami proses nasalisasi menjadi *m*, seperti *piceun* → *miceun* ‘membuang’. Kata-kata yang berhuruf awal selain yang sudah tersebut akan mengalami proses nasalisasi menjadi *ng*, seperti *omong* → *ngomong* ‘berbicara’.

Bahasa Indonesia tidak memiliki sistem nasalisasi seperti bahasa Daerah. Kosakata bahasa Daerah yang diserap bahasa Indonesia idealnya hanya kosakata dasar tanpa nasalisasinya. Namun, pada penutur jati bahasa yang diserap timbul anggapan adanya pergeseran konsep atau makna dari kosakata tersebut. Di sisi lain, jika kosakata tersebut dipertahankan dengan nasalisasinya dengan asumsi diserap sebagai satu kata utuh, kosakata bahasa Indonesia akan memiliki banyak kosakata bernasal seperti itu. Sebagai contoh, kata *burit* dari bahasa Sunda sudah tercatat dalam KBBI Edisi Ketiga. Kata itu sebetulnya tidak perlu dicatat dalam KBBI karena tidak unik dan bersinonim dengan *sore*, *senja*, yang unik justru kata turunannya, yaitu *ngabuburit*. Ketika kata *ngabuburit* diekstraksi menjadi kata dasar *burit* kemudian diberi imbuhan *me-* menjadi *memburit* untuk konsep *ngabuburit* jelas terasa “aneh” bahkan bisa ambigu dengan kata yang sama yang memiliki konotasi negatif, yaitu *memburit* ‘bersetubuh sesama lelaki’. Akhirnya, kata *ngabuburit* dicatat dalam KBBI IV dengan bentuk dan makna yang sama dengan bahasa sumber, tentu dengan asumsi imbuhan *nga-* tidak dianggap sebagai imbuhan tetapi dianggap melekat satu kata dengan kata dasarnya.

2.3 Kendala Semantis

Bahasa-bahasa Daerah di Indonesia secara linguistik berkerabat karena berasal dari satu rumpun bahasa. kekerabatan itu mewujud dalam beragam hal, seperti fonologi, morfologi, dan semantik. Beberapa kata yang secara semantik bersinonim memiliki bentuk yang berbeda. Beberapa kata yang bentuknya mirip atau sama memiliki makna yang berbeda.

Kesinoniman dapat menjadi pemer kaya bahasa Indonesia, meskipun dari sudut pandang perkembangan konsep bahasa, tidak berpengaruh terhadap pemer kayaan konsep. Setakat ini, bahasa Daerah dimanfaatkan untuk memadankan istilah asing dengan syarat antara lain lebih singkat dari pilihan padanan yang tersedia. Secara umum, kriteria lebih singkat ini tentu dapat digunakan untuk memper kaya kosakata umum bahasa Indonesia.

Sebagai contoh, bahasa Indonesia memiliki kosakata *cuci tangan*, *cuci muka*, sementara bahasa Sunda memiliki kosakata untuk kedua konsep itu, *sibanyo*, *sibeungeut* yang lebih pendek suku katanya. Jadi, kedua kata itu dapat ditawarkan sebagai sinonim, tetapi tentu tidak menambah perbendaharaan konsep bahasa Indonesia.

Contoh lain adalah kedua kata yang kerap digunakan oleh pengguna bahasa, yaitu *keukeuh* dan *kadeudeuh*. Kekerapan penggunaan kata *keukeuh* cukup tinggi terutama di kalangan remaja, tetapi kata itu belum berterima dalam bahasa Indonesia karena secara semantis dianggap sama dengan kata *kukuh*. Sama halnya dengan kata kokoh yang dirujuk silang ke *kukuh*. Padahal, kata *kukuh*, kokoh, dan *keukeuh* yang berkerabat dekat bentuk dan maknanya, bisa saja dimaknai dan digunakan untuk konteks yang berbeda.

III. SIMPULAN

Bahasa Daerah, secara legal telah diamanatkan sebagai sumber pemer kaya bahasa Indonesia. Namun, perbedaan sistem dalam bahasa Daerah dan Bahasa Indonesia dapat menjadi kendala dalam upaya pemer kayaan bahasa Indonesia yang bersumber dari bahasa Daerah. Oleh karena itu, diperlukan sistem atau pola penyerapan yang lebih sistematis dan menyeluruh untuk “memperlakukan” kosakata-kosakata bahasa Daerah yang layak untuk menjadi warga bahasa Indonesia.

Daftar Pustaka

- Budiwiyanto, Adi. 2012. *Kontribusi Kosakata Bahasa Daerah dalam Bahasa Indonesia*. <http://badanbahasa.kemdikbud.go.id/lamanbahasa/content/kontribusi-kosakata-bahasa-daerah-dalam-bahasa-indonesia>. Diakses 1 Juli 2016, pukul 8.49.
- Coolsma, S. 1983. *Tata Bahasa Sunda*. Jakarta: Djambatan
<http://www.ethnologue.com/country/ID>
- Peraturan Pemerintah Republik Indonesia Nomor 57 Tahun 2014 tentang Pengembangan, Pembinaan, dan Pelindungan Bahasa dan Sastra, serta Peningkatan Fungsi Bahasa Indonesia.
- Pusat Bahasa. 2008. *Pedoman Umum Pembentukan Istilah Edisi Ketiga*. Jakarta: Pusat Bahasa.
- Pusat Bahasa. 2012. *Kamus Besar Bahasa Indonesia Edisi Keempat*. Jakarta: PT Gramedia Pustaka Utama.

Evaluasi Kamus Bilingual Kamus Bahasa Daerah *Using-Indonesia* sebagai Upaya Pemertahanan Eksistensi Bahasa Using di Kabupaten Banyuwangi

Fajar Erikha, S.Psi. dan Satwiko Budiono, S.Hum.

Departemen Linguistik, Fakultas Ilmu Pengetahuan Budaya, Universitas Indonesia
fajar.erikha51@ui.ac.id dan satwiko.iko@gmail.com

Abstrak

Sejak kemunculannya pada tahun 2002, *Kamus Bahasa Daerah Using-Indonesia* yang dibuat oleh Hasan Ali sudah sulit ditemukan di beberapa wilayah di Banyuwangi. Padahal, pada tahun 2007 Pemerintah Kabupaten Banyuwangi telah menetapkan bahwa muatan lokal untuk jenjang sekolah dasar dan menengah adalah bahasa Using. Hal tersebut seharusnya membuat keberadaan kamus bilingual Using-Indonesia menjadi semakin dibutuhkan masyarakat. Melihat kondisi demikian, sekiranya diperlukannya pembaharuan *Kamus Bahasa Daerah Using-Indonesia* itu sendiri selain mendistribusikannya kembali ke masyarakat. Maka dari itu, pada penelitian ini akan mencoba melihat kembali struktur dari kamus bilingual itu sendiri dilihat dengan menggunakan kriteria kamus bilingual dari beberapa ahli, seperti Atkins & Rundell (2008) dan Kridalaksana (2003). Evaluasi ini perlu dilakukan karena latar belakang dari pembuat *Kamus Bahasa Daerah Using-Indonesia* yang merupakan seorang linguis otodidak (Arps, 2010: 237) sehingga bisa saja ada beberapa kriteria kamus bilingual yang masih belum terpenuhi dari kriteria kamus bilingual secara teori linguistik. Selain itu, struktur Kamus Bahasa Daerah Using-Indonesia juga akan dibandingkan dengan kriteria kamus bilingual dari *Summer Institute of Linguistics International* (SIL International) dalam buku *Making Dictionaries* (2000). Tentu saja, manfaat dari penelitian ini adalah dapat menjadi bahan evaluasi untuk membuat *Kamus Bahasa Daerah Using-Indonesia* yang baru dengan disesuaikan pada kegunaannya di masyarakat.

Kata kunci: evaluasi, struktur, leksikografi, kamus bilingual, dan Kamus Bahasa Daerah Using-Indonesia.

I. PENDAHULUAN

Kamus Bahasa Daerah Using-Indonesia (2002) mempunyai peranan penting di Kabupaten Banyuwangi mulai sejak awal kemunculannya yang ditulis oleh Hasan Ali. Selain sebagai salah satu hasil kodifikasi dari bahasa Using sendiri, kamus bilingual tersebut mempunyai andil dalam menyebarluaskan bahasa Using di Kabupaten Banyuwangi. Pasalnya, tidak semua masyarakat yang tinggal di Kabupaten Banyuwangi merupakan masyarakat Using. Dalam buku *Geografi Dialek Banyuwangi*, Soetoko (1981) mengungkapkan bahwa masyarakat yang tinggal di Kabupaten Banyuwangi sangat beragam. Misalnya, dari suku Jawa, Bali, Bugis, dan Madura. Terlebih lagi, Pemerintah Kabupaten Banyuwangi (2007) menetapkan bahwa muatan lokal untuk jenjang sekolah dasar dan menengah adalah bahasa Using. Ini senada dengan temuan Lew (2016) yang menyatakan bahwa kamus merupakan alat yang belajar menulis dan belajar berbahasa kedua yang efektif. Hal tersebut membuat keberadaan *Kamus Bahasa Daerah Using-Indonesia* semakin dibutuhkan masyarakat untuk dapat mempelajari bahasa Using. Namun, kondisi di lapangan justru memperlihatkan hal yang sebaliknya. Kamus Bahasa Daerah Using-Indonesia sekarang ini sudah susah dicari di toko buku manapun di Kabupaten Banyuwangi. Kamus bilingual tersebut sudah tidak dicetak kembali. Dapat dikatakan di sini, sangat sedikit masyarakat yang mempunyai kamus bilingual tersebut. Padahal, masyarakat sangat membutuhkan *Kamus Bahasa Daerah Using-Indonesia* untuk menjadi dasar acuan mengenal bahasa Using.

Kondisi demikian membuat diperlukannya pendistribusian kembali kamus bilingual Using-Indonesia ke masyarakat di Kabupaten Banyuwangi dengan adanya evaluasi dalam hal strukturnya. Evaluasi kamus ini dirasa penting selain karena memang tahun terbitan dari *Kamus Bahasa Daerah Using-Indonesia* yang sudah lama sejak tahun 2002, alasan lain harus adanya evaluasi disebabkan penyusun kamus bilingual ini merupakan seorang linguist otodidak (Arps, 010: 237). Hal tersebut sekiranya sedikit bertentangan dengan persyaratan kemampuan penyusun kamus yang diungkapkan oleh Kridalaksana dalam buku yang berjudul *Rintisan Kajian Leksikologi dan Leksikografi*. Dalam hal ini, Kridalaksana (2003: xii) menyebutkan bahwa ada beberapa persyaratan kemampuan penyusun kamus. Salah satu persyaratan kemampuan penyusun kamus tersebut adalah kemampuan linguistik yang memadai sesuai tingkat tanggung jawab dalam program dan bidang yang akan disajikan. Dengan kata lain, sebuah kamus idealnya dibuat oleh seorang linguist terlatih yang memiliki kepekaan bahasa yang baik. Selain itu, pembuatan kamus ini juga harus didasarkan melalui penataran untuk melatih para penyusun kamus mulai dari penelitian leksikologi dan leksikografi yang mencakup dari korpus yang ada sampai ke pemeriksaan naskah akhir. Dengan begitu, Hasan Ali selaku pembuat kamus bilingual Using-Indonesia dirasa belum memenuhi persyaratan kemampuan pembuat kamus yang memadai karena ia tergolong linguist otodidak sehingga perlu dilakukan evaluasi yang didasarkan pada kriteria kamus bilingual berdasarkan teori leksikologi dan leksikografi.

Sehubungan dengan hal tersebut, penulis ingin melakukan evaluasi *Kamus Bahasa Daerah Using-Indonesia* yang disesuaikan atau dicocokkan dengan kriteria kamus bilingual menurut beberapa ahli linguistik. Beberapa ahli tersebut adalah Atkins & Rundell (2008) dalam buku *The Oxford Guide to Practical Lexicography* dan Kridalaksana (2003) dalam buku *Rintisan Kajian Leksikologi dan Leksikografi*. Tidak hanya itu, pada tulisan ini pun evaluasi akan dilihat berdasarkan kriteria kamus bilingual dari SIL Internasional dalam buku *Making Dictionaries* (2000). Pendekatan kriteria kamus bilingual didasarkan oleh ketiga sumber yang telah dijelaskan untuk membuat evaluasi kamus semakin komprehensif dan menyeluruh. Adanya evaluasi ini dapat menjadi pijakan awal atau bahan referensi jika *Kamus Bahasa Daerah Using-Indonesia* akan direvisi dan disebarluaskan kembali ke masyarakat di Kabupaten Banyuwangi. Evaluasi tersebut juga dimaksudkan agar masyarakat dipermudah saat mempelajari dan menggunakan kamus bilingual ini.

Selain itu, kebutuhan akan evaluasi kamus ini disebabkan oleh adanya beberapa kendala dalam menggunakan sebuah atau beberapa kamus. Salah satu kendala yang sering dihadapi adalah seringkali isi kamus dibentuk hanya sebatas pertimbangan para leksikografernya. Beberapa keputusan yang dibuat para leksikografer ini rawan mengarah kepada subjektivitas penyusun kamus (Lauder, 2010: 222). Bahkan, Piotrowski (1994) membedakan pembuat kamus karya linguist dan karya bukan linguist. Menurutnya, kamus karya bukan linguist cenderung superfisial dan berisikan sejumlah data dengan cakupan yang luas, sedangkan kamus karya linguist cenderung meneliti dalam cakupan yang terbatas, lebih intensif, dan telaah kasus secara mendalam. Dari hal tersebut, dapat dikatakan bahwa kamus dari bukan linguist kurang mendalam seperti kamus karya linguist. Tentu saja, hal tersebut disebabkan pembuat kamus yang bukan linguist ingin memasukkan semua hal yang dirasa perlu sebagai bagian dari pendokumentasian bahasa. Namun, hal tersebut membuat kamus menjadi tidak mendalam dan kurang spesifik.

Maka dari itu, pada tulisan ini dapat dirumuskan permasalahan penelitian berupa bagaimana evaluasi kamus bilingual Using-Indonesia sebagai upaya pemertahanan eksistensi bahasa Using di Kabupaten Banyuwangi dengan menggunakan pendekatan kriteria pembuatan kamus dari teori leksikografi yang ada dan apa saja masukan yang dapat dijadikan bahan referensi jika *Kamus Bahasa Using-Indonesia* akan direvisi di kemudian hari. Hal tersebut membuat penelitian ini bertujuan untuk mengidentifikasi komposisi standar kamus bilingual sesuai teori leksikografi dengan *Kamus Bahasa Using-Indonesia*. Selain itu, penelitian ini juga dapat menjelaskan kekuatan dan kelemahan yang terdapat pada kamus bilingual Using-Indonesia yang dibuat oleh linguist otodidak.

II. KERANGKA TEORI DAN METODE PENELITIAN

Jika dilihat berdasarkan jenis kamus, Kamus Bahasa Using-Indonesia tergolong ke dalam kamus dwibahasa atau kamus bilingual. Kamus bilingual adalah kamus yang berisi lema-lema dari dua bahasa berbeda namun memiliki makna yang setara satu sama

lainnya, lengkap dengan konteks yang melingkupinya (Coward & Grimes, 2000: 70). Kamus bilingual juga merupakan produk leksikografi dasar selain jenis-jenis kamus yang lain (Piotrowski, 1994). Pada dasarnya, kamus bilingual digunakan bagi pemelajar bahasa kedua. Oleh karena itu, kamus jenis ini dibuat sesederhana mungkin dan mudah digunakan. Selain itu, Humblé (2001) juga memaparkan bahwa kamus bilingual harus dibuat lebih lengkap dengan menyertakan definisi yang komprehensif dari masing-masing lema yang dipaparkan. Hal tersebut menandakan kamus ini tidak hanya ditujukan kepada pemelajar bahasa kedua saja tetapi juga pengguna bahasa pertama.

Berkaitan dengan hal tersebut, *Kamus Daerah Using-Indonesia* juga sekiranya sesuai dengan pendapat Humblé bahwa kamus bilingual ini tidak hanya ditujukan kepada pemelajar bahasa kedua saja tetapi juga untuk pengguna bahasa pertama. Untuk itu, diperlukan pemahaman mengenai kriteria pembuatan kamus bilingual supaya masyarakat mudah menggunakan dan isi kamus sesuai dengan kebutuhannya. Kriteria pembuatan kamus bilingual sendiri dipaparkan oleh beberapa ahli. Misalnya, Atkins & Rundell (2008) dan Kridalaksana (2003). Selain itu, kriteria pembuatan kamus bilingual berdasarkan pedoman SIL Internasional juga akan dilihat agar evaluasi sekiranya lebih komprehensif dan menyeluruh. Menurut Atkins & Rundell (2008: 176), bagian dari kamus dibedakan menjadi dua, yaitu materi depan dan belakang kamus serta konten dari penyusunan lema. *Pertama*, hal yang dimaksud dengan materi depan dan belakang kamus lebih mengarah kepada bagian di luar lema. Lebih spesifik lagi, hal ini berkaitan dengan materi apa saja yang harus ada di dalam kamus selain daftar kosakata atau lemanya. Sebenarnya masalah konten depan dan belakang kamus ini harus disesuaikan dengan kebutuhan dan sasaran kamus terlebih dahulu. Ini disebabkan kebutuhan yang berbeda sehingga membuat dan sasaran berbeda pula pada konten depan dan belakang kamus.

Lebih lanjut, Atkins & Rundell menjelaskan bahwa secara umum konten depan kamus memuat kata pengantar, ucapan terima kasih, penjelasan singkatan, label, dan kode yang digunakan dalam kamus. Selain itu, dapat pula ditambahkan penjelasan mengenai sejarah bahasa yang tentu saja harus disesuaikan dengan kebutuhan dan sasaran kamus. Tidak hanya itu, bagaimana cara menggunakan kamus atau petunjuk kamus juga harus ada di bagian depan kamus. Dalam hal ini, petunjuk kamus merupakan bagian yang paling penting untuk dapat mengarahkan pengguna supaya dapat memahami isi dari kamus. Petunjuk penggunaan kamus ini juga sekiranya harus memuat penjelasan terkait struktur lema sehingga pengguna dapat dengan jelas mengetahui nama dan komponen apa yang ada di dalam lema. Berikut contoh yang diberikan Atkins & Rundell (2008: 177).

Sementara itu, untuk bagian belakang kamus materi yang disarankan ada pada kamus secara umum adalah tabel kata, angka, berat, langkah-langkah, unsur kimia, angka romawi, dan pedoman umum ejaan bahasa tersebut. Selain itu, informasi tambahan, seperti variasi regional, peta dialek bahasa, dan menyediakan alat bantu belajar yang berguna sebagai panduan. Semua hal tersebut dirasa Atkins & Rundell sebagai pelengkap dan membuat keutuhan kamus.

	IPA pronunciation (see p. xvii)	variant spelling (applicable to whole entry)	
headword in bold	cosy /'kəʊzi/ <i>adj., n., & v. (US cozy)</i> • <i>adj. (cosier,</i>		comparative and
roman type	cosiest 1 comfortable and warm; snug. 2 <i>derog.</i>		superlative forms
plural inflection of	<i>friendly</i> • <i>n. (pl. -ies)</i> 1 a cover to keep something		of adjective
noun (see p. xviii)	hot, esp. a teapot or a boiled egg. 2 <i>Brit. a canopied</i>		inflected forms of
	corner seat for two. • <i>v.tr. (-ies, -ied)</i> (often foll. by		verb (see p. xviii)
indicating	<i>along</i>) <i>colloq.</i> reassure, esp. deceptively. □ cosy up		
subsumed idioms,	<i>to US colloq.</i> 1 ingratiate oneself with. 2 snuggle		
phrasal verbs, and	up to. □ cosily <i>adv.</i> cosiness <i>n.</i> [18 th c. from Scots,		
derivatives	of unknown origin]		
	subsumed derivatives		

Gambar 1. Contoh penjelasan konten lema pada bagian petunjuk kamus

Selanjutnya, pada konten penyusunan lema Atkins & Rundell (2008: 491) mengungkapkan bahwa konten lema memiliki beberapa gaya yang bisa diikuti. Salah satu gaya penulisan konten lema ini seharusnya disesuaikan lagi dengan kebutuhan dan sasaran sehingga pengguna tidak susah atau bingung ketika melihat kamusnya. Dalam hal ini, ada lima konten *multiword expressions* (MWEs) pada lema. Kelima konten MWE pada lema tersebut adalah idiom, kolokasi, kata majemuk, frase verba, dan kontruksi pendukung verba. Berikut contoh dua gaya penulisan konten lema berdasarkan kamus Inggris-Perancis yang berbeda seperti yang tertera di bawah ini.

<p>shrug /ʃrɑːɡ/</p> <p>A <i>n</i> (also ~ of the shoulders) <i>haussement m d'épaules</i>; to give a ~ hausser les épaules.</p> <p>B <i>utr</i> (p <i>prés etc -gg-</i>) (also ~ one's shoulders) <i>hausser les épaules fpl.</i></p> <p>Phrasal verb • shrug off ► ~ off [sth], ~ [sth] off ignorer [problem, rumour].</p> <p>OHFD-3 (2001)</p>	<p>shrug /ʃrɑːɡ/ 1 <i>n</i> <i>haussement m d'épaules</i> • to give a ~ of contempt hausser les épaules (en signe) de mépris • ... he said with a ~ ... dit-il en haussant les épaules or avec un haussement d'épaules 2 <i>vti to ~ (one's shoulders)</i> :hausser les épaules</p> <p>shrug off <i>vt sep</i> [+ suggestion, warning] dédaigner, faire fi de; [+ remark] ignorer, ne pas relever; [+ infection, cold] se débarrasser de</p> <p>CRFD-8 (2006)</p>
<p>shower /'ʃaʊə(r)/</p> <p>A <i>n</i> 1 (for washing) <i>douche f</i>; to have or take a ~ [...] 2 Meteorol <i>averse f</i>; [...] 3 (of confetti, sparks, fragments) <i>pluie f</i> (of de); [...] 4 US <i>bridal/baby ~</i> [...] 5 GB <i>pej</i> (gang) <i>bande f</i>.</p> <p>B <i>modif</i> [<i>cubicle, curtain, head, rail, spray</i>] de douche.</p> <p>C <i>utr</i> 1 (wash) <i>doucher</i> [...]; 2 to ~ sth on or over sb/sth [...] 3 <i>fig</i> to ~ sb with sth [...].</p> <p>D <i>vi</i> 1 (person) <i>prendre une douche</i>; 2 <i>petals/sparks ~ed on me</i> [...]</p> <p>shower: ~ <i>attachment n</i> <i>douchette f</i> de lavabo; ~ <i>cap n</i> <i>bonnet m</i> de douche; ~ <i>proof adj</i> <i>impermeabilisé</i>; ~ <i>unit n</i> <i>douche f</i>; ~ <i>room n</i> (<i>private</i>) <i>salle f</i> de bains (avec douche); (<i>public</i>) <i>douches fpl.</i></p> <p>OHFD-3 (2001)</p>	<p>shower /'ʃaʊə(r)/ 1 [<i>of rain</i>] <i>averse f</i> (<i>fig</i>) [<i>of blows</i>] <i>volée f</i> [...]</p> <p>2 <i>douche f</i> to have or take a ~ <i>prendre une douche</i> [...]</p> <p>3 (<i>Brit ** pej = people</i>) <i>bande f</i> de crétins *</p> <p>4 (<i>before wedding etc</i>) to give a ~ for sb <i>organiser</i> [...]</p> <p>VT (fig) to ~ sb with gifts [...]</p> <p>VI 1 (= <i>wash</i>) <i>se doucher, prendre une douche</i></p> <p>2 (= <i>fall</i>) <i>hailstones ~ed (down) on to the car</i> [...]</p> <p>COMP shower attachment <i>N</i> <i>douchette f</i> à main, <i>douchette f</i> de lavabo</p> <p>shower cap <i>N</i> <i>bonnet m</i> de douche</p> <p>shower cubicle <i>N</i> <i>cabine f</i> de douche</p> <p>shower curtain <i>N</i> <i>rideau m</i> de douche</p> <p>shower gel <i>N</i> <i>gel m</i> douche</p> <p>shower stall <i>N</i> → <i>shower cubicle</i></p> <p>shower unit <i>N</i> <i>bloc-douche m</i></p> <p>CRFD-8 (2006)</p>

Gambar 2. Contoh gaya konten MWE dari dua kamus yang berbeda

Berdasarkan gambar di atas dapat terlihat bahwa terdapat perbedaan konten lema pada dua kamus yang berbeda. Perbedaan lebih kepada frasa verba, kontruksi pembantu verba, dan lambang kolokasinya. Untuk perbedaan frasa verba, antara OHFD-3 dan CRFD-8 hanya terletak pada penyusunan atau formatnya saja. Pada OHFD-3, format yang digunakan untuk frasa verba melebur menjadi satu dalam satu paragraf dan kata entri depannya tidak dituliskan kembali. Hal tersebut berbeda dengan CRFD-8 yang menampilkan frasa verba dengan disusun ke bawah lengkap dengan kata entri depannya. Kemudian pada kontruksi pembantu verba perbedaan terletak dari ada tidaknya contoh kalimat. Hanya CRFD-8 yang mempunyai contoh kalimat. Selain itu, lambang kolokasinya pun berbeda. Lambang kolokasi pada OHFD-3 menggunakan huruf, sedangkan kolokasi pada CRFD-8 menggunakan angka.

Di pihak lain, Kridalaksana (2003: xii) menggolongkan kriteria kamus, baik kamus monolingual dan kamus bilingual, menjadi dua bagian. Bagian tersebut dibagi atas mikrostruktur dan makrostruktur. Mikrostruktur menyangkut susunan informasi dalam entri kamus dan semua informasi berupa format dan perincian entri tergantung pada pengguna yang dituju. Mikrostruktur meliputi lema (kata kepala), definisi dalam kamus, kelas kata, dan sublema dengan informasinya. Misalnya, derivasi, infleksi, lafal, variasi historis, sosial dan geografis, gaya bahasa, konteks penggunaan bahasa, sinonim dan antonym, serta etimologi yang ditandai dengan label singkatan atau lambang. Sementara itu, makrostruktur bersangkutan dengan susunan entri dalam kamus. Makrostruktur juga meliputi susunan abjad, rujuk silang, kata majemuk, gabungan leksem, tipografis, gambar atau diagram sebagai ilustrasi, dan petunjuk penggunaan kamus, sejarah bahasa, daftar nama negara di dunia, dan informasi lain yang dianggap perlu oleh pengguna. Perihal ilustrasi pun merupakan hal penting untuk mempermudah pengguna (Granger & Lefer, 2016)

Tidak hanya itu, kriteria kamus bilingual juga terdapat pada pedoman pembuatan kamus dari SIL International. Pada buku *Making Dictionaries* (2000) yang diterbitkan oleh SIL *International*, kriteria kamus bilingual juga meliputi isi dan luar isi. Maksud dari isi adalah informasi struktur dalam entri leksikal, sedangkan luar isi adalah bagian selain lema seperti pengantar kamus. Coward & Grimes (2000: 178) mengungkapkan bahwa pengantar kamus harus dibuat untuk mengelaborasi informasi yang ada di dalam kamus sehingga pengguna dapat dengan mudah menggunakan kamus tersebut. Lebih spesifik lagi, Coward & Grimes memberikan beberapa aspek yang sekiranya harus ada atau disarankan ada pada bagian pengantar kamus seperti daftar di bawah ini:

1. Identifikasi sasaran dan tujuan penggunaan kamus. Selain itu, disarankan pula untuk memberikan keseluruhan informasi kamus. Misalnya, memberikan urutan alfabet, memberikan jumlah entri kamus, dan sebagainya.
2. Memberikan keterangan singkat terkait lokasi bahasa, jumlah populasi etnis kelompok, jumlah penutur bahasa, dan konteks regional bahasa tersebut berada.

3. Memberikan penjelasan peristiwa sejarah, seperti perang, migrasi, maupun hal-hal yang dianggap perlu untuk menjelaskan perubahan atau variasi bahasa sehingga pembaca tidak salah menafsirkan informasi yang terdapat dalam kamus.
4. Memberikan ulasan singkat tentang nama bahasa atau nama alternatifnya.
5. Menyebutkan klasifikasi linguistik, seperti hubungan kekerabatan bahasa, induk bahasa, dan sebagainya.
6. Daftar sejumlah karya bahasa yang telah dipublikasikan sebelumnya
7. Menyediakan profil sosiolinguistik singkat, termasuk dialek, register sosial, pola bicara berdasarkan gender, pendidikan, atau apapun yang akan membantu pengguna kamus.
8. Menyediakan peta dalam konteks regional dan peta dialek untuk membantu pengguna kamus memahami informasi variasi dialek.
9. Memberikan gambaran singkat fonologi, panduan untuk pengucapan, dan panduan untuk ortografi yang digunakan dalam kamus.
10. Memberikan gambaran singkat dari tata bahasa.
11. Memberikan sketsa etnografi singkat untuk membantu pengguna kamus menafsirkan masukan pada kamus.
12. Memberikan panduan untuk label dan singkatan yang digunakan dalam kamus.
13. Memberikan penjelasan cara membaca entri kamus
14. Menyediakan bagian yang menjelaskan bagaimana menggunakan rujukan silang
15. Menyediakan bibliografi semua referensi yang dikenal mulai dari segi bahasa, budaya, dan sejarah yang digunakan pada kamus.

Di sisi lain, Coward & Grimes (2000: 99) menyebutkan pula bahwa dalam membuat kamus ada beberapa struktur informasi dalam entri leksikal yang harus ada atau terpenuhi. Beberapa struktur informasi dalam entri leksikal yang harus ada pada sebuah kamus adalah prinsip pemilihan lema atau kata kepala, pemilihan contoh kalimat, pemisahan perbedaan kata dan perbedaan makna, kategori semantik, dan informasi dialek. Pemilihan lema atau kata kepala ini didasarkan pada akar kata dan afiks yang mengikutinya. Pemilihan contoh kalimat harus ada pada setiap makna kata sehingga pengguna kamus dapat mengetahui konteks dari kata tersebut. Pemisahan perbedaan kata dan perbedaan makna berdasarkan homonimi dan polisemi ini penting untuk menggambarkan situasi kebahasaan pada kamus sehingga pengguna tidak akan salah bila mengkaitkannya dengan konteks yang ada di masyarakat. Kategori semantik pun perlu untuk ditinjau lebih lanjut untuk dapat mengkategorikan kata menjadi sebuah lema dan informasi dialek dapat dimaksudkan untuk mengetahui asal dialek dari lema yang ada sehingga pengguna kamus akan mengetahui sumber dari entri leksikal tersebut tergolong ke dalam dialek yang mana jika bahasa tersebut mempunyai banyak dialek.

Dari segi metode penelitian, metode pemerolehan data didapat dari *Kamus Using-Indonesia* (2002). Data ini diambil karena penulis menduga belum ada penelitian sejenis sebelumnya. Penulis juga berharap bahwa dengan menggunakan data ini akan dapat

memberikan kontribusi dalam mengangkat salah satu hasil kodifikasi bahasa Using itu sendiri. Data ini pun diambil karena melihat keadaan kamus bilingual ini sudah sulit ditemukan. Sementara itu, metode penelitian yang digunakan pada penelitian ini adalah metode kualitatif. Metode kualitatif adalah metode yang berhubungan dengan ide, persepsi, pendapat, atau kepercayaan orang yang diteliti. Semuanya tidak dapat diukur dengan angka (Basuki, 2006: 78). Jadi, dalam tulisan ini data diolah menurut interpretasi penulis tanpa adanya penghitungan yang bersifat angka. Tulisan ini juga lebih kepada penelitian korpus data dari kepustakaan dan bukan penelitian lapangan.

Langkah pertama yang dilakukan dalam penelitian ini adalah melihat keseluruhan *Kamus Bahasa Daerah Using-Indonesia* untuk menentukan topik apa yang akan dibahas. Setelah mendapatkan topik atas bahasan yang akan digunakan pada penelitian ini, penulis mencari teori yang relevan terkait dengan topik yang ditemukan sebelumnya. Dalam hal ini, topik yang akan diangkat penulis lebih cenderung mengarah pada kriteria kamus bilingual. Hal ini disebabkan setelah melihat dengan kamus bilingual yang lain terdapat perbedaan konten maupun isi kamus sehingga menarik rasa penasaran penulis akan komposisi standar kamus bilingual itu sendiri dari berbagai sumber yang kompeten di ranah perkamusan. Berikutnya, melakukan identifikasi atas kriteria kamus bilingual dari Atkins & Rundell (2008), Kridalaksana (2003), dan Coward & Grimes (2000). Dari identifikasi tersebut nantinya diketahui kekuatan dan kelemahan atau aspek yang sudah terpenuhi dan belum terpenuhi dalam *Kamus Bahasa Daerah Using-Indonesia* sehingga dapat memberikan masukan konten atau isi apa saja yang harus ditambah dan dipertahankan jika kamus bilingual ini akan direvisi serta disebarluaskan kembali ke masyarakat di Kabupaten Banyuwangi.

III. ANALISIS PENELITIAN

Seperti penjelasan pada kerangka teori, analisis penelitian ini didasarkan pada kriteria kamus bilingual Atkins & Rundell (2008), Kridalaksana (2003), dan Coward & Grimes (2000). Secara umum ketiganya memiliki kriteria yang relatif sama. Hanya saja perbedaan di antara ketiganya dalam hal perincian dan penyebutan istilah-istilah saja. Oleh karena itu kriteria dari masing-masing ahli akan disandingkan dengan konten pada *Kamus Bahasa Daerah Using-Indonesia*. Berdasarkan penyandingan tersebut diharapkan akan menampilkan kriteria mana saja yang sudah terpenuhi dan kriteria mana yang belum terpenuhi. Hasilnya akan dijadikan sebagai bahan pertimbangan saat melakukan perevisian kamus di kemudian hari.

Untuk mengawali bagian analisis, kriteria kamus bilingual Coward & Grimes (2000) akan dipaparkan terlebih dahulu. Dalam hal ini, Coward & Grimes membagi kriteria kamus bilingual menjadi dua bagian. Bagian pertama adalah kriteria pengantar kamus, sedangkan bagian kedua adalah kriteria isi kamus. Berikut ini perbandingan kriteria pengantar kamus yang ditetapkan dengan kriteria pada *Kamus Bahasa Daerah Using-Indonesia*.

No	Kriteria Pengantar Kamus	Kamus Bahasa Using-Indonesia	
		Ketersediaan	Keterangan
1	Tujuan, sasaran penggunaan, dan gambaran penyusunan kamus (misalnya urutan lema disusun berdasarkan alfabetis)	✓	Kecuali jumlah entri kamus.
2	Keterangan bahasa: lokasi pemakaian, jumlah populasi, etnis pemakai, jumlah penutur, dan konteks regional daerah penutur	✓	Lokasi ada, pemakaian oleh siapa ada, etnis ada, jumlah penutur tidak, konteks regional ada
3	Peristiwa bersejarah yang melatari perkembangan bahasa: ditafsirkan melalui simbol \et (etimologi) dan \bw borrowed word atau kata pinjaman)	✓	Etimologi diberi dalam bentuk label nama Bahasa asalnya. Misal: A (Arab), JK (Jawa Kuna), S (Sanskerta)
4	Ulasan singkat tentang penamaan bahasa atau nama alternatif bahasa tersebut, jika memang isu ini dianggap relevan	✓	Penjelasan dicantumkan melalui kata pengantar oleh Perwakilan Kabupaten Banyuwangi
5	Klasifikasi linguistik: diperdebatkan atau tidak	—	Tidak ada penjelasan. Padahal terdapat penjelasan linguistik lain yang menyatakan bahwa bahasa Using bukan bahasa tetapi bentuk dialek bahasa Jawa Kuna (Budiono, 2015)
6	Daftar sejumlah karya bahasa yang telah dipublikasikan sebelumnya	✓	Terdapat beberapa karya seperti disertasi Bahasa Using di Kabupaten Banyuwangi (1987), karya sastra Sri Tanjung dalam Ensiklopedia Indonesia (1987), dan Kalangwan (1985)

7	<p>Profil sosiolinguistik singkat, termasuk aneka dialek, daftar kelompok masyarakat, daftar leksikal yang tabu, pola tuturan yang berbeda pada lintas gender/usia, ragam tuturan penutur yang berpendidikan.</p> <p>Misalnya: \us (usage atau penggunaan), \va (varian), \oe (restrictions), \lf SynD (sinonim dialektal), \lf SynR (sinonim register), \lf SynT (sinonim tabu), dan \lf SynL (sinonim pinjaman yang sudah berasimilasi)</p>	—	Tidak ada penjelasan
8	Peta persebaran bahasa dan dialek	—	Tidak ada penjelasan
9	Bagan fonologi singkat, panduan pelafalan, dan panduan ortografi berikut dengan sejumlah contoh sederhana	✓	Seluruhnya dimuat secara lengkap dan ringkas pada kamus ini
10	Gambaran tata bahasa singkat	—	Tata bahasa tidak ada. Namun, dalam pustaka acuan “Pedoman Umum Ejaan Bahasa Using” tertera informasi acuan dari Tata Bahasa Baku Bahasa Using (Hasan Ali, 1990) yang belum pernah diterbitkan
11	Gambaran singkat etnografis. Misalnya: struktur sosial, budaya material, ekonomi, agrikultur, dan kosmologi.	—	Tidak ada penjelasan
12	Pencantuman label-label dan singkatan	✓	Terdapat informasi seperti perbedaan representasi huruf

13	Pencantuman cara membaca entri kamus. Misalnya: informasi apa yang dipaparkan pertama kali, perbedaan representasi huruf (huruf tebal, miring, dll), makna hierarki dari entri, penandaan homonim, dll).	✓	Terdapat informasi seperti perbedaan representasi huruf
14	Pencantuman bagaimana menggunakan rujukan silang (reversed finderlist)	✓	Terdapat penjelasan (pada hal x)
15	Bibliografi seluruh informasi yang termaktub di kamus	✓	Terdapat bibliografi

Tabel 1. Perbandingan *Kamus Bahasa Daerah Using-Indonesia* dengan Kriteria Pengantar Isi Kamus (Coward & Grimes, 2000)

Tabel di atas menunjukkan bahwa kriteria pengantar kamus yang sudah terpenuhi dari *Kamus Bahasa Daerah Using-Indonesia* yaitu urutan alfabetis, identifikasi sasaran & tujuan kamus, keterangan bahasa, penjelasan sejarah bahasa, panduan pengucapan, panduan label & singkatan, dan bibliografi. Semua hal tersebut ada pada *Kamus Bahasa Daerah Using-Indonesia* walaupun tidak secara khusus dijelaskan. Maksudnya, kriteria seperti urutan alfabetis, panduan pengucapan dan panduan label & singkatan dalam *Kamus Bahasa Daerah Using-Indonesia* termasuk ke dalam atau berada di bagian petunjuk penggunaan kamus sebelum bagian daftar lema. Hal yang sama juga diperlihatkan pada identifikasi sasaran & tujuan, keterangan bahasa, dan penjelasan sejarah bahasa. Ketiga hal itu berada dalam satu bagian, yaitu bagian pengantar. Selain itu, identifikasi sasaran & tujuan, dan keterangan bahasa juga terdapat pada bagian kata sambutan kepala pusat bahasa dan sambutan Bupati Banyuwangi. Tidak lupa, bibliografi pun dicantumkan pada *Kamus Bahasa Daerah Using-Indonesia* setelah daftar lema.

Di sisi lain, kriteria yang belum terpenuhi adalah informasi jumlah entri kamus, profil sosiolinguistik, peta dialek, sketsa etnografi, penjelasan cara baca entri, dan petunjuk rujuk silang. Jumlah entri kamus ini penting sebagai informasi dasar kamus sehingga dapat memudahkan proses revisi. Sebenarnya jumlah entri kamus cukup disebutkan bersama kata pengantar layaknya informasi keterangan bahasa dan penjelasan sejarah bahasa yang sudah ada dalam kamus. Kemudian, profil sosiolinguistik, peta dialek, dan sketsa etnografi pun seharusnya dicantumkan pada *Kamus Bahasa Daerah Using-Indonesia* supaya pengguna mendapatkan gambaran terkait kondisi kebahasaan dan kondisi geografis bahasa Using. Sebagai perbandingan, aspek ini sudah ada pada kamus bilingual *Kamus Pengantar Bahasa Pantar Barat* (Holton, 2008).

Hal lain yang sudah ada tetapi kurang diperjelas pada *Kamus Bahasa Daerah Using-Indonesia*, sesuai dengan kriteria pengantar kamus dari Coward & Grimes adalah penjelasan cara membaca entri dan petunjuk rujuk silang. Pada kamus ini penjelasan mengenai hal-hal yang ada pada bagian entri sudah dijelaskan pada bagian petunjuk penggunaan kamus. Akan tetapi, jika didasarkan pada tujuan kamus yang penggunaannya semua kalangan, baik linguist maupun bukan linguist, hal tersebut belum cukup menjelaskan apa maksud dari hal-hal yang ada pada lema. Tidak ada penjelasan yang dapat memudahkan masyarakat umum, terutama mereka yang tidak menggeluti bidang linguistik secara khusus. Seharusnya ada penjelasan cara membaca entri seperti contoh penjelasan konten lema pada bagian petunjuk kamus yang telah diterangkan Atkins & Rundell (2008), pada bagian kerangka teori. Kondisi serupa juga terdapat pada kriteria petunjuk rujuk silang. Pada *Kamus Bahasa Daerah Using-Indonesia*, informasi rujuk silang sudah dimuat untuk mengarahkan pengguna kamus ke daftar entri lema yang benar. Namun sayangnya, petunjuk rujuk silang ini tidak dibuat secara khusus sehingga dapat memudahkan pengguna kamus agar tidak kesulitan.

No	Kriteria Isi Kamus	Kamus Bahasa Using-Indonesia	
		Ketersediaan	Keterangan
1	Lema	✓	Kecuali jumlah entri kamus.
2	Contoh Kalimat	—	Lokasi ada, pemakaian oleh siapa ada, etnis ada, jumlah penutur tidak, konteks regional ada
3	Perbedaan Kata	—	Etimologi diberi dalam bentuk label nama Bahasa asalnya. Misal: A (Arab), JK (Jawa Kuna), S (Sanskerta)
4	Homonimi vs Polisemi	—	Penjelasan dicantumkan melalui kata pengantar oleh Perwakilan Kabupaten Banyuwangi
5	Perbedaan Makna	✓	Tidak ada penjelasan. Padahal terdapat penjelasan linguistik lain yang menyatakan bahwa bahasa Using bukan bahasa tetapi bentuk dialek bahasa Jawa Kuna (Budiono, 2015)

6	Informasi Dialek	—	Terdapat beberapa karya seperti disertasi Bahasa Using di Kabupaten Banyuwangi (1987), karya sastra Sri Tanjung dalam Ensiklopedia Indonesia (1987), dan Kalangwan (1985)
---	------------------	---	---

Tabel 2. Perbandingan Kamus *Using-Indonesia* dengan Kriteria Isi Kamus (Coward & Grimes, 2000)

Dengan melihat kriteria isi kamus di atas, dapat diketahui bahwa kriteria isi *Kamus Bahasa Daerah Using-Indonesia* sebagian besar belum terpenuhi. Ini dibuktikan dengan hanya kriteria lema dan perbedaan makna saja yang terdapat pada *Kamus Bahasa Daerah Using-Indonesia*. Meskipun demikian, perbedaan makna tersebut juga belum konsisten ada pada seluruh lema atau mungkin saja memang hanya beberapa saja yang memiliki perbedaan makna. Ketidakkonsistenan juga terdapat pada kriteria contoh kalimat dan afiksasi. Sebenarnya, kedua kriteria tersebut sudah ada pada *Kamus Bahasa Using-Indonesia*. Akan tetapi, frekuensi kemunculannya yang sedikit atau inkonsisten pada seluruh lema sehingga penulis menggolongkan kedua kriteria ini belum terpenuhi dan menjadi bahan pertimbangan untuk revisi kamus di kemudian hari. Selain itu, perbedaan kata dan informasi dialek juga belum ada. Informasi dialek merupakan hal penting untuk mengenali variasi dari bahasa Using di Kabupaten Banyuwangi mengingat kabupaten ini tergolong luas dibandingkan daerah lainnya di sekitarnya.

Selanjutnya, struktur dan konten pada *Kamus Bahasa Daerah Using-Indonesia* akan dicocokkan dengan kriteria kamus bilingual dari Kridalaksana (2003). Dalam hal ini, Kridalaksana menerangkan bahwa kamus memiliki bagian makrostruktur dan mikrostruktur. Di bawah ini akan ditampilkan terlebih dahulu tabel perbandingan Kamus Bahasa Using-Indonesia dengan kriteria makrostruktur.

No	Kriteria Makrostruktur	Kamus Bahasa Using-Indonesia	
		Ketersediaan	Keterangan
1	Susunan Abjad	✓	Terdapat informasi susunan abjad pada awal petunjuk penggunaan kamus
2	Rujuk Silang	✓	Dicantumkan di bawah tanda baca
3	Kata Majemuk	✓	Dicantumkan dengan jelas pada Pedoman Umum Ejaan Bahasa Using, lampiran kamus
4	Gabungan Leksem	—	Tidak ada penjelasan

5	Tipografis	✓	Pencantuman penggunaan huruh tegak dan hurung miring pada <i>Petunjuk Penggunaan Kamus</i> ; Pencantuman penggunaan huruf tebal dan miring pada <i>Pedoman Umum Ejaan Bahasa Using</i>
6	Petunjuk Penggunaan	✓	Mencakup informasi susunan abjad, rujuk silang, dan tipografis,
7	Sejarah Bahasa	✓	Terdapat informasi sejarah bahasa dan sastra yang sangat singkat; bersifat pengenalan semata
8	Daftar Nama Negara	—	Tidak ada penjelasan

Tabel 3. Perbandingan Kamus *Using-Indonesia* dengan Kriteria Makrostruktur (Kridalaksana, 2003)

Bila dicocokkan dengan kriteria makrostruktur Kridalaksana (2003), *Kamus Bahasa Daerah Using-Indonesia* ini sudah ada yang sesuai dan ada pula yang belum. Kriteria makrostruktur yang sudah sesuai adalah susunan abjad, kata majemuk, rujuk silang, tipografis, petunjuk penggunaan kamus, dan sejarah bahasa. Berbeda dari criteria Coward & Grimes (2000), Kridalaksana tidak menyebutkan petunjuk rujuk silang secara spesifik sehingga dapat diartikan bahwa hanya muatan rujuk silang saja yang harus ada. Kemudian, susunan abjad dan tipografis melebur menjadi satu pada bagian petunjuk penggunaan kamus. Meskipun Kridalaksana membedakan susunan abjad, tipografis, dan petunjuk penggunaan kamus, tetapi pada *Kamus Bahasa Daerah Using-Indonesia* tidak membedakannya menjadi bagian yang berdiri sendiri pada satu bagian.

Di samping itu, kriteria makrostruktur yang belum sesuai adalah belum adanya kata majemuk, gabungan leksem, dan daftar nama negara. Untuk kata majemuk dan gabungan leksem perlu ditelusuri terlebih dahulu apakah pada bahasa Using memiliki kedua kriteria tersebut atau tidak. Jika memang ada, seharusnya kata majemuk dan gabungan leksem harus dimasukkan untuk menambah entri leksikal dan wawasan pembaca tentang khazanah bahasa Using. Kemudian, pada daftar nama negara yang belum ada pada kamus dirasa penulis wajar dan belum terlalu mendesak untuk dimasukkan mengingat kamus ini merupakan kamus bahasa daerah. Mungkin saja, jika nanti akan ada revisi sebaiknya memasukkan daftar nama daerah yang ada di Indonesia. Selain turut melakukan penginventarisasian daerah, memasukkan daftar nama daerah di Indonesia juga akan menjadi inovasi yang baru pada dunia perkamusan, khususnya kamus bahasa daerah jika memang ada penyebutan nama daerah yang berbeda bila disesuaikan dengan struktur bahasa Using.

Setelah melihat perbandingan kriteria makrostruktur kamus bilingual dari Kridalaksana dengan *Kamus Bahasa Daerah Using-Indonesia*, tidak lengkap rasanya bila tidak melihat pula perbandingan mikrostruktur kamus bilingualnya. Hal ini diharapkan akan dapat membuat pemahaman akan struktur kamus menjadi lebih lengkap atau komprehensif. Berikut di bawah ini kriteria mikrostruktur yang dibandingkan dengan *Kamus Bahasa Daerah Using-Indonesia*.

No	Kriteria Makrostruktur	Kamus Bahasa Using-Indonesia	
		Ketersediaan	Keterangan
1	Definisi	✓	Terdapat definisi pada setiap entri leksikal
2	Kelas Kata	✓	Terdapat informasi kelas kata pada setiap entri leksikal
3	Sublema dan Informasinya	—	Tidak terdapat sublema dan informasi yang meliputinya
4	Lafal	✓	Terdapat cara pelafalan pada setiap entri leksikal
5	Variasi Sosial	—	Tidak terdapat informasi variasi sosial penggunaan bahasa
6	Gaya Bahasa	✓	Terdapat informasi gaya bahasa

Tabel 4. Perbandingan Kamus *Using-Indonesia* dengan Kriteria Mikrostruktur (Kridalaksana, 2003)

Dari beberapa kriteria di atas dapat terlihat bahwa sebagian besar mikrostruktur berdasarkan Kridalaksana telah terpenuhi. Meskipun demikian, tetap ada beberapa kriteria yang masih belum terpenuhi. Kriteria yang terpenuhi adalah pendefinisian lema, kelas kata, lafal, dan gaya bahasa. Kondisi tersebut merupakan keunggulan dari *Kamus Bahasa Daerah Using-Indonesia*. Bila dibandingkan dengan kamus bilingual lainnya, penulis menilai kamus bilingual ini memuat susunan lema yang terbilang banyak. Apalagi kamus ini baru pertama kali diterbitkan. Beberapa kriteria yang terpenuhi pun secara konsisten terdapat hampir di setiap lemannya. Hal ini membuat revisi atas kamus menjadi tidak terllu banyak dalam hal susunan lema. Dengan begitu, kriteria yang belum terpenuhi seperti sublema beserta informasinya dan variasi sosial tinggal ditambahkan saja untuk menambah kelengkapan susunan lema.

Berbeda dengan Coward & Grimes (2000) dan Kridalaksana (2003), Atkins & Rundel (2008) menggolongkan struktur kamus terdiri dari bagian depan & belakang kamus dan konten lema kamus. Meskipun pada dasarnya sama saja dengan teori terdahulunya, Atkins & Rundel sudah menentukan bagian luar lema dengan posisi depan dan belakang kamus. Adanya pembagian tersebut, sekiranya akan memudahkan pembuat kamus untuk

menentukan posisi bagian luar lema. Berikut perbandingan kriteria depan dan belakang kamus dengan *Kamus Bahasa Daerah Using-Indonesia*.

No	Kriteria Depan & Belakang Kamus	Kamus Bahasa Using-Indonesia	
		Ketersediaan	Keterangan
	<i>Depan Kamus</i>		
1	Ucapan Terima Kasih	✓	
2	Kelas Kata	✓	Terdapat informasi kelas kata pada setiap entri leksikal
3	Penjelasan Singkat	✓	Tidak terdapat sublema dan informasi yang meliputinya
4	Label	✓	Terdapat cara pelafalan pada setiap entri leksikal
5	Kode	✓	Tidak terdapat informasi variasi sosial penggunaan bahasa
6	Sejarah Bahasa	✓	
7	Petunjuk Kamus	✓	
8	Komponen Lema	✓	
	<i>Belakang Kamus</i>		
1	Pedoman Ejaan	✓	Terdapat pedoman ejaan pada bagian akhir kamus (lampiran)
2	Variasi Regional	—	Tidak terdapat informasi variasi regional
3	Peta Dialek Bahasa	—	Terdapat informasi gaya bahasa

Tabel 5. Perbandingan Kamus *Using-Indonesia* dengan Kriteria Depan & Belakang Kamus (Atkins & Rundell, 2008)

Sesuai dengan kriteria depan dan belakang kamus berdasarkan Atkins & Rundell (2008), *Kamus Bahasa Using-Indonesia* termasuk ke dalam kamus yang sudah memenuhi kriteria depan dan belakang kamus, terutama kriteria depan kamus. Hanya variasi regional dan peta dialek bahasa saja yang belum termuat pada *Kamus Bahasa Daerah Using-Indonesia*. Sama seperti sebelumnya, kriteria depan kamus, seperti komponen lema, kode dan label dimasukkan menjadi satu pada bagian petunjuk penggunaan kamus. Sementara itu, penjelasan singkat bahasa, sejarah bahasa, dan ucapan terima kasih pun termuat pada satu bagian di bagian pengantar. Bahkan, ucapan terima kasih dan penjelasan singkat bahasa juga termuat pada bagian kata sambutan kepala pusat bahasa dan kata sambutan bupati Banyuwangi. Adanya pengulangan ucapan terima kasih dan penjelasan singkat bahasa menandakan bahwa kamus bilingual ini memang sangat

dibutuhkan bagi masyarakat Banyuwangi dan sebagai upaya pemertahanan eksistensi bahasa Using itu sendiri.

Sebaliknya, jika pada kriteria depan dan belakang kamus berdasarkan Atkins & Rundell *Kamus Bahasa Daerah Using-Indonesia* sudah terpenuhi, pada kriteria konten lema justru belum terpenuhi. Hal ini harus dijadikan evaluasi untuk para pembuat *Kamus Bahasa Daerah Using-Indonesia* selanjutnya supaya lebih memerhatikan konten lema dengan lebih cermat. Berikut perbandingan kriteria konten lema dengan Kamus Bahasa Daerah Using-Indonesia seperti yang tertera di bawah ini.

No	Kriteria Konten Lema	Kamus Bahasa Using-Indonesia	
		Ketersediaan	Keterangan
1	Idiom	—	Tidak terdapat informasi idiom
2	Kolokasi	—	Tidak terdapat informasi kolokasi
3	Kata Majemuk	✓	Terdapat informasi kata majemuk
4	Frase Verba	—	Tidak terdapat informasi frase verba
5	Fonologi	✓	Terdapat informasi fonologi
6	Kalimat	—	Tidak terdapat informasi kalimat

Tabel 6. Perbandingan Kamus *Using-Indonesia* dengan Kriteria Konten Lema Kamus (Atkins & Rundell, 2008)

Pada bagian konten lema kamus, kriteria seperti fonologi dan kata majemuk sudah terpenuhi dan kriteria idiom, kolokasi, frase verba, dan kalimat belum terpenuhi. Adanya kondisi demikian dirasa penulis wajar karena kamus bilingual ini baru pertama kali diterbitkan sehingga pembuat kamus lebih mementingkan inventarisasi lema dan belum secara mendalam memerhatikan isi konten lema atau susunan lema. Terlebih lagi, penyusun kamus merupakan linguist otodidak dan bukan linguist konvensional. Meskipun demikian, kriteria fonologi yang sudah ada ini cukup membuat penulis terkesan mengingat kriteria fonologi ini penting untuk mempermudah pembaca bukan pemelajar bahasa pertama mengucapkan bahasa Using sesuai dengan masyarakat Using itu sendiri.

Sementara itu, kriteria konten lema yang belum ada pada *Kamus Bahasa Daerah Using-Indonesia* ini harus dijadikan bahan evaluasi dan pertimbangan apakah perlu diadakan atau tidak pada kamus bilingual ini. Hal tersebut sekiranya selain membutuhkan adanya penelitian lebih lanjut, juga dapat dijadikan dasar sebelum membuat kamus bilingual bahasa daerah. Evaluasi tersebut diharapkan dapat membuat *Kamus Bahasa*

Daerah Using-Indonesia ini menjadi lebih lengkap dan sesuai dengan kebutuhan masyarakat maupun kebutuhan penelitian yang tertarik untuk meneliti bahasa Using selain dari berbagai dokumentasi bahasa Using dari berbagai media cetak yang sekarang ini gencar dilakukan oleh masyarakat Banyuwangi.

IV. SIMPULAN

Berdasarkan pemaparan di atas, dapat disimpulkan bahwa *Kamus Bahasa Daerah Using-Indonesia* tergolong ke dalam kamus bilingual bahasa daerah yang sudah hampir mendekati kriteria-kriteria kamus atau kamus bilingual. Ini dilihat dari segi penjelasan luar lema maupun susunan lemanya itu sendiri jika dibandingkan dengan kamus bilingual bahasa daerah lainnya. Poin positif lain yaitu penyusun kamus ini merupakan linguist otodidak (Arps, 2010). Kamus yang terbit pada 2002 ini belum pernah direvisi hingga saat ini. Oleh karena itu, jika kamus ini akan direvisi atau dievaluasi, penulis berharap kamus ini menjadi semakin paripurna. Bahkan, jika evaluasi kamus nantinya dilakukan dengan menggunakan tiga kriteria (Atkins & Rundell, 2008; Coward & Grimes, 2000; Kridalaksana, 2003), bukan tidak mungkin *Kamus Bahasa Daerah Using-Indonesia* ini akan menjadi percontohan sebagai kamus bilingual bahasa daerah bagi kamus-kamus bilingual bahasa daerah lainnya.

Sesuai dengan judul tulisan ini, berbagai sudut pandang beberapa ahli yang telah disebutkan sekiranya dapat menjadi bahan evaluasi yang komprehensif dan terbaru mengingat teori leksikografi terus berkembang dari waktu ke waktu. Perkembangan ini bisa saja melampaui standar baku pembuatan kamus yang telah ditetapkan Badan Bahasa selaku pemegang otoritas kebijakan bahasa di Indonesia. Ini yang mendasari perlu adanya evaluasi pada kamus sehingga struktur kamus menjadi lebih paripurna dan lebih sesuai dengan kegunaannya di masyarakat. Selain itu, evaluasi ini akan memudahkan masyarakat menggunakan kamus dan mampu merekam perkembangan maupun perubahan bahasa yang terjadi pada bahasa Using.

Maka dari itu, pada bagian di bawah ini merupakan poin-poin evaluasi atau usulan pelengkapan bagi *Kamus Bahasa Daerah Using-Indonesia*. Poin-poin ini merupakan kriteria yang belum dipenuhi berdasarkan kriteria kamus bilingual dari Atkins & Rundell (2008), Kridalaksana (2003), dan Coward & Grimes (2000). Lebih spesifik lagi, evaluasi ini merupakan bagian luar lema dan susunan lema jika nantinya *Kamus Bahasa Daerah Using-Indonesia* akan direvisi.

No	Kriteria Luar Lema	Kriteria Susunan Lema
1	Jumlah entri	Contoh Kalimat
2	Profil Sociolinguistik	Perbedaan Kata
3	Peta Dialek	Afiksasi

4	Cara Membaca Entri	Informasi Dialek
5	Petunjuk Rujuk Silang	Kata Majemuk
6		Gabungan Leksem
7		Variasi Sosial
8		Idiom
9		Kolokas
10		Frase Verba

Daftar Pustaka

- Ali, Hasan. 2002. *Kamus Bahasa Using-Indonesia*. Banyuwangi: Dewan Kesenian Blambangan.
- Arps, Bernard. 2010. Terwujudnya Bahasa Using di Banyuwangi dan Peranan Media Elektronik di Dalamnya (Selayang Pandang 1970-2009). Dalam Mikihiro Moriyaman dan Manneke Budiman (Ed.) *Geliat Bahasa Selaras Zaman: Perubahan bahasa-bahasa di Indonesia pasca-Orde Baru*. Jakarta: Kepustakaan Populer Gramedia.
- Atkins, B.T. & Rundell, Michael. 2008. *The Oxford Guide to Practical Lexicography*. New York: Oxford Univesity Press.
- Budiono, Satwiko. 2015. *Variasi Bahasa di Kabupaten Banyuwangi: Penelitian Dialektologi*. Skripsi: Fakultas Ilmu Pengetahuan Budaya Universitas Indonesia.
- Coward, David. F. & Grimes, Charles. E. 2000. *Making Dictionaries: A Guide to Lexicography and the Multi-Dictionary Formatter*. North Carolina: SIL International.
- Grenger, Sylviane dan Lefer, Marie-Aude, 2016. From general to learner's bilingual dictionaries: Towards a more effective fulfillment of advanced learner' phraseological needs. In *Int J Lexicography* Dipublikasikan secara daring pada 2 Juni, 2016 doi:10.1093/ijl/ecw022
- Holton, Gary, dan Mahalalel Lamma Holy. 2008. *Kamus Pengantar Bahasa Pantar Barat*. Kerjasama antara Universitas Alaska dengan Unit Bahasa dan Budaya, GMT.
- Humblé, Philippe. 2001. *Dictionaries and Language Learners*. Frankfurt: Haag und Herchen.
- Lauder, Allan. F. 2010. Data for Lexicography: The Central Role of The Corpus. *Wacana*, Vol. 12. No.2. 219-242.
- Lew, Robert. 2016. Can a dictionary help you write better? A user study of an active bilingual dictionary for Polish learners of English. *Int J Lexicography* dipublikasikan secara daring pertama sekali pada 2 Juni, 2016 doi:10.1093/ijl/ecw024

- Kridalaksana, Harimurti. 2003. Sambutan Ilmiah Pusat Leksikologi dan Leksikografi FIB UI dalam *Rintisan Kajian Leksikologi dan Leksikografi*. Lilie Suratminto dan Munawar Holil (Ed.). Depok: FIB UI.
- Pemerintah Kabupaten Banyuwangi. 2007. *Peraturan Daerah Kabupaten Banyuwangi Nomor 5 tentang Pembelajaran Bahasa Daerah Pada jenjang Pendidikan Dasar*.
- Piotrowski, Tadeusz. 1994. *Problems in Bilingual Lexicography*. Wroclaw: -----.
- Soetoko, dkk. 1981. *Geografi Dialek Banyuwangi*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan
- Sulistyo, Basuki. 2006. *Metode Penelitian*. Jakarta: Wedatama Widya Sastra dan Fakultas Ilmu Pengetahuan Budaya Universitas Indonesia.

Sejarah Perkembangan Leksikografi Tu Dialek di Brunei Darussalam

Hajah Rosmariah binti Haji Alli

Dewan Bahasa dan Pustaka, Brunei

maria_alli@hotmail.com

Abstrak

Leksikografi merupakan proses menyusun kamus termasuklah kosa kata atau perbendaharaan kata. Kosa kata atau perbendaharaan kata akan melambangkan kekayaan bahasa dan budaya bagi setiap masyarakat. Keseluruhan kosa kata tersebut akan dirakam dan dimasukkan ke dalam kamus. Sejarah Perkamusan di Brunei Darussalam telah bermula awal abad ke-19. Dewan Bahasa dan Pustaka telah ditubuhkan bagi mengangkat martabat bahasa, sastera, dan budaya melayu. Penubuhan Unit Perkamusan di Dewan Bahasa dan Pustaka telah diamanahkan bagi menanai misi memasyarakatkan serta membudayakan bahasa Melayu selaras dengan hasrat Bahagian Pembinaan dan Pengembangan Bahasa mendaulatkan bahasa Melayu sebagai Bahasa Rasmi Negara. Unit Perkamusan juga berfungsi menangani projek penyusunan pelbagai kamus yang menjurus kepada pengayaan bahasa Melayu. Dewan Bahasa dan Pustaka telah berusaha mulai awal tahun 1971 dengan pembentukan sebuah jawatankuasa yang menyelenggarakan penyusunan kamus Bahasa Melayu Brunei. Dewan Bahasa dan Pustaka melalui Unit Perkamusan juga telah berjaya menerbitkan beberapa buah kamus bermula dengan penyusunan Kamus Kanak-kanak terbit pada tahun 1986 sehingga menerbitkan Daftar Leksikal 7 Dialek Brunei Darussalam pada tahun 2011. Antara objektifnya ialah untuk mengongsikan perkembangan leksikografi di Brunei Darussalam kepada orang awam. Metodologi penyelidikan deskriptif dan skopnya memfokuskan lapangan dan produk kamus yang dihasilkan. Antara produk kamus yang sudah terbit ialah Kamus Kanak-kanak, Kamus Kata dan Ungkapan Am, Kamus Bahasa Melayu Nusantara, Kamus Tutong, Kamus Kedayan, Kamus Bahasa Melayu Brunei, dan Daftar Leksikal 7 Dialek Brunei Darussalam.

I. PENDAHULUAN

Menurut R.R.K Hartmann(1993:4) memberi takrif mengenai kamus ialah buku rujukan atau senarai perkataan (biasanya mengikut abjad) dengan padanan maknanya, sebutan, ejaan, atau persamaan atau padanan dalam bahasa-bahasa dan leksikografi ialah proses yang berkaitan dengan fenomena yang boleh dicerapi (iaitu kosa kata, atau leksis, atau leksikon) dan prinsip teoretis (leksikologi) untuk menanganinya. Leksikografi merupakan proses menyusun kamus termasuklah kosa kata atau perbendaharaan kata. Kosa kata atau perbendaharaan kata akan melambangkan kekayaan bahasa dan budaya bagi setiap masyarakat. Keseluruhan kosa kata tersebut akan dirakam dan dimasukkan ke dalam kamus. Di Negara Brunei Darussalam terdapat tujuh puak jati yang diiktiraf dalam perlembagaan 1959 iaitu puak Melayu Brunei, Belait, Tutong, Kedayan, Murut, Dusun, dan Bisaya.

Sejarah Perkamusan di Brunei Darussalam telah bermula awal abad ke-19. Dewan Bahasa dan Pustaka telah ditubuhkan bagi mengangkat martabat bahasa, sastera, dan budaya melayu. Penubuhan Unit Perkamusan di Dewan Bahasa dan Pustaka telah diamanahkan bagi menanai misi memasyarakatkan serta membudayakan bahasa Melayu selaras dengan hasrat Bahagian Pembinaan dan Pengembangan Bahasa mendaulkan bahasa Melayu sebagai Bahasa Rasmi Negara. Unit Perkamusan juga berfungsi menangani projek penyusunan pelbagai kamus yang menjurus kepada pengayaan dan pengembangan bahasa Melayu. Dewan Bahasa dan Pustaka telah berusaha mulai awal tahun 1971 dengan pembentukan sebuah jawatankuasa yang menyelenggarakan penyusunan kamus Bahasa Melayu Brunei. Antara beberapa matlamat penerbitan kamus ini ialah sebagai suatu dokumentasi yang merakamkan kata-kata bahasa Melayu Brunei dan pada masa itu bahan bercetak agak terbatas penerbitannya. Kamus ini dicetak juga bertujuan sebagai penyelamat supaya kata-kata tersebut tidak pupus daripada diketahui oleh generasi muda.

II. OBJEKTIF KAJIAN

1. Mengongsikan perkembangan leksikografi di Brunei Darussalam kepada orang awam.
2. Mempertingkat pengetahuan belia mengenai dialek yang terdapat di Brunei Darussalam supaya tidak pupus.
3. Menggalakkan para belia untuk mengetahui pelbagai bahasa terutama dialek yang terdapat di negara ini.
4. Mendokumentasi dan mengkoordinasi kosa kata dan konteks bahasa Melayu Brunei dan dialek-dialek di Negara Brunei Darussalam dalam bentuk daftar dan kamus.

III. KAEDAH PENYELIDIKAN

Penyelidikan ini hanya menggunakan kaedah penyelidikan lapangan kepustakaan deskriptif.

IV. ANALISIS PENELITIAN

Perkembangan leksikografi boleh dilihat dengan hasil penerbitan beberapa buah kamus di bawah ini:

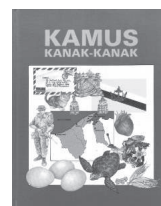
Bil	Nama Kamus	Tahun Terbit	Jumlah Entri
1	Kamus Kanak-kanak	Cetakan Pertama 1986; Cetakan Kedua 1990	895
2	Kamus Bahasa Melayu Brunei	Cetakan Pertama 1991; Cetakan Kedua 1998; Edisi Kedua 2007	15,537
3	Kamus Tutong-Melayu Melayu-Tutong	Cetakan Pertama 1991; Cetakan Kedua 1998	2000
4	Kamus Kata dan Ungkapan Am	1994	200
5	Kamus Bahasa Melayu Nusantara	Cetakan Pertama 2003; Edisi Kedua Cetakan Pertama	105,000
6	Kamus Kedayan-Melayu Melayu-Kedayan	2006	3506
7	Daftar Leksikal 7 Dialek Brunei Darussalam	2011	2000

4.1 Kamus-kamus Yang Sudah Terbit

Kamus Kanak-kanak



Cetakan Pertama 1986

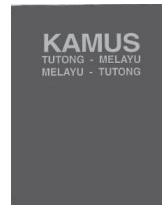


Cetakan Kedua (terbitan semula)
1990

Kamus Tutog-Melayu Melayu-Tutong



Cetakan Pertama 1991



Cetakan Kedua 1998

Kamus Kata dan Ungkapan Am



Cetakan Pertama 1994

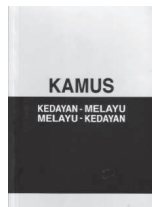


Cetakan Kedua
(Edisi Perbaikan) 2003



Edisi Kedua
Cetakan Kedua 2007

Kamus Kedayan-Melayu Melayu-Kedayan



Cetakan Pertama 2006

Kamus Bahasa Melayu Nusantara



Cetakan Pertama 2003

Kamus Bahasa Melayu Brunei



Cetakan Pertama 1991



Cetakan Kedua 1998



Edisi Kedua
Cetakan Pertama 2007

Dewan Bahasa dan Pustaka melalui Unit Perkamusan telah berjaya menghasilkan beberapa buah kamus iaitu:

1. Kamus Kanak-Kanak
2. Kamus Bahasa Melayu Brunei
3. Kamus Kata dan Ungkapan Am
4. Kamus Tutong-Melayu Melayu-Tutong
5. Kamus Kedayan-Melayu Melayu-Kedayan
6. Kamus Bahasa Melayu Nusantara
7. Daftar Leksikal 7 Dialek Brunei Darussalam

4.1.1 Kamus Kanak-Kanak

Kamus ini terbit pada tahun 1986 cetakan pertama dan cetakan kedua pada tahun 1990. Antara objektif kamus ini ialah untuk memenuhi keperluan pembelajaran kanak-kanak di bawah umur 12 tahun dan untuk membimbing kanak-kanak mempelajari bahasa Melayu dengan betul.

Kamus ini mengandungi informasi yang bermanfaat kepada kanak-kanak kerana kata masukannya meliputi kata yang terdapat dalam buku-buku teks sekolah rendah, kata-kata umum, nama-nama burung, ikan, dan binatang. Entri kamus ini juga mudah dan sesuai untuk kanak-kanak, termasuk kata masukan, contoh penggunaan ayat, frasa, simpulan bahasa, dan ilustrasi berwarna. Kamus ini juga sebagai rujukan di sekolah-sekolah rendah.

4.1.2 Kamus Bahasa Melayu Brunei

Kamus Bahasa Melayu Brunei cetakan Pertama terbit pada tahun 1991 dan cetakan kedua terbit pada tahun 1998, dan edisi kedua terbit pada tahun 2007. Antara objektif sebagai suatu dokumentasi yang merakamkan kosa kata bahasa Melayu Brunei yang agak terbatas penggunaannya, ia juga sebagai penyelamat kosa kata tersebut supaya tidak pupus daripada diketahui oleh masyarakat, dan juga sebagai bahan rujukan kepada para penyelidik yang hendak membuat kajian ilmiah.

4.1.3 Kamus Kata dan Ungkapan Am

Kamus ini mengandungi kata dan ungkapan am yang sama ejaan dan sebutan, tetapi membawa konotasi yang berbeza di Brunei Darussalam, Indonesia, dan Malaysia. Kata dan ungkapan am yang mempunyai makna tambahan yang membawa kelainan makna daripada bahasa Melayu, bahasa Indonesia, dan bahasa Malaysia. Bahasa Melayu Brunei merupakan dialek yang sangat popular digunakan sebagai bahasa perhubungan lisan dalam kalangan penduduk Negara Brunei Darussalam.

Kamus ini bertujuan untuk melaksanakan salah satu keputusan Sidang ke-21 Majlis Bahasa Indonesia-Malaysia (MBIM), pada April 1984 dan sebagai sumbangan Negara Brunei Darussalam setelah memasuki majlis tersebut yang telah berubah menjadi Majlis Bahasa Brunei Darussalam-Indonesia-Malaysia (MABBIM), November 1985. Kamus ini telah diterbitkan cetakan Pertama pada tahun 1994. Cetakan Kedua (Edisi perbaikan) terbit 2003 tujuan penerbitannya ialah sempena menyambut 30 tahun penubuhan Majlis Bahasa Brunei Darussalam, Indonesia, dan Malaysia (MABBIM).

Antara objektifnya ialah sebagai bahan rujukan dalam Sidang MABBIM dan juga membantu pengguna bahasa dalam mengenal pasti kata dan ungkapan am yang sudah mantap digunakan di ketiga-tiga buah negara.

4.1.4. Kamus Tutong-Melayu Melayu-Tutong

Kamus ini terbit pada 1991 iaitu Cetakan Pertama dan telah terbit cetakan kedua pada 1998. Antara objektif kamus ini ialah supaya bahasa atau dialek berkenaan akan diketahui khususnya oleh masyarakat, supaya bahasa ini tidak akan pupus daripada pengetahuan umum, dan ia dapat dijadikan asas kepada peminat bahasa bagi penerbitan tatabahasa bahasa Tutong.

Antara informasi yang terkandung dalam kamus ini ialah setiap entri bahasa Tutong dieja dalam bentuk fonemik dan diberi fonetik dan padanan dalam bahasa Melayu, untuk member kemudahan kepada pengguna kamus ini, rujuk silang Melayu-Tutong disediakan supaya orang-orang yang tidak mengetahui dialek Tutong dapat memahaminya, dan pelaksanaan kamus ini ialah melalui penyelidikan dan wawancara dengan penutur-penutur asli puak Tutong.

4.1.5 Kamus Kedayan-Melayu Melayu-Kedayan

Kamus ini terbit pada 2006 dan antara objektifnya ialah mengumpul kosa kata peribumi supaya tidak pupus dan memberi pendedahan kepada generasi baru tentang bahasa Kedayan dan ia juga sebagai khazanah budaya Melayu yang boleh dijadikan rujukan umum dan bahan kajian kepada para penyelidik bahasa.

Antara informasi kamus ini ialah kamus ini merakamkan kosa kata dialek puak Kedayan sebagai salah satu puak jati Melayu Brunei, penyusunan kamus ini dilaksanakan melalui penyelidikan dan wawancara dengan penutur-penutur asli bahasa Kedayan dari

beberapa buah kampong yang berumur dalam lingkungan 40 tahun hingga 65 tahun, kamus ini memuatkan 3236 entri Kedayan-Melayu dan rujuk silang Melayu-Kedayan sebanyak 3506 entri. Setiap entri bahasa Kedayan dieja dalam bentuk fonemik dan diberi fonetik, dan entri Kedayan-Melayu yang tiada padanannya dalam bahasa Melayu dihuraikan secara ringkas.

4.1.6 Kamus Bahasa Melayu Nusantara

Projek penyusunan Kamus Bahasa Melayu Nusantara merupakan salah satu resolusi Dialog Kalimantan Ketiga yang diadakan di Brunei Darussalam pada Oktober 1992. Kamus ini terbit cetakan pertama pada 2003 dan cetakan kedua pada 2011. Antara objektifnya ialah memahami makna kata yang digunakan oleh ketiga-tiga buah negara, mengelakkan penggunaan kata yang mempunyai makna berkonotasi, mengetengahkan bahasa Melayu Brunei dalam sebuah penerbitan yang lebih luas penyebarannya, dan sebagai khazanah bahasa Melayu di samping sebagai lambing kemegahan kepada bangsa Melayu terhadap bahasa.

Kamus ini memuat kosa kata bahasa Melayu Brunei, bahasa Indonesia, dan bahasa Malaysia. Penyusunannya dilaksanakan dengan kerjasama pakar dari Negara Brunei Darussalam, Indonesia, dan Malaysia. Kamus ini mengandungi kira-kira 105 000 entri: lema 47 000, sublema 26 000, peribahasa 39 000, kiasan 600, dan frasa atau kiasan 27 000. Penyusunan entri yang berbentuk “mengahwinkan” ketiga-tiga bahasa anggota MABBIM, memudahkan pengguna daripada membuka dua atau tiga kamus untuk merujuk makna kata. Penyusunan kamus ini membantu pengkaji bahasa menyelidiki kosa kata ketiga-tiga bahasa sama ada dari segi tatabahasa, ejaan, imbuhan, sebutan dan penulisan kata.

4.1.7 Daftar Leksikal 7 Dialek Brunei Darussalam

Daftar Leksikal 7 Dialek Brunei Darussalam merupakan pendokumentasian kosa kata tujuh-tujuh puak jati Negara Brunei Darussalam (Dialek Melayu Brunei, Dialek Kedayan, Dialek Dusun, Dialek Tutong, Dialek Murut, Dialek Belait, dan Dialek Bisaya) yang tercatat dalam Perlembagaan Negeri Brunei 1959. Data yang terkumpul berjumlah 2000 entri bahasa Melayu yang disinonimkan dengan leksikal yang dituturkan oleh 7 puak jati Brunei.

Antara objektifnya ialah untuk mengangkat bahasa puak-puak jati yang tercatat dalam Perlembagaan Negeri Brunei 1959 bagi memajukan dan mengayakan bahasa Melayu, mendokumentasi bahasa dan dialek peribumi untuk mengelak kepupusan, dan memperkaya korpus linguistik di Negara Brunei Darussalam, Kajian lapangan daftar ini ialah dengan menggunakan kajian akustik dan auditori.

Penyelidikan lapangan lanjutan bagi Daftar Leksikal 7 Dialek Brunei Darussalam untuk menambah korpus dan edisi perbaikan juga telah dibuat pada December 2015 hingga April 2016.

Hasil dapatan penyelidikan lapangan lanjutan ialah:

Bil.	Dialek	Jumlah entri
1	Dialek Melayu Brunei	123
2	Dialek Bisaya	212
3	Dialek Kedayan	229
4	Dialek Belait	634
5	Dialek Murut	339
6	Dialek Tutong	351
7	Dialek Dusun	230
Jumlah		2118

Metodologi penyelidikan bagi Daftar Leksikal 7 Dialek Brunei Darussalam ini ialah:

1. Temu bual berdasarkan Swadesh
2. Rakaman menggunakan perisian Wave Surfer (W.S) dan perisian Sound Filing System (S.F.S)
3. Transkripsi fonetik dan fonemik
4. Rakaman wacana spontan
5. Penggambaran

Kajian lapangan ini dilakukan dalam 3 tahap oleh 3 kumpulan. Setiap kumpulan akan menemu bual 3 orang informan dari 3 zon bagi setiap bahasa berdasarkan jadual. Lapangan tahap pertama merupakan lapangan untuk menemu bual informan secara kelompok atau kumpulan. Melalui temu bual ini, seorang informan perempuan dan seorang informan lelaki akan dipilih sebagai informan yang terjamin tuturnya untuk dirakam menggunakan perisian Wave Surfer pada tahap lapangan kedua. Rakaman pertama ini akan menggunakan perakam suara.

Penyelidikan lapangan tahap ini berpandukan Swadesh. Melalui lapangan ini dapatan (kata dan padanan) akan ditranskripsi menggunakan lambang fonetik (IPA). Di samping padanan kata, contoh-contoh pemakaian kata juga akan digali secara semantik dan pragmatik bahasa yang dirakam.

Lapangan tahap kedua akan mengaplikasi rakaman akustik menggunakan perisian Wave Surfer dan perisian Sound Filing System (S.F.S). Lapangan tahap ketiga pula penyelidik akan menyebut kembali dapatan kata dalam lapangan sebelumnya. Ini bertujuan untuk menyemak transkripsi fonetik dan menyemak kembali linguistic bahasa yang dikaji, iaitu fonologi, semantik, morfologi, dan sebagainya.

V. KAJIAN LEPAS

Kajian lepas yang pernah dibuat ialah Sejarah Perkamusan di Negara Brunei Darussalam satu tugas dan tuntutan oleh Dr. Haji Jaludin bin Haji Chuchu. Kertas kerja yang dibentangkan semasa Seminar Leksikografi Sempena Pelancaran Kamus Bahasa Melayu Nusantara 2003.

VI. CADANGAN

Pihak Dewan Bahasa dan Pustaka, Brunei perlu menjalin hubungan kerjasama dengan Kementerian Pendidikan untuk menerapkan dialek-dialek di Brunei menjadi salah satu mata pelajaran di sekolah rendah dan menengah. Tujuannya untuk memberikan galakan kepada para murid dan pelajar untuk mengetahui dialek-dialek yang terdapat di negara ini supaya tidak pupus ditelan zaman.

VII. KESIMPULAN

Sejarah perkembangan leksikografi di Brunei bermula dengan terbitnya Kamus Kanak-Kanak pada 1986. Ini telah menunjukkan dengan perkembangannya itu menjadi pemangkin atau platform kepada Dewan Bahasa dan Pustaka dalam upaya memupuk dan memperdaya perkembangan bahasa atau dialek di negara ini kepada masyarakat khususnya para belia, dan generasi baru. Penubuhan Unit Perkamusan di Dewan Bahasa dan Pustaka telah diamanahkan bagi menanai misi memasyarakatkan serta membudayakan bahasa Melayu selaras dengan hasrat Bahagian Pembinaan dan Pengembangan Bahasa mendaulatkan bahasa Melayu sebagai Bahasa Rasmi Negara Kementerian Pendidikan juga perlu memainkan peranan untuk mencapai visi dan misi Dewan Bahasa dan Pustaka dan menerapkan dialek-dialek menjadi mata pelajaran di sekolah-sekolah.

BIBLIOGRAFI

- Daftar Leksikal 7 Dialek Brunei Darussalam*, 2011. Brunei: Dewan Bahasa dan Pustaka
- Jaludin Haji Chuchu, 2003. “Sejarah Perkamusan Di Negara Brunei Darussalam Satu Tugas dan Tuntutan” dlm Seminar Leksikografi Sempena Pelancaran Kamus Bahasa Melayu Nusantara, hlm. 1-11. Negara Brunei Darussalam: Pusat Persidangan Antarabangsa Berakas.
- Kamus Bahasa Melayu Nusantara*, 2003. Brunei: Dewan Bahasa dan Pustaka
- Kamus Kedayan-Melayu Melayu-Kedayan*, 2006. Brunei: Dewan Bahasa dan Pustaka
- Kamus Tutong-Melayu Melayu-Tutong*, 1991. Brunei: Dewan Bahasa dan Pustaka
- R.R.K. Hartmann (ed.), 1993. *Leksikografi: Prinsip dan Amalan* penterjemah Zainab Ahmad. Kuala Lumpur: Dewan Bahasa dan Pustaka, Kementerian Pendidikan Malaysia.

Pembentukan Kamus Pewayangan Bahasa Jawa–Indonesia Berbasis Korpus

Ika Nurfarida, S.S. dan Nita Suryawati, S.Hum.

Magister Ilmu Linguistik Universitas Airlangga
kaa.farida@gmail.com, nita.suryawati@yahoo.co.id

Abstrak

Kamus memiliki peran yang cukup vital sebagai alat yang dapat membantu kita dalam memahami kosakata yang sukar dipahami. Dengan kamus, kosakata seseorang dapat berkembang. Terkait dengan peran penting sebuah kamus, maka di dalam studi ini, peneliti bermaksud mengembangkan kamus bahasa daerah, khususnya kamus pewayangan yang berbentuk kamus dwibahasa Jawa-Indonesia. Bahasa yang digunakan dalam pentas pewayangan yaitu bahasa Jawa campuran yang terdiri dari Jawa Ngoko, Jawa Krama dan Jawa Inggil. Kamus pewayangan yang akan dibentuk oleh peneliti merupakan sebuah kamus pewayangan yang berbentuk kamus saku (*pocket dictionary*) dengan jumlah entri kata sebanyak 500. Tujuan dibuatnya kamus ini tidak hanya diperuntukkan bagi mereka yang ingin melanjutkan studinya di bidang seni tradisional wayang kulit namun juga untuk masyarakat luas yang gemar terhadap seni tradisional wayang kulit sehingga dapat membantu mereka memahami kosakata yang terdapat dalam pentas seni pewayangan. Adapun tahap-tahap dalam pembuatan kamus di dalam penelitian ini meliputi beberapa tahap yaitu: (1) mengumpulkan kosakata yang telah ditranskrip dalam sebuah pentas seni wayang kulit selama semalam suntuk yang disimpan dalam bentuk file txt (plain text), (2) memasukkan hasil transkripsi ke alat *Ant Cont* 3.4 untuk mendapatkan kosakata tertinggi yang sering muncul dalam pentas seni wayang kulit, (3) menyeleksi kosakata tertinggi sebanyak 500, (4) mencari arti yang terdapat dalam bahasa pewayangan, (5) melakukan cek kosakata pada corpora. uni-leipzig.de, dan yang terakhir menyusun secara manual dalam piranti lunak *Lexique Pro* 3.6 untuk membentuk kamus saku bahasa pewayangan. Hasil yang didapat dari penelitian ini yaitu kamus saku (*pocket dictionary*) pewayangan dwibahasa Jawa-Indonesia yang berisi kosakata yang digunakan dalam pentas pewayangan.

Keywords: kamus, pewayangan, bahasa Jawa, bahasa Indonesia, korpus

I. PENDAHULUAN

1.1 Latar Belakang

Bangsa Indonesia dikenal sebagai bangsa yang kaya dengan budaya dan bahasa daerahnya. Bentangan pulau dari Sabang hingga Merauke memiliki ke khas-an bahasa daerah dan budayanya masing-masing. Salah satu fenomena yang menarik mengenai bahasa daerah di Indonesia adalah mengenai jumlah penutur bahasa Jawa. Dikatakan bahwa bahasa Jawa merupakan bahasa daerah yang memiliki penutur paling banyak diantara bahasa daerah lainnya. Dikutip dari artikel pada laman Badan Bahasa yang ditulis oleh Sugiyono dengan judul “*Pelindungan Bahasa Daerah dalam Kerangka Kebijakan Nasional Kebahasaan*” menyebutkan bahwa di antara bahasa Indonesia, terdapat tiga bahasa daerah yang memiliki jumlah penutur lebih dari 10 juta jiwa yakni bahasa Jawa (84,3 juta jiwa), bahasa Sunda (34 juta jiwa) dan bahasa Madura (13,6 juta jiwa). Hal tersebut sejalan dengan estimasi dari ethnologue.com yang menyebutkan bahwa Bahasa Jawa telah menempati peringkat ke 11 di dunia berdasarkan jumlah penutur setelah Bahasa Jerman pada urutan ke 10 dengan jumlah penutur sebanyak 90,3 juta jiwa.

Bahasa Jawa sebagai bahasa daerah yang cukup populer keberadaanya, sangat erat kaitannya dengan seni tradisional wayang kulit, dimana dalam pementasannya menggunakan bahasa Jawa sebagai bahasa pengantar. Seni wayang kulit yang merupakan warisan luhur budaya Jawa sarat dengan kekayaan kosakata. Hal ini dikarenakan wujud bahasa Jawa merupakan rangkaian dari variasi bahasa Jawa seperti Jawa *Ngoko*, *Madya*, dan *Krama* yang tidak dimiliki oleh bahasa-bahasa daerah lain di Indonesia. Variasi-variasi tersebut muncul dan digunakan seorang Dalang selama pentas seni wayang kulit berlangsung. Adapun Bahasa Jawa kuno terkadang muncul di sela-sela pertunjukan yang menimbulkan perpaduan indah dengan bahasa Jawa. Dalam seni wayang kulit, seorang Dalang memegang peran yang penting sebagai penggerak wayang selama pertunjukkan dengan tanpa disadari ia menghasilkan kosakata yang banyak dan beraneka ragam. Melalui makalah ini penulis mencoba menggali kekayaan kosakata dalam seni wayang kulit dengan cara membentuk sebuah kamus pewayangan dwibahasa Jawa-Indonesia yang nantinya dapat digunakan sebagai alat belajar bagi pemula yang ingin mendalami studi dalam bidang seni wayang kulit dan juga bagi masyarakat luas secara umum. Selain itu pembuatan kamus ini juga diharapkan dapat membantu mendokumentasikan bahasa daerah khususnya bahasa Jawa yang terdapat dalam pewayangan sehingga diharapkan mampu membantu menjaga kelestariannya.

1.2 Masalah Penelitian

Berdasarkan fenomena mengenai bahasa Jawa pada wayang kulit, dapat dirumuskan permasalahan yang akan dikaji dalam makalah ini. Pernyataan tersebut dapat dijabarkan ke dalam pertanyaan sebagai berikut:

- (1) Apa saja kosakata yang muncul dalam seni wayang kulit sebagai lemma dalam kamus pewayangan bahasa Jawa-Indonesia?

1.3 Tujuan Penelitian

Beranjak dari fenomena kebahasaan pada wayang kulit yang telah dirumuskan dalam masalah penelitian tersebut, penelitian ini bertujuan sebagai berikut:

- (1) Memperoleh kosakata yang muncul dalam seni wayang kulit yang selanjutnya digunakan untuk entri kamus pewayangan bahasa Jawa-Indonesia.

1.4 Tinjauan Pustaka

Studi mengenai perkamusan pernah dilakukan sebelumnya oleh Kwary, Rusnaningtyas dan Jurianto (2007) dan Kwary (2010). Kwary, Rusnaningtyas dan Jurianto mengkaji Pengembangan Kamus Elektronik Akuntansi Berbasis Korpus. Dari penelitian tersebut diperoleh bahwa kamus elektronik akuntansi keuangan Inggris-Indonesia yang dihasilkan berupa *freeware* dan *user friendly* dimana kamus elektronik tersebut dapat disalin oleh siapa saja tanpa harus membayar biaya royalti. Disamping itu, sifat kamus elektronik yang *user friendly* memungkinkan untuk tidak hanya digunakan namun juga mudah diperbaharui oleh siapapun. Sementara Kwary mengkaji kamus bilingual pada pengolahan bahasa. Dari penelitian tersebut dimaksudkan untuk mengetahui kamus cetak dan online Inggris-Indonesia terpopuler diantara orang Indonesia, kedua menunjukkan bagaimana kamus tersebut dimungkinkan memiliki efek negatif pada usaha pengolahan bahasa. Selanjutnya yang ketiga, mendemonstrasikan bagaimana cara mengimplementasikan pendekatan baru tersebut.

Kedua penelitian diatas sama-sama meneliti mengenai kamus. Apabila Kwary, Rusnaningtyas, dan Jurianto (2007) mengembangkan sebuah kamus akuntansi elektronik yang berupa software, maka Kwary (2010) mengkaji kamus cetak dan online yang paling populer di Indonesia beserta dampak negatif dan cara pengembangannya di masa depan. Studi ini pun mengkaji pembentukan sebuah kamus Bahasa Jawa-Indonesia. Dalam studi ini, kamus yang dibentuk merupakan kamus yang memiliki istilah khusus dalam Bahasa Jawa, khususnya dalam bidang pewayangan. Tujuan dibentuknya kamus pewayangan Bahasa Jawa-Indonesia diharapkan nantinya akan dapat dimanfaatkan oleh pengguna baik sebagai pemula pembelajar seni pewayangan maupun masyarakat secara umum. Kamus yang akan dibuat ini bila digunakan untuk pemula yang ingin belajar seni pewayangan dapat membantu dalam memahami makna kata yang banyak digunakan dalam pentas wayang kulit. Sedangkan, untuk masyarakat umum dapat digunakan dalam memahami makna kata selama pentas pewayangan dikarenakan beberapa kata ada yang jarang digunakan dalam keseharian.

II. KERANGKA TEORI

Beberapa pakar memberikan definisi mengenai kamus. Kridalaksana (1982), mendefinisikan sebagai buku referensi yang memuat daftar kata atau gabungan kata dengan keterangan mengenai berbagai segi maknanya dan penggunaannya dalam bahasa; biasanya disusun menurut abjad (dalam tradisi Yunani-Romawi menurut urutan abjad Yunani-Romawi), kemudian menurut abjad bahasa bersangkutan; dalam tradisi Arab menurut urutan jumlah konsonan. Selain itu, kamus juga dapat didefinisikan sebagai tipe paling umum sebagai referensi kerja, pertama kali digunakan sebagai “JUDUL” oleh bangsa Inggris dan Latin dalam sebuah kamus yang berjudul *Dictionary of Syr Thomas Elyot Knight* di London pada tahun 1538 dan kamus ekabahasa *English Dictionarie; or, An Interpreter of Hard English Words* di London pada tahun 1623 oleh Henry Cockeram (Hartmann & James, 1998).

Di Inggris, kamus komprehensif yang pertama dibuat adalah *Dictionary of the English Language* pada tahun 1755 oleh Samuel Johnson. Jumlah kosakata yang terdapat dalam kamus tersebut sebanyak 40.000 entri kata. Di benua Amerika, seorang bernama Noah Webster mengikuti jejak Johnson dengan menyusun sebuah kamus dua volume yang berjudul *An American Dictionary of the English Language* pada tahun 1828 yang kemudian kosakata yang terdapat dalam kamus tersebut berkembang lebih banyak sekitar 7.000 entri kata.

Di Indonesia, perkembangan kamus dimulai pada tahun 1951 dengan diterbitkannya *Kamus Indonesia* yang ditulis oleh E. St. Harahap. Berdasarkan *Kamus Besar Bahasa Indonesia IV* (2008), sejarah perkamusan di Indonesia dimulai dari sejarah leksikografi dengan mendaftar atau membuat glosarium ke kamus – kamus dwibahasa, kemudian ke kamus – kamus ekabahasa. Menurut catatan, karya leksikografi tertua dalam sejarah studi bahasa di Indonesia ialah daftar kata Cina-Melayu pada permulaan abad ke-15, yang berisi 500 lema. Daftar kata Italia-Melayu yang disusun Pigafetta (1522) termasuk pula karya leksikografi yang awal, sedangkan kamus tertua dalam sejarah Bahasa Indonesia ialah *Spraeck ende wordboek, Inde Malaysche ende Madagaskarsche Talen met vele Arabische ende Turksche Woorden* (1603) karangan Frederick de Houtman, dan *Vocabularium ofte Woortboek naer order vanden Alphabet in't Duytsch Maleysch ende Maleysche-Duytsch* (1623) karangan Casper Wiltens dan Sebastian Danckaerts. Kamus – kamus Melayu itu jelas lebih tua daripada *Lexicon Javanum* (1841), anonim, yang naskahnya tersimpan di perpustakaan Vatikan, yang dianggap sebagai kamus Jawa tertua, yakni *Nederduitsch-Maleisch en Soendasch Woordenboek* (1841) oleh A. De Wilde. Daftar kata dan kamus pelopor yang multibahasa atau dwibahasa itu kemudian diikuti oleh pelbagai daftar kata dan kamus lain yang beraneka formatnya.

Minat pada bahasa dan perkamusan pada zaman kolonial itu terbatas pada orang asing saja. Kamus yang disusun pun pada umumnya kamus bahasa asing-bahasa di Indonesia atau bahasa di Indonesai-asing. (Yang dimaksud dengan *bahasa di Indonesia*

ialah bahasa Melayu, Jawa, Bali, Sunda, Makassar, dan lain-lain – bukan *bahasa Indonesia!*). Satu – satunya kekecualian yang harus dicatat di sini ialah kamus Melayu-Jawa yang berjudul *Baoesastra Melajoe-Djawa* (1916) karangan R. Sastrasoeganda, sebagai kamus dwibahasa pertama disusun oleh putra Indonesia.

Untuk kamus ekabahasa pertama disusun oleh putra Indonesia ialah *Kitab Pengetahuan Bahasa*, yaitu *Kamus Loghat Melayu-Johor-Padang-Riau-Lingga penggal yang pertama* oleh Raja Ali Haji dari Kepulauan Riau. Pada tahun 1345 Hijriah (1928 Masehi), kamus itu tercatat pada buku yang dicetak oleh Al Ahmadiyah Press Singapura. Karena Raja Ali Haji hidup pada pertengahan abad ke-19, dapat dipastikan bahwa buku tersebut sudah beredar sekurang-kurangnya sudah berbentuk naskah – pada abad ke-19, seperti halnya buku pelajaran ejaan dan tata bahasanya, *Bustanulkatibina* (dicetak tahun 1273 H atau 1857 M). Dipandang dari teknik leksikografi sekarang, *Kitab Pengetahuan Bahasa* itu tidak dapat disebut kamus murni, tetapi boleh dianggap sebagai kamus ensiklopedis untuk pelajar. Sementara itu, *Baoesastra Djawa* (1930) karangan W.J.S. Poerwadarminta, C.. Hardjasoedarma, dan J.C. Poedjosoedira dapat dianggap sebagai pelopor perkamusan ekabahasa bahasa Jawa, seperti halnya *Kamoes Bahasa Soenda* (1948) karangan R. Satjadibrata.

Kamus memiliki keunikan masing – masing yaitu disesuaikan dengan tipe penggunaannya dan digunakan untuk apa. Sedangkan bagian pokok dari kamus yang mana termasuk cerminan kamus yang baik yaitu dapat membantu menentukan apa yang akan dimasukkan ke dalam kamus dan bagaimana materi harus terstruktur. Berdasarkan Atkins dan Rundell (2008), kebanyakan kamus memiliki dua komponen utama yaitu: kata A-Z (atau setara dengan bahasa yang tidak menggunakan alfabet Romawi) dan semua bahan yang bersifat non-linear yang dapat mengategorikan sebagai *front matter* dan *back matter*. *Front matter* dan *back matter* serta kata A-Z merupakan bagian kamus yang biasanya wajib ada dalam kamus. *Front matter* biasanya berisi kata pengantar dan ucapan terima kasih, semacam pengantar kamus tersebut, penjelasan singkat, label dan kode yang digunakan dalam teks. Selain itu *front matter* dapat berisi mini esai yang berisi tentang aspek – aspek tertentu dari bahasa tergantung dari pangsa pasar yang dituju. *Back matter* (materi akhir) sering mencakup daftar seperti, daftar kata kerja, nomer, berat dan ukuran, unsur – unsur kimia, penomoran Romawi, buku – buku dari Alkitab dan lain – lain. Selain itu juga dapat berisi peta, diagram, dan materi lainnya disesuaikan dengan kebutuhan pengguna kamus tersebut. Dalam kamus dwibahasa dapat mencakup tentang *faux amis* dan panduan praktis untuk berbagai aspek kehidupan tentang negara pengguna dua bahasa. kategori selanjutnya yaitu kata A-Z, dalam pemilihan pengisian didorong oleh profil pengguna, target pasar dari kamus, pesaingnya di pasar, dan akibat dari biaya dan anggaran.

2.1 Tipe Kamus

Ada berbagai tipe kamus dilihat dari entri dalam kamus yang dapat menjadi ciri khas dari kamus. Sebagian besar para leksikografer berfokus pada tipe kamus yang akan dibentuk di mana akan menjadi tipe apa kamus tersebut. Dalam Atkins dan Rundell (2008) terdapat 4 tipe entri yaitu standar leksikal, gramatikal, kata – kata singkatan (*abbreviation*), dan kata ensiklopedia. Berikut beberapa penjelasan tentang tipe kamus berdasarkan entrinya:

2.1.1 Entri Standar Leksikal

Entri standar leksikal yaitu dengan memasukkan kata berdasarkan kelas katanya (*part of speech*) di mana kelas kata sangat penting dalam sebuah pembentukan kamus. Setiap kata yang memiliki makna dapat disebut leksikal karena kata – kata leksikal membawa definisi makna yang penuh dan kontribusinya di dalam kalimat untuk memperkuat makna dari kalimat tersebut. Leksikografer memasukkan *headwords* ke dalam kamus yang selanjutnya diberi klasifikasi kelas katanya yang mana dalam Bahasa Indonesia terdapat verba, nomina, adjektiva, adverbialia, numeralia, partikel, dan pronomina. Entri standar leksikal dapat dilihat dalam gambar di bawah ini:

bersabar v bersikap tenang (tt pikiran, perasaan): *kita harus ~ dl menghadapi setiap masalah;*
menyabarkan v menenangkan perasaan (pikiran dsb); menenteramkan hati: *dia sedang ~ hatinya yg sedang gundah;*
penyabar n orang yg bersikap tenang (tidak terburu nafsu dan tidak lekas marah): *orang yg ~ tidak cepat marah dl menangani setiap persoalan;*
tersabar a paling tenang (tidak mudah putus asa, marah, jengkel, dsb): *ia termasuk gadis ~ di antara teman-temannya se-*

Gambar 1. dengan lema “sabar”

Dari gambar 1 yang merupakan kata “sabar” diambil dari *Kamus Besar Bahasa Indonesia IV* menunjukkan bahwa ada turunan kata yang terbentuk setelah adanya suffix. Dari kata – kata tersebut memiliki kelas kata yang berbeda, yang ditunjukkan dengan lingkaran merah yang terdapat dalam gambar 1.

Selain kelas kata dalam entri standar leksikal berupa IPA atau cara baca secara fonetik dalam bahasa tujuan. Seperti pada kamus Indonesia-Inggris terdapat cara baca kata tersebut dalam bahasa Inggris bila kamus tersebut dwibahasa.

2.1.2 Entri Kata Singkatan (*Abbreviation Entry*)

Entri kata singkatan, tidak banyak membawa informasi karena kata – kata tersebut jarang memiliki lebih dari satu arti kata dan jarang memiliki lebih dari satu wordclass. Kata singkatan menyerupai memasukkan nama yang tepat. Namun, entri kata – kata singkatan harus memiliki referensi silang dalam bentuk penuh. Biasanya karena alasan tempat yang disediakan, alasan yang diberikan hanya sekali baik dalam memasukkan singkatan atau yang dalam bentuk utuh.

2.1.3 Entri Kata Gramatikal

Entri berdasarkan gramatikal lebih mementingkan fungsi dari kata tersebut bila diterapkan ke dalam kalimat. Dalam kamus gramatikal, tidak hanya menyajikan fungsi kata dalam kalimat tetapi merancang cara paling baik untuk menyajikan informasi agar pengguna dapat memahami dengan mudah dan benar. Selain itu harus mempertimbangkan ketrampilan dan kebutuhan pengguna. *Wordclass* dan *subclass* melakukan fungsi berbeda, tidak ada struktur yang dirancang untuk entri bahasa. Dalam pembentukannya fungsi *headword* adalah pembahas dan pemberi contoh dalam cara memasukkannya, dan karena fungsi yang mungkin juga dilakukan oleh konjungsi lainnya jadi informasi tentang kata tersebut masuk ke dalam bagian kamus. Kamus yang menggunakan gramatikal entri yaitu kamus penutur asli dan kamus pembelajaran (ekabahasa dan dwibahasa) yang sangat bergantung pada catatan penggunaan ketika berhubungan dengan kata – kata gramatikal.

2.1.4 Entri Ensiklopedis

Yaitu memasukkan keterangan tentang penamaan yang tepat yang tentunya lebih ringkas dari entri leksikal dan entri gramatikal. Dalam kamus monolingual untuk hal – hal yang berbau ensiklopedis akan lebih bervariasi jumlah informasi yang diberikan. Sedangkan dalam kamus bilingual terjemahan sederhana sudah cukup memberikan informasi yang dibutuhkan pengguna. Entri ensiklopedis dapat berupa nama dari tokoh yang disingkat, sehingga memerlukan catatan khusus untuk menjelaskan.

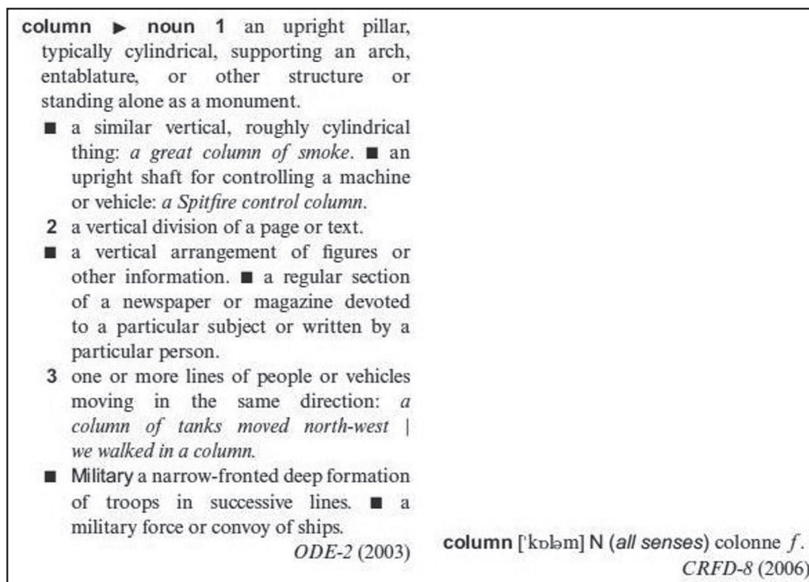
Dari berbagai macam kamus yang dijelaskan di atas, penulis mencoba membuat kamus yang bersifat standar leksikal entri karena lebih praktis digunakan dan sudah mencakup beberapa contoh penggunaannya. Kamus pewayangan ini berupa kamus dwibahasa yaitu Jawa-Indonesia yang sudah terdapat keterangan dari leksikal dan contoh secara gramatikal.

2.2 Penulisan Entri

Kamus pewayangan dwibahasa Jawa-Indonesia yang dibentuk oleh penulis dengan standar leksikal entri mengikuti aturan dalam proses penulisan entrinya. Atkins dan Rundell (2008) menyatakan bahwa ada beberapa hal yang perlu diperhatikan oleh leksikografer apabila menulis entri pada sebuah kamus yakni; menentukan arti (sense), memberikan terjemahan, memilih contoh, mengindikasikan arti, dan melengkapi entri. Berikut penjelasan mengenai aturan dalam penulisan entri diatas:

2.2.1 Menentukan Arti

Dalam menentukan arti dari *headword* pada kamus dwibahasa akan sangat berbeda dengan kamus ekabahasa. Apabila kamus ekabahasa menyesuaikan arti dengan *headword* (dimana arti kata/senses) yang ditulis sebagai arti dari *headword* dapat berupa lebih dari satu definisi yang masih berkorelasi dengan arti dari *headword* tersebut), maka kamus dwibahasa hanya menggunakan padanan kata atau kata yang setara dengan arti dari *headword* nya saja.

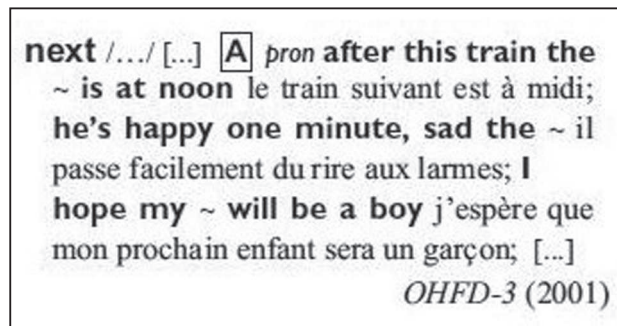


Gambar 2. Dengan lema “column”

Dua contoh diatas merupakan contoh kamus ekabahasa (ODE-2) dan kamus dwibahasa (CRFD-8). Kamus dwibahasa tidak terlalu terpaku pada bahasa sumber (*Source Language/SL*) dalam memberikan arti pada lema, tetapi secara sederhana memberikan arti kata untuk dipahami. Arti kata pada kamus dwibahasa lebih didasarkan pada bahasa sasaran (*Target Language/TL*) daripada bahasa sumber pada headword selama konten semantik pada unit leksikal terkait cukup mirip atau setara.

2.2.2 Memberikan Terjemahan

Dalam memberikan terjemahan arti dari headword pada kamus dwibahasa yakni memberikan ide yang jelas mengenai arti kata secara langsung terhadap *headword* kepada pengguna dengan mengetahui batas-batas terhadap terjemahan yang tidak sesuai, sehingga dapat mengungkapkan konsep dari *headword* dengan jelas. Adapun jenis terjemahan yang dapat digunakan sebagai pedoman para leksikografer adalah terjemahan langsung dan komponen terjemahan lain. Dalam menggunakan terjemahan langsung pada pemberian arti kata pada *headword*, pilihan kata yang digunakan dalam bahasa sasaran harus yang benar-benar setara dengan bahasa sumber. Sedangkan dalam menggunakan komponen terjemahan lain adalah sebuah teknik yang dapat digunakan dalam memberikan arti kata jika tidak didapatkan kata yang menjadi terjemahan langsung dari headword. Penggunaan arti kata yang paling dekat (*Near-Equivalent*) dan daftar istilah (*Glosses*) dapat dipilih sebagai strategi dalam memberikan arti kata. Kata “**Foreign**” pada CRFD-5 diartikan dengan memberikan kedekatan arti kata menjadi “**Foreign Secretary**” yang berarti ministre dalam bahasa Perancis. Adapun strategi yang paling bagus dalam memberikan arti kata yang tidak didapatkan terjemahannya secara langsung adalah dengan memberikan terjemahan berupa contoh dalam kalimat dengan disertai arti dalam bahasa sasarannya. Kasus tersebut dapat dilihat pada gambar berikut.



Gambar 3. Dengan lema “next”

Kata *next* pada kamus OHFD-3 diatas dapat diberikan arti dengan memberikan contoh kalimat dan diterjemahkan langsung ke dalam bahasa sasarannya yaitu Perancis. Startegi tersebut dapat memiliki kegunaan dalam mengatasi ketidakhadiran arti kata dari terjemahan langsung yang memang tidak ada.

2.2.3 Memilih Contoh

Pemberian contoh dari kata utama sebagai lema (*headword*) yang bertujuan untuk memberikan kemudahan kepada pengguna kamus, maka para leksikografer perlu memperhatikan beberapa aspek seperti arti kata yang seperti apakah yang akan diterjemahkan dari kata utama (*headword*), meyakinkan pengguna terhadap penggunaan

terjemahan langsung, melengkapi (atau mengganti) terjemahan langsung; bilamana terjemahan tersebut tidak dapat digunakan, atau sederhananya bila pengguna kamus membutuhkan kata yang berbeda, dan yang terakhir adalah menentukan arti dari kata yang merupakan bahasa sasaran (TL) yang memiliki definisi lebih dari satu.

2.2.4 Mengindikasi Arti

Sebelum diputuskan untuk memberikan arti pada kata yang akan dientri, maka melakukan indikasi terhadap arti kata sangatlah penting. Hal ini dimaksudkan guna mengantisipasi arti kata yang bersifat komplikasi dan sulit untuk diberikan artinya. Ada tiga tipe utama yakni *specifiers*, *kolokator*, dan juga *domain label* yang disebutkan oleh Atkins dan Rundell (2008) yang dapat digunakan sebagai cara mengindikasi arti kata. *Specifiers* dijelaskan berwujud sebagai *sinomin* atau *paraphrase* dalam menjelaskan arti sebuah kata utama (*headword*), sedangkan *kolokator* berupa kata yang selalu mengikuti kata utama (*headword*), dan yang terakhir adalah *domain label* yang berfungsi untuk membantu penutur bahasa sumber dalam memahami konsep dari *domain label headword* seperti yang dicontohkan pada label “*Ind*” untuk kata “*Industry*” pada kata gabungan kata “*design office*”, sehingga penutur bahasa sumber memahami konsep *design* kantor seperti apakah yang dimaksud.

2.2.5 Melengkapi Entri

Tahap ini merupakan tahap akhir dalam penulisan entri sebuah kamus. Pada tahap ini seluruh proses penentuan arti (*sense*) pada kamus, pemberian contoh, hingga memberikan terjemahan telah selesai dilakukan. Untuk memastikan bahwa entri seluruh kosakata dengan definisinya telah dilakukan dengan baik, maka ada baiknya untuk membaca dan meninjau kembali hasil entri pada kamus. Waktu terbaik untuk membaca kembali adalah setelah satu atau dua minggu setelah proses entri selesai.

III. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode deskriptif kualitatif yang bertujuan untuk memberikan penjelasan secara detail tentang permasalahan dalam penelitian terkait dengan data dan fonema. Adapun teknik pengolahan data dalam penelitian ini yakni melalui beberapa tahap; Pertama, mentranskrip rekaman dalam bentuk tulisan pada *Ms. Word* dan disimpan dalam bentuk file txt (plain text), Kedua, memasukkan hasil transkripsi ke alat *Ant Cont 3.4* untuk mendapatkan kosakata tertinggi yang sering muncul dalam pentas seni wayang kulit sebanyak 500 kata. Ketiga, menyeleksi kosakata tertinggi sebanyak 30 kosakata sebagai sampel kosakata yang berkaitan dengan bidang pewayangan. Untuk memastikan kosakata yang akan dipilih sebagai lema, penulis terlebih dahulu melakukan cek kosa kata melalui <http://corpora.uni-leipzig.de> untuk

memastikan bahwa kosakata tersebut merupakan kosakata berfrekuensi tinggi. Keempat mencari arti dari kosakata yang telah dipilih. Dalam hal ini, penulis secara langsung terjun ke lapangan menemui seorang informan yang tidak lain adalah seorang Dalang wayang kulit guna membantu memberikan arti kata. Selanjutnya, langkah terakhir yang dilakukan adalah menyusun secara manual satu per satu kosakata dalam piranti lunak *Lexique Pro 3.6* untuk membentuk kamus saku (*pocket dictionary*) pewayangan.

IV. ANALISIS PENELITIAN

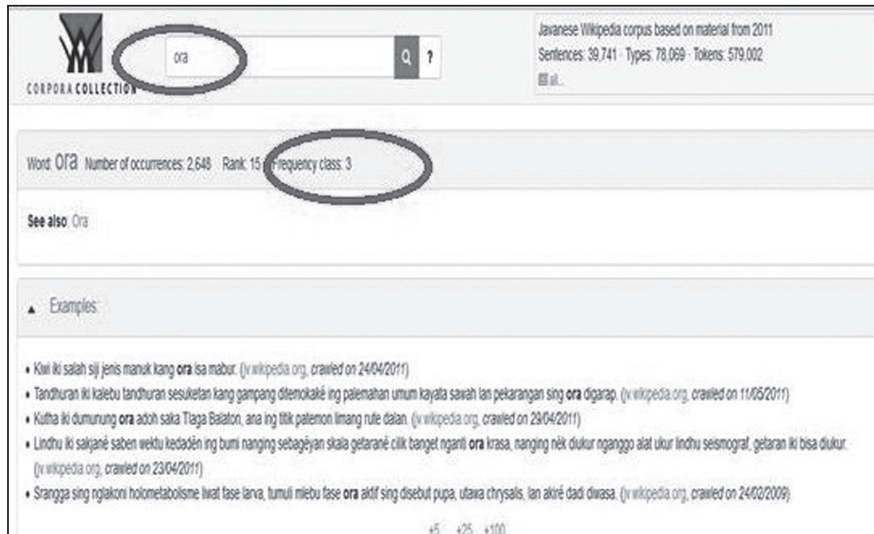
Pembuatan kamus berbasis korpus sebagai salah satu metodologi dalam pemilahan data yang dilakukan oleh penulis dengan cara mengambil kosakata yang berasal dari pementasan wayang kulit. Oleh karenanya terdapat kosakata yang menggunakan Bahasa Jawa ataupun Bahasa Jawa Kuno. Bahasa-bahasa tersebut tetap dilestarikan dalam pementasan wayang kulit karena dapat digunakan sebagai media pembelajaran. Berdasarkan data yang didapat dari piranti lunak *Ant Conc 3.4*, maka kosakata yang muncul dapat dilihat sebagai berikut:

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Word Types: 2524		Word Tokens: 8679		Search Hits: 0		
Rank	Freq	Word	Lemma Word Form(s)			
1	161	kula				
2	101	aku				
3	93	tak				
4	80	ora				
5	75	kakang				
6	73	amarta				
7	72	apa				
8	72	kowe				
9	69	prabu				
10	57	yen				
11	56	negara				
12	56	raden				
13	54	ana				
14	53	inggih				
15	53	ndara				
16	52	sing				
17	45	va				

Gambar 4. Kosakata pada piranti lunak *AntConc 3.4*

Hasil dari transkripsi yang telah diproses ke dalam piranti lunak *AntConc 3.4* diperoleh total kosakata sebanyak 2.524 kata. Gambar diatas merupakan wujud kosakata yang muncul secara urut dari frekuensi tinggi ke frekuensi yang paling rendah. Terlihat bahwa kata seperti *kula*, *aku*, *tak*, *ora* berada pada urutan paling atas. Sehubungan dengan pemilihan kosakata yang akan dimasukkan ke dalam entri, maka kata-kata tersebut tidak begitu saja dipilih sebagai lema. Terlebih dahulu dilakukan cek kosakata satu persatu di <http://corpora.uni-leipzig.de> guna memastikan bahwa kosakata tersebut tidak termasuk sebagai kata yang bersifat umum di dalam penggunaannya sehari-hari.

Apabila ditemukan bahwa kata tersebut memiliki kelas frekuensi rendah, maka tidak akan dipilih sebagai lema. Hal tersebut mencerminkan bahwa dengan rendahnya frekuensi kelas sebuah kata, maka kata tersebut merupakan kata yang umum digunakan. Berikut contoh kosakata yang telah dilakukan cek pada corpora.uni-leipzig.



Gambar 5. Kata “ora” dengan kelas frekuensi rendah

Kata “ora” sebagai salah satu contoh kosakata yang dibuktikan sebagai kata yang memiliki kelas frekuensi rendah dengan menunjukkan angka 3. Jika ditinjau kembali kata tersebut, maka dalam percakapan sehari-hari kata “ora” bersifat umum dan sering dijumpai dalam percakapan sehari-hari. Dengan melakukan cek kosakata pada corpora.uni-leipzig.de dapat membantu penulis dalam pemilihan kosakata secara akurat. Sehingga dalam proses entri kata, penulis tidak memasukkan kata “ora” sebagai lema dalam kamus pewayangan. Selanjutnya kata yang memiliki kelas frekuensi tinggi juga dilakukan cek kosakata. Adapun contoh kata tersebut dapat dilihat sebagai berikut.



Gambar 6. Kata “kakang” dengan kelas berfrekuensi tinggi

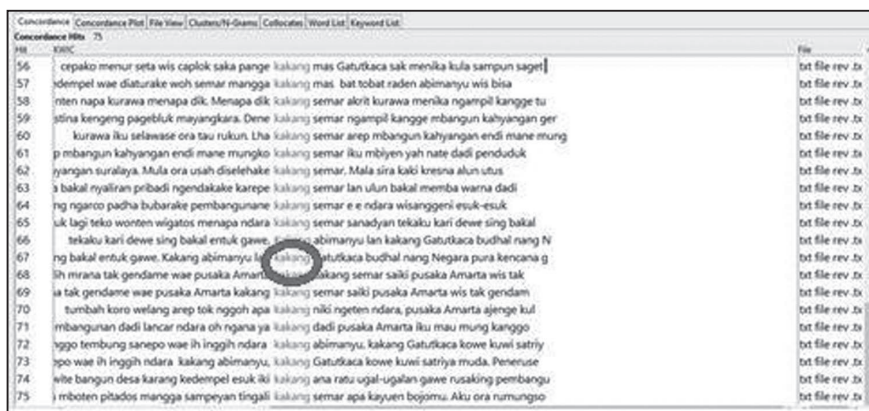


Gambar 7. Kata “pagebluk” dengan kelas frekuensi tinggi

Kata “kakang” dan “pagebluk” ditemukan sebagai kata yang memiliki kelas frekuensi kata yang tinggi yakni 11 dan 13. Hal ini mencerminkan bahwa kata tersebut sedikit digunakan dalam percakapan sehari-hari, bersifat khusus dan digunakan dalam keperluan secara khusus dalam pentas seni wayang kulit. Kata-kata tersebut yang kemudian dipilih sebagai lema dalam kamus pewayangan bahasa Jawa-Indonesia. Tahap selanjutnya setelah lema yang dipilih telah terkumpul adalah dengan memberikan arti kata terhadap lema. Pemberian arti yang dilakukan oleh penulis yaitu berdasarkan entri standar leksikal. Penulis memasukkan kata berdasarkan kelas katanya (*part of speech*) kemudian memberikan definisi maknanya. Lema atau headwords selanjutnya diberi klasifikasi kelas katanya seperti verba, nomina, adjektiva, adverbial, numeralia, partikel, dan pronomina.



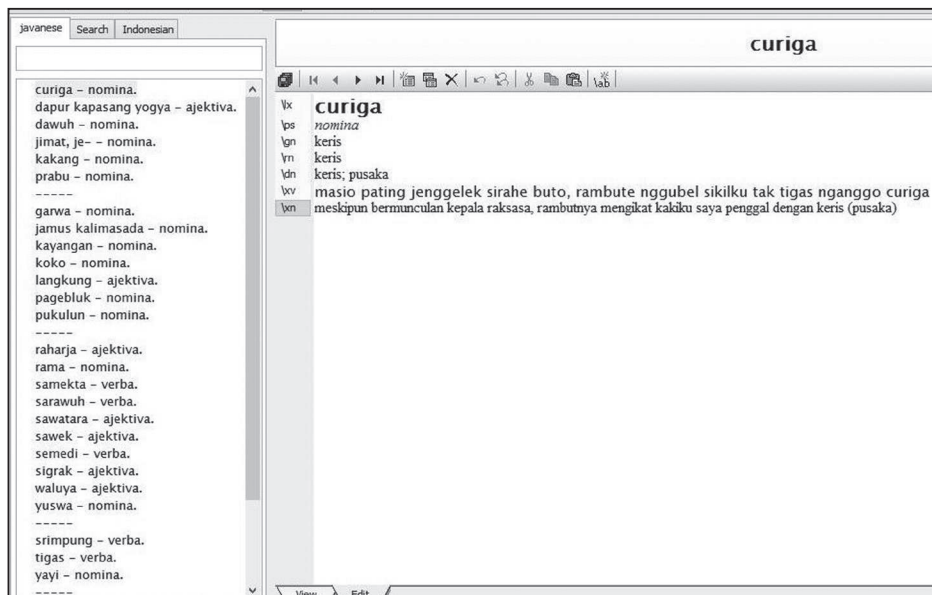
Gambar 8. lema “ulun” merupakan nomina



Gambar 9. Kata “kakang” dalam Concordance

Setelah lema sudah dientri dan diberikan kelas katanya, maka penulis memberikan contoh kalimat dalam entri tersebut. Contoh kalimat yang dimasukkan dalam entri merupakan kumpulan kalimat yang terdapat dalam concordance. Kalimat tersebut disederhanakan kembali kedalam kalimat guna memudahkan pengguna kamus mudah dalam memahami katanya. Seperti contoh kata “kakang” yang terdapat pada Gambar. 9 diatas dijadikan contoh kalimat dalam entri lema “kakang” pada kamus. Jika kalimat dinilai mudah, maka tidak dilakukan penyederhanan seperti kata “kakang” yang terdapat dalam kalimat “*kakang Abimanyu lan kakang Gatutkaca budhal nang Negara Pura Kencana*” dan kemudian diterjemahkan kedalam bahasa Indonesia menjadi “*kakak Abimanyu dan kakak Gatutkaca berangkat ke Negara Pura Kencana.*”

Kosakata yang telah diseleksi oleh penulis sebanyak 500 kata kemudian satu persatu mulai dilakukan entri ke dalam piranti lunak *Lexique Pro 3.6* untuk mulai dibuat kamus. Proses entri kosakata setelah dimasukkan satu persatu ke dalam piranti lunak *lexique pro* dapat dilihat sebagai berikut:



Gambar 10. Hasil entri dalam Lexique Pro 3.6

Tabel di atas merupakan contoh entri kosakata yang telah dimasukkan satu persatu ke dalam piranti lunak *lexique pro* yang secara otomatis akan tersusun secara alfabetikal. Kosakata yang terdapat dalam piranti lunak di atas terlihat bahwa kosakata yang dipilih sebagai lemma secara khusus berupa kosakata yang berhubungan dengan bidang pewayangan. Seluruh lemma yang telah dientri satu persatu kemudian akan disimpan ke dalam bentuk Ms. word yang nantinya akan dibuat sebagai kamus cetak pewayangan bahasa Jawa-Indonesia. Adapaun wujud kamus cetak yang telah dibentuk ke dalam Ms. word dapat dilihat pada gambar sebagai berikut:

G – g	
<p>garwa <i>nomina</i>. seseorang yang dinikahi (laki-laki atau perempuan). duh rama kula menika mikiraken garwa kula raden Tetuka wahai ayah sesungguhnya saya sedang memikirkan suami saya raden Tetuka.</p>	
J – j	
<p>jimat, je- <i>nomina</i>. barang-barang yang dianggap memiliki kekuatan. pangestunipun kanjeng koko prabu mugi andadosaken jejimat memohon doanya yang mulia kakak prabu semoga menjadikan azimat.</p>	<p>jamus kalimasada <i>nomina</i>. sebutan untuk nama pusaka. kaya ora gampang nyolong Jamus Kalimasada sepertinya tidak mudah mencuri Jamus Kalimasada.</p>
K – k	
<p>kakang <i>nomina</i>. kakak; saudara laki-laki yang lebih tua. kakang Abimanyu lan kakang Gatotkaca budhal nang negara Pura Kencana kakak Abimanyu dan kakak Gatotkaca berangkat ke negara Pura Kencana.</p>	<p>kayangan <i>nomina</i>. tempat tinggal para dewa; surgaloka. monggo kula dereaken tindak dateng kayangan mari saya antarkan berangkat ke surgaloka.</p> <p>koko <i>nomina</i>. kakak. koko prabu sami raharja sarawuh paduka wonten negari Ngastina? kakak prabu apakah semua selamat atas kedatanganmu di negara Ngastina?</p>

Gambar 11. Hasil Lexique Pro dalam Microsofr Word

Gambar di atas merupakan wujud kamus cetak pewayangan bahasa Jawa-Indonesia yang telah diolah menjadi sebuah kamus cetak yang nantinya dapat digunakan oleh pengguna. Kamus pewayangan bahasa Jawa-Indonesia yang dibentuk oleh penulis tidak hanya memberikan arti dalam bentuk sinonim dan uraian singkat, namun kamus ini dibuat disertai contoh penggunaan kata dalam sebuah kalimat. Hal ini bertujuan untuk memberikan pemahaman yang lebih dalam tentang penggunaannya dalam sebuah kalimat, sehingga pengguna tidak hanya memahami arti namun juga dapat mengaplikasikan dalam bentuk kalimat yang nantinya dapat digunakan sebagai penunjang wawasan baik secara lisan maupun tulisan.

V. SIMPULAN

Dewasa ini, kesenian wayang kulit jarang diadakan dan jarang pula orang yang mengerti dengan bahasa yang digunakan selama pertunjukan wayang kulit. Ketidapahaman masyarakat dikarenakan bahasa yang digunakan berbeda dengan bahasa yang digunakan sehari – hari. Pertunjukannya jarang diadakan karena biaya pengadaan yang mahal dan sudah jarang generasi muda yang menyukai pertunjukan wayang kulit. Alasan tersebut dapat dibuktikan dengan sudah jarang pemnetasan wayang kulit dan penonton yang rata-rata didominasi orang tua.

Bahasa yang digunakan selama pementasan wayang kulit terdiri dari Bahasa Jawa dan Bahasa Jawa Kuno. Bahasa tersebut telah dipelajari secara turun temurun dan telah dibukukan untuk mempermudah dalam pembelajaran. Untuk Bahasa Jawa telah adanya

kamus Bahasa Jawa-Indonesia ataupun Bahasa Jawa Monolingual yang diterbitkan oleh beberapa institusi. Sedangkan penulis mencoba untuk membuat contoh dan cara dalam pembuatan bahasa yang digunakan dalam kesenian wayang kulit. Bahasa yang digunakan berbeda menurut makna leksikal maupun secara grammatikal penggunaannya. Bila dalam kamus Bahasa Jawa-Indonesia atau Bahasa Jawa Monolingual berbentuk kumpulan kosakata dari bahasa tersebut, tetapi dalam kamus pewayangan Jawa-Indonesia yang penulis buat tidak dapat berupa satu kosakata. Dalam kamus ini penulis menemukan beberapa kata yang tidak dapat berdiri sendiri karena tidak memiliki arti bahkan dapat berarti berbeda dalam penerapan kesenian wayang kulit. Kamus pewayangan ini merupakan kamus teknik yang dihasilkan dari linguistik terapan yaitu leksikografi.

Dari pembuatan kamus ini diharapkan dapat membantu pelajar yang mempelajari kesenian tradisional dan masyarakat luas. Pelajar yang mempelajari kesenian tradisional diharapkan dengan mengetahui makna dari kosakata yang dipelajarinya sehingga dapat mempertahankan dan mengembangkan kesenian tradisional seperti wayang kulit yang sudah mulai jarang dikenal dan diminati. Sedangkan untuk masyarakat umum, dengan adanya kamus ini diharapkan dapat membantu untuk mengerti kosakata yang digunakan selama pertunjukan wayang kulit. Semoga dengan adanya kamus ini dapat membantu pelajar maupun masyarakat luas.

Daftar Pustaka

- Atkins, B.T., Rundell, Michael. 2008. *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press.
- Hartmann, R. R. K., & James, G. (1998). *Dictionary of lexicography*. Psychology Press.
- Kwary, D. A. (2010). *Bilingual Dictionaries in Language Cultivation*.
- Kwary, D. A, Rusnaningtyas, E, & Jurianto. (2007). *Pengembangan Kamus Elektronik Akuntansi Berbasis Korpus*.
- Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Sugiono, Dendy. (2008). *Kamus Besar Bahasa Indonesia Pusat Bahasa*: Edisi Keempat. Jakarta: PT. Gramedia Pustaka Utama
- Wilkins, D.A. 1972. *Linguistics and Language Teaching*. London: Edward Arnold.
- <http://badanbahasa.kemdikbud.go.id/lamanbahasa/content/pelindungan-bahasa-daerah-dalam-kerangka-kebijakan-nasional-kebahasaan>
- <http://corpora.uni-leipzig.de>
- <https://www.ethnologue.com/statistics/size>

Redesain Kamus Bahasa Jawa Berdasarkan Tingkatan Bahasa Jawa (Bahasa Jawa Ngoko, Krama Madya, Krama Inggil)

Afwin Sulistiawati, S.Pd.

Universitas Airlangga
afwince@gmail.com

Abstrak

Penelitian ini dilatarbelakangi oleh realita bahwa kamus-kamus bahasa Indonesia-bahasa Jawa atau bahasa Jawa-bahasa Indonesia yang sudah ada hanya menyuguhkan arti kata saja. Misalnya dari kata berbahasa Jawa diartikan ke dalam bahasa Indonesia atau sebaliknya. Sedangkan pengguna kamus sering kali menemukan kesulitan ketika akan menggunakan bahasa Jawa karena arti yang ada di dalam kamus hanyalah satu jenis bahasa Jawa saja. Sedangkan bahasa Jawa sendiri memiliki tiga tingkatan bahasa yang bahkan memiliki kata-kata yang berbeda dari setiap tingkatan bahasa Jawa tersebut. Adanya tiga tingkatan tersebut berdasarkan tingkat kesopanan. Misalnya dalam bahasa Indonesia kata ‘pakai’ bisa diartikan menjadi 3 kata yang berbeda sesuai tingkatan bahasa Jawa, yakni ‘anggo’, ‘angge’, ‘agem’ (bahasa Jawa Ngoko, Krama Madya, Krama Inggil). Hal itulah yang belum ada dalam perkamusan bahasa Jawa padahal dalam penggunaannya, itu sangat diperlukan. Bila bahasa Inggris memiliki perubahan kata kerja dipengaruhi oleh waktu, dalam bahasa Jawa terdapat perubahan kata dipengaruhi oleh tingkat kesopanan atau tingkatan bahasa Jawa.

Oleh karena itu, penelitian ini bertujuan mengumpulkan sebanyak-banyaknya kata yang memiliki perubahan berdasarkan tingkatan bahasa Jawa untuk kemudian dimasukkan ke dalam kamus. Artinya, jika ingin mengetahui penggunaan suatu kata dalam bahasa Jawa, maka kamus akan bisa membantu dengan cara menyuguhkan kosakata-kosakata dalam tiga tingkatan bahasa Jawa. Kosakata-kosakata itu didapat dari 10 sumber data (manula di Yogyakarta yang dipercaya sebagai pengguna bahasa Jawa sesuai dengan tiga tingkatan bahasa Jawa [empunya bahasa Jawa]). Data penelitian ini berupa hasil perumusan kosakata sesuai tingkat kesopanan atau tingkatan bahasa Jawa. Secara akademik, penelitian ini diharapkan bermanfaat bagi

pengembang leksikografi, terlebih bagi pembuat dan pengguna kamus bahasa Jawa. Secara praktis, penelitian ini diharapkan dapat menambah pengetahuan di bidang kosakata berdasar tiga tingkatan bahasa Jawa. Dari hasil analisis, diperoleh bahwa tidak semua kata memiliki perbedaan kosakata berdasar tingkatan bahasa Jawa dan telah didapatkan 430 kosakata yang dapat disajikan di dalam kamus. Data-data tersebut siap dimasukkan ke dalam kamus bahasa Jawa untuk membantu para penggunanya.

Keyword: bahasa Jawa Ngoko, Krama Madya, Krama Inggil

I. PENDAHULUAN

1.1 Latar Belakang

Kamus-kamus bahasa Indonesia-bahasa Jawa atau bahasa Jawa-bahasa Indonesia yang sudah ada hanya menyuguhkan arti kata saja. Baik kamus *online*, *offline*, maupun kamus cetak. Pengertian itu misalnya dari kata berbahasa Jawa diartikan ke dalam bahasa Indonesia atau sebaliknya. Misalnya kata ‘mata’ di dalam kamus berarti netra; mripat; pandeleng. Pengertian semacam itu hanya menunjukkan kosakata dalam beda bahasa saja, misal bahasa Jawa dialihkan ke bahasa Indonesia atau sebaliknya, tanpa ada tambahan pengetahuan mengenai kosakata berdasarkan tingkatan bahasa Jawa, devinisi, atau contoh penggunaannya. Sedangkan pengguna kamus sering kali menemukan kesulitan ketika akan menggunakan bahasa Jawa karena arti yang ada di dalam kamus hanyalah satu jenis bahasa Jawa saja.

Kalaupun ada aplikasi kamus yang bisa dibeli di Play-Store yang menyuguhkan tingkatan kata dalam bahasa Jawa tetapi hanya kata-katanya saja yang disuguhkan. Di dalamnya tidak terdapat contoh atau definisi dari kata yang sedang dicari sehingga belum cukup membantu para pengguna kamus bahasa Jawa-bahasa Indonesia atau bahasa Indonesia-bahasa Jawa.

Bahasa Jawa sendiri memiliki tiga tingkatan bahasa yang bahkan memiliki kata-kata yang berbeda dari setiap tingkatan bahasa Jawa tersebut. Adanya tiga tingkatan tersebut berdasarkan tingkat kesopanan, yakni bahasa Jawa Ngoko yang biasa digunakan dalam keadaan santai dan cenderung tidak sopan, Krama Madya yang cenderung setengah sopan, dan Krama Inggil yang bisa dikatakan sangat sopan. Pembagian tingkatan bahasa Jawa tersebut merupakan salah satu budaya berbahasa yang ada di Jawa.

Dalam bahasa Jawa ada pepatah “*Ajining dhiri dumunung ana ing lathi*”, artinya bahasa merupakan cerminan jati diri si pembicara. Sapir (dalam Blount, 1974) menyatakan bahwa kandungan tiap budaya terwujud di dalam bahasanya. Oleh karena itu tingkatan

bahasa Jawa ini dapat dikatakan sebagai bagian dari budaya Jawa. Contoh tingkatannya misalnya dalam bahasa Indonesia kata ‘pakai’ bisa diartikan menjadi 3 kata yang berbeda sesuai tingkatan bahasa Jawa, yakni ‘anggo’, ‘angge’, ‘agem’ (bahasa Jawa Ngoko, Krama Madya, Krama Inggil) (Nugroho, TT: 80). Hal itulah yang belum ada dalam perkamusan bahasa Jawa padahal dalam penggunaannya, itu sangat diperlukan. Oleh karena itu penelitian ini perlu dilakukan untuk memenuhi kebutuhan pengguna kamus bahasa Jawa dan untuk mengembangkan leksikografi bahasa daerah sehingga rumusan masalah dalam penelitian ini adalah bagaimana bentuk redesain kamus bahasa Jawa berdasarkan tingkatan bahasa Jawa (bahasa Jawa Ngoko, Krama Madya, Krama Inggil).

1.2 Rumusan Masalah

Berdasar pada judul di muka, rumusan masalah yang akan dibahas dalam penelitian ini adalah bagaimana bentuk redesain kamus bahasa Jawa berdasarkan tingkatan bahasa Jawa (bahasa Jawa Ngoko, Krama Madya, Krama Inggil).

1.3 Tujuan

Tujuan dari diadakannya penelitian ini adalah untuk mendeskripsikan dan menunjukkan bentuk redesain kamus bahasa Jawa berdasarkan tingkatan bahasa Jawa (bahasa Jawa Ngoko, Krama Madya, Krama Inggil) sehingga penelitian ini akan bisa bermanfaat bagi pengguna kamus.

1.4 Tinjauan Pustaka

Penelitian terdahulu yang relevan dengan penelitian ini adalah penelitian yang dilakukan oleh Agustinus Ngadiman dengan judul *Tingkat Tutur Bahasa Jawa: Wujud Kesantunan Manusia Jawa (Dulu dan Sekarang)* pada tahun 2006. Kesamaan dengan penelitian tersebut ialah sama-sama meneliti tentang bahasa Jawa dan bahkan kesantunannya tetapi fokus penelitiannya berbeda, penelitian tersebut terfokus pada aspek tindak tutur masyarakat Jawa sedang penelitian ini terfokus pada tingkatan bahasa Jawa di dalam kamus. Penelitian ini merupakan penelitian yang berjenis penelitian kualitatif dengan menyajikan hasil penelitian berupa kata-kata dalam perkamusan.

Sumber data yang digunakan dalam penelitian ini adalah 10 manula yang ada di Yogyakarta dan merukan empunya bahasa Jawa karena merekalah yang menggunakan tiga tingkatan bahasa Jawa itu dalam kehidupan sehari-harinya. Sedangkan data yang dipakai dalam penelitian ini adalah tuturan mengenai pengartian kata dalam bahasa Indonesia ke bahasa Jawa atau sebaliknya, mereka juga menunjukkan bagaimana kosakata-kosakata sesuai dengan tingkatan bahasa Jawa. Teknik pengumpulan data dalam penelitian ini adalah teknik rekam, simak dan transkripsi.

II. KERANGKA TEORI DAN METODE PENELITIAN

2.1 Kerangka Teori

Leksikografi sering diartikan sebagai cabang ilmu bahasa yang mempelajari teknik penyusunan kamus. Kegiatan yang terlibat dalam ilmu leksikografi di antaranya adalah perancangan, kompilasi, penggunaan, serta evaluasi suatu kamus. Namun terdapat juga pandangan bahwa leksikografi merupakan satu cabang linguistik terapan dan garapannya mencakup teknik dan metodologi penyusunan, seperti yang ditegaskan oleh Kridalaksana (1985). Juga ada pendapat lain mengenai leksikografi yakni leksikografi merupakan satu cabang linguistik terapan yang mencakup pemerhatian, pengumpulan data, pemilihan data dan membuat pemerian unit dan stok perkataan serta kombinasi perkataan (frasa) dalam satu atau lebih bahasa.

Selain pendapat-pendapat di muka, leksikografi juga mencakup perkembangan dan deskripsi teori dan kaidah yang menjadi dasar kegiatan penyusunan kamus. Ini melibatkan bentuk leksikal unit semantik dalam bahasa, biasanya disusun berdasar abjad. Malkiel (dalam Sariyan, 2014: 363) juga mengemukakan bahwa leksikografi, dalam makna yang sempit, merupakan subbagian linguistik terapan yang berkaitan dengan teknik mengumpulkan data leksikal dan mempersembahkannya dalam bentuk kamus. Oleh karena itu, berterima jika leksikografi diartikan sebagai cabang ilmu bahasa yang mempelajari teknik penyusunan kamus.

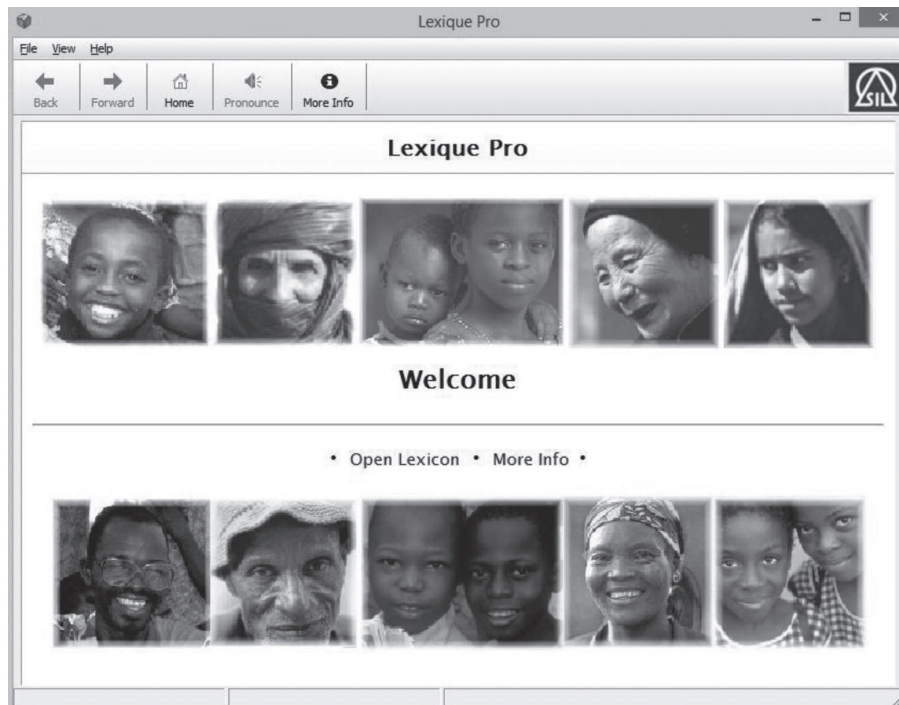
Adapun definisi kamus juga banyak dikemukakan oleh para ahli. Salah satunya adalah Kridalaksana (1985), mendefinisikan kamus sebagai buku referensi yang memuat daftar kata atau gabungan kata dengan keterangan mengenai berbagai segi maknanya dan penggunaannya dalam bahasa; biasanya disusun menurut abjad (dalam tradisi Yunani-Romawi menurut urutan abjad Yunani-Romawi), kemudian menurut abjad bahasa bersangkutan; dalam tradisi Arab menurut urutan jumlah konsonan.

Pembuatan kamus dapat dilakukan dengan berbagai cara. Dalam penelitian ini menggunakan software Lexique Pro. *Lexique Pro is an interactive lexicon viewer, with hyperlinks between entries, category views, dictionary reversal, search, and export tools. It can be configured to display your Toolbox/Shoebox database in a user-friendly format so that you can distribute it to others* (http://www-01.sil.org/computing/catalog/show_software.asp?id=92). Pemakaian Lexique Pro ini karena Lexique Pro merupakan penampil leksikon interaktif, dengan sistem *hyperlink* antara entri, tampilan kategori, pembalikan kamus, pencarian, dan alat-alat ekspor dalam bentuk kamus cetak ataupun kamus online. Hal ini dapat dikonfigurasi untuk menampilkan database dalam format yang sederhana dan memudahkan sehingga pengguna Lexique Pro dapat mendistribusikannya kepada orang lain.

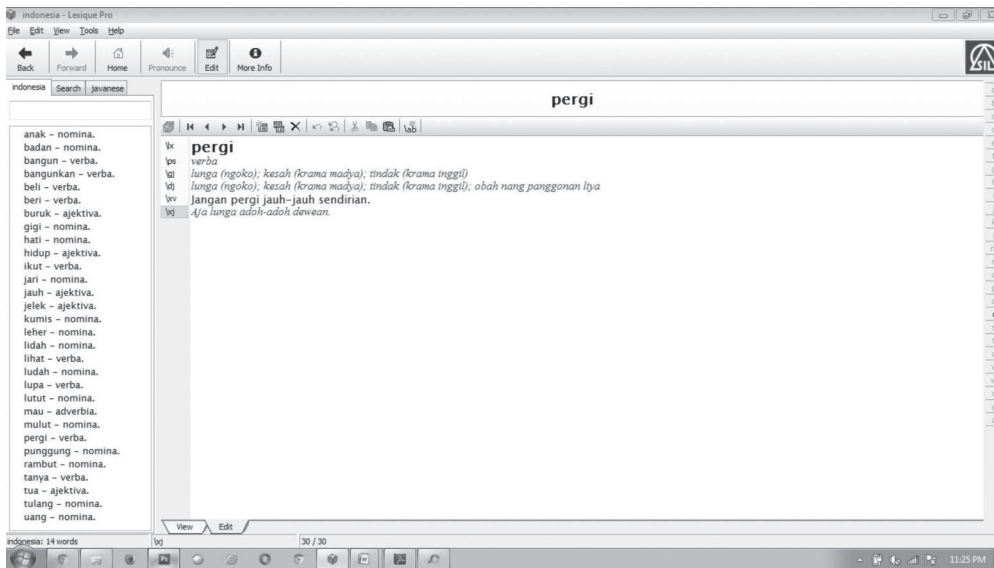
2.2 Metode Penelitian

Analisis data yang digunakan yakni metode deskriptif kualitatif yang terdiri atas tiga proses yakni reduksi data, penyajian data, dan penarikan simpulan. Penelitian kualitatif menurut Moleong (2007:6) adalah penelitian yang bermaksud untuk memahami fenomena tentang apa yang dialami oleh subjek penelitian misalnya perilaku, persepsi, motivasi, tindakan, tuturan, dan lainnya secara holistik, dan dengan cara deskripsi dalam bentuk katakata dan bahasa, pada suatu konteks khusus yang alamiah dan dengan memanfaatkan berbagai metode alamiah. Data yang disajikan dalam pembahasan hanya data sampel berupa 30 data sebagai contoh bentuk kamusnya.

Penelitian ini menggunakan *software* atau alat Leksikografi berupa Lexique Pro Versi 3.6. Lexique Pro merupakan *software* atau alat untuk membuat kamus dengan berbagai pilihan bahasa bahkan bahasa daerah. Penelitian ini menggunakan bahasa Indonesia sebagai bahasa sumber dan bahasa Jawa sebagai bahasa untuk pendefinisian atau pengartian. Wujud Lexique Pro dapat dilihat pada gambar berikut ini.



Gambar tersebut adalah tampilan awal Lexique Pro. Kemudian pengguna bisa mengatur bahasa-bahasa yang akan digunakan sesuai kebutuhan. Karena penelitian ini adalah redesain bahasa Jawa maka bahasa yang digunakan adalah bahasa Indonesia-bahasa Jawa dan akan muncul pada contoh berikut.



Setelah data-data sudah dimasukan sesuai dengan kodenya, misal *gn* berarti *gloss*, *lx* berarti *lexeme*, *dn* berarti *devinition*, dan lainnya. Yang menentukan jenis dan banyaknya kode-kode itupun pengguna Lexique Pro karena penggunalah yang mengatur apa saja yang akan ada dalam kamus yang sedang dibuatnya. Setelah proses pemasukan data kemudian data-data dalam Lexique Pro itu bisa diekspor atau diubah dalam bentuk kamus cetak (dokumen) atau kamus *online*.

III. ANALISIS PENELITIAN

Desain berarti kerangka bentuk (KBBI, 2007). Redesain berarti membangun kembali suatu kerangka bentuk, dalam hal ini kerangka bentuk kamus bahasa Indonesia-bahasa Jawa atau sebaliknya. Kamus yang sudah ada biasanya hanya menunjukkan arti kata dalam bahasa yang dituju, misalnya ketika ingin mencari kata '*makan*' dalam kamus bahasa Indonesia-bahasa Jawa maka akan muncul '*mbadhog, mangan, maem, dahar*', atau kata '*depan*' dalam bahasa Indonesia akan diartikan ke dalam bahasa Jawa di dalam kamus menjadi kata '*arsa, ngarsa, andhap*'. Jika kasus yang ada semacam itu, pengguna kamus akan merasa bingung ketika akan menggunakan bahasa Jawa yang santun sesuai tiga tingkatan bahasa Jawa.

Oleh karena alasan di atas, perlulah diadakan redesain kamus yang ada dengan cara menambahkan keterangan tingkatannya sehingga pengguna kamus dan pengguna bahasa Jawa akan terbantu melalui kamus. Data-data mentah yang sudah didapat dari narasumber yang berupa empunya dan pengguna bahasa Jawa dengan tiga tingkatan akan diubah menjadi bentuk kamus. Dari 10 sumber data dapat diperoleh 430 kosakata yang dapat diuraikan tingkatan bahasa Jawanya. Di antara data-data mentahnya ialah

No	Bahasa Indonesia	Bahasa Jawa		
		Ngoko	Krama Madya	Krama Inggil
1	Pergi	Lunga	Kesah	Tindak
2	Perempuan	Wadon	Estri	Estri
3	Lupa	Lali	Supe	Kalimengan
4	Uang	Duwit	Yatra	Artha
5	Mau	Gelem	Purun	Kersa
6	Badan	Awak	Badhan	Salira
...

Berdasar beberapa data mentah di muka, dapat diketahui bahwa tidak semua kata memiliki kosakata yang berbeda pada setiap tingkatan bahasa Jawanya. Hal tersebut wajar terjadi karena tidak semua kata memiliki kesantunan yang berdeda, akan tetapi ada beberapa kata yang memiliki tingkat kesantunan yang sama dengan tingkatan bahasa Jawa yang lain. Misalnya kata ‘estri’ dalam bahasa Jawa Krama Madya sama dengan bahasa Jawa Krama Inggil yang artinya adalah ‘perempuan’. Juga terjadi pada beberapa kata yang lain. Adapun contoh dalam bentuk kamus yang sudah diproses dengan Lexique Pro adalah sebagai berikut.

anak

ikut

A – a

anak *nomina. anak (ngoko); yoga (krama madya); putra (krama inggil); keturunan nomer loro. Bu Andri punya anak dua. Bu Andri anduweni yoga kale.*

B – b

badan *nomina. awak (ngoko); badan (krama madya); salira (krama inggil); wujud sekabehane menungsa. Badan kakak tinggi dan kurus. Awake kakak dhuwur lan kuru.*

bangun *verba. tangi (ngoko); tangi (krama madya); wungu (krama inggil); ngadhek utawa lungga soko turu. Adik menangis ketika bangun tidur. Adik nangis wektu tangi turu.*

bangunkan *verba. gugah (ngoko); gigah (krama madya); wungu (krama inggil); nggarai wong liyo ngadhek utawa lungga soko turu. Aku membangunkan ibu karena ada tamu. Aku gugah ibu amerga ana dayo.*

beli *verba. tuku (ngoko); tumbas (krama madya); mundhut (krama inggil); olehi apa-apa amerga ditukar nganggo duwit. Aku beli baju baru ketika lebaran. Aku tuku klambi anyar wektu riyaya.*

beri *verba. aweh (ngoko); sukani (krama madya); paringi (krama inggil); nyerahake apa-apa kanggo wong liya. Ada teman yang suka memberi. Ana kanca sing seneng aweh-aweh.*

buruk *ajektiva. ala (ngoko); awon (krama madya); awon (krama inggil); ora kepenak didhelok. Kelakuan buruk tidak boeh dilakukan. Kelakuan ala ora oleh dilakoni.*

G – g

gigi *nomina. untu (ngoko); waos (krama madya); waja (krama inggil); balung cilik-cilik sajeroning cangkem. **Aku punya obat sakit gigi.** Aku dhuwe obat lara untu.*

H – h

hati *nomina. ati (ngoko); manah (krama madya); (peng)galih (krama inggil); organ ing jero awak warna abang ing dhuwur weteng. **Rahayu suka membicarakan hati.** Rahayu remen ngomongke ati.*

hidup *ajektiva. urip (ngoko); gesang (krama madya); sugeng (krama inggil); isih ana tur obah. **Kita hidup cuma sekali.** Awak dhewe urip mung sepisan.*

I – i

ikut *verba. anut (ngoko); tumut (krama madya); nderek (krama inggil); ngelakoni penggawean kaya sing dilakoni wong liya. **Aku ikut ibu pergi ke pasar.** Aku tumut ibuk kesah peken.*

jari

mulut

J – j

jari *nomina. dirji (ngoko); driji (krama madya); racikan (krama inggil); bagian lima ing pucuk tangan utawa sikil. **Jari kakakku baru sembuh dari gatal-gatal.** Driji kakakku nembe waras saka gatal-gatel.*

jauh *ajektiva. adoh (ngoko); tebih (krama madya); tebih (krama inggil); dowo jarake. **Rumah Adi jauh dari sini.** Omahe Adi adoh soko kene.*

jelek *ajektiva. ala (ngoko); awon (krama madya); awon (krama inggil); ora kepenak didhelok. **Kata Obi, baju ini tidak jelek.** Jare Obi, klambi iki ora ala.*

K – k

kumis *nomina. brengos (ngoko); brengos (krama madya); gumbala; rawis (krama inggil); rambut ing dhuwur cangkem. **Kumis Mas Bambang tebal sekali.** Brengose Mas Bambang kandhel banget.*

L – l

leher *nomina. gulu (ngoko); gulu (krama madya); jangga (krama inggil); bagian awak kanggo nyambung endas lan awak. Leherku panjang. Guluku dowo.*

lidah *nomina. ilat (ngoko); ilat (krama madya); lidhah (krama inggil); bagian awak ing jero cangkem agawe ngerasakake raos; agawe ngomong. Lidah untuk merasakan makanan yang enak. Ilat kanggo ngerasakake panganan ingkang echo.*

lihat *verba. tonton (ngoko); tingal (krama madya); priksa (krama inggil); agawe mata kanggo nambah kaweruhan. Baju Anda tidak terlihat. Klambi sampeyan mboten ketingal.*

ludah *nomina. idu (ngoko); idu (krama madya); kecoh (krama inggil); banyu metu soko lambe. Jangan mengeluarkan ludah di sembarang tempat. Aja ngetokake idu ing sembarang panggonan.*

lupa *verba. lali (ngoko); supe (krama madya); kalimengan (krama inggil); ucul soko keilingan. Jangan lupa dengan orang tuamu. Aja lali karo wong tuwamu.*

lutut *nomina. dhengkul (ngoko); dhengkul (krama madya); jengku (krama inggil); bagian awak wetara pupu lan kentol. Kata Bapak, Lutut itu untuk menggerakkan kaki. Jare Bapak, dhengkul iku agawe nggerakake sikil.*

M – m

mau *adverbia. gelem (ngoko); purun (krama madya); kersa (krama inggil); karep lan temen-temen seneng. Kamu mau makan ikan. Kowe gelem mangon iwak.*

mulut *nomina. cangkem (ngoko); cangkem (krama madya); tutuk (krama inggil); bagian awak agawe ngomong, mangan, ngombe. Mulut adikku warna merah. Cangkeme adikku warna abang.*

P – p

pergi *verba. lunga (ngoko); kesah (krama madya); tindak (krama inggil); obah nang panggonan liya. Jangan pergi jauh-jauh sendirian. Aja lunga adoh-adoh dewan.*

uang *nomina. geger (ngoko); geger (krama madya); pingkeran (krama inggil); bagian buri awak soko gulu nganti bokong. Kemarin punggung paman sudah sembuh. Wengi geger e paman wes waras.*

R – r

rambut *nomina. rambut (ngoko); rambut (krama madya); rikma (krama inggil); wulu ing sirae menungsa. Rambutmu warna hitam. Rambutmu warna ireng.*

T – t

tanya *verba. takon (ngoko); taken (krama madya); nyuwun pirsu (krama inggil); anjaluk katerangan utawa jawaban. Budi tanya kepada Pak Guru. Budi takon menyang Pak Guru.*

tua *ajektiva. tuwa (ngoko); sepuh (krama madya); yuswa (krama inggil); urip wes suwe. Paman Ona sudah tua sekarang. Paman Ona wes sepuh saiki.*

tulang *nomina. balung (ngoko); balung (krama madya); tosan (krama inggil); ragangan ing jero awak. Manusia bisa berdiri karena punya tulang. Menungsa isa ngadhek amarga dhuwe balung.*

uang *nomina. duwit (ngoko); yatra (krama madya); artha (krama inggil); alat resmi (kertas utawa koin) sing diwedalake pemerintah kanggo tuku lan tukar barang. Aku punya uang seratus ribu. Aku dhuwe duwit satus ewu.*

Dalam contoh kamus yang berisi 30 lema tersebut dapat dilihat bahwa setiap kata memiliki kosakata masing-masing sesuai dengan tingkatan bahasa Jawanya meskipun ada beberapa yang kosakatanya sama. Selain itu juga terdapat definisi yang bisa menjelaskan arti atau makna kata. Juga diberikan contoh kalimat dalam bahasa Indonesia sebagai bahasa sumber dan bahasa Jawa sebagai bahasa tujuan.

Jika sebuah kamus dapat membantu para penggunanya, juga dapat memberi penjelasan mengenai arti kata dan penggunaannya, maka kamus juga akan menjadi sumber belajar yang memadai untuk dimiliki dan juga untuk terus dikembangkan. Kamus juga merupakan hasil dokumentasi mengenai bahasa yang dimiliki oleh suatu masyarakat, terlebih kamus bahasa daerah sehingga dapat dikatakan pembuatan kamus sama dengan proses pelestarian suatu bahasa daerah. Penggunaan bahasa Jawa dengan tiga tingkatan (bahasa Jawa Ngoko, Krama Madya, dan Krama Inggil) merupakan warisan budaya berbahasa pada masyarakat Jawa.

IV. SIMPULAN

Dari 430 data yang sudah diidentifikasi berdasarkan tingkatan bahasa Jawanya, ternyata tidak semua kosakata yang memiliki tingkatan bahasa Jawa itu berbeda sesuai tingkatannya, tetapi ada beberapa kata yang sama. Misal kata ‘estri’ dalam bahasa Jawa Krama Madya sama dengan bahasa Jawa Krama Inggil yang artinya adalah ‘perempuan’. Faktanya, lebih banyak jumlah kosakata yang berbeda setiap tingkatan bahasa Jawanya. Pemberian tiga tingkatan bahasa Jawa dalam kamus bahasa Jawa-bahasa Indonesia atau sebaliknya dapat membantu pengguna kamus dan pengguna bahasa Jawa juga merupakan suatu sarana pembelajaran serta pelestarian bahasa daerah di Indonesia.

Daftar Pustaka

- Alwi, Hasan. 2007. *Kamus Besar Bahasa Indonesia*. Jakarta: Balai Pustaka.
- Blount, B. G. (ed). 1974. *Language Culture and Society*. Cambridge: Winthrop.
- http://www-01.sil.org/computing/catalog/show_software.asp?id=92 (diunduh tanggal 10 Juli 2016 pukul 13:00 WIB)
- Kridalaksana, Harimurti. 1985. *Tata bahasa deskriptif bahasa Indonesia: Sintaksis*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Moleong, L. J. 2007. *Metodologi Penelitian Kualitatif*. Bandung: PT Remaja Rosdakarya Offset.
- Ngadiman, Agustinus. 2006. *Tingkat Tutur Bahasa Jawa: Wujud Kesantunan Manusia Jawa (Dulu dan Sekarang)*. Situs Sutresna Jawa: <http://ki-demang.com/kbj5/index.php/makalah-kunci/1132-09-tingkat-tutur-bahasa-jawa-wujud-kesantunan-manusia-jawa>
- Nugroho, G. Setyo, dkk. (Tanpa Tahun). *Tatanan Anyar PINTER BASA JAWI PEPAK*. Surabaya: CV Pustaka Agung Harapan.
- Sariyan, Awang. 2014. *Teras Pendidikan Bahasa Melayu*. Selangor: PTS Malaysia.

Lema Budaya: Ketersebaran dan Keterwakilan Semantis

Ahmad Fadly

Universitas Muhammadiyah Jakarta

ahmadfadly2901@gmail.com

Abstrak

Kemajuan suatu bahasa tercermin dari kekayaan kosakata sehingga tidak terdapat celah atau kekosongan leksikal, baik pada bidang iptek maupun budaya. Dalam konteks Indonesia kekayaan itu diukur dari jumlah dan kelengkapan lema (entri) yang bersumber—salah satunya—dari kontribusi bahasa daerah mengingat bahwa bahasa-bahasa daerah di Indonesia sangatlah kaya akan budaya. Meskipun demikian, kenyataannya lema budaya KBBI masih didominasi oleh serapan dari kosakata dan ungkapan asing.

Dalam penyusunan KBBI, terutama terkait dengan lema budaya, perlu diambil beberapa langkah berikut. *Pertama*, penyusunan korpus bahasa daerah tidak hanya terpusat di pulau atau masyarakat Jawa, tetapi juga tersebar ke seluruh wilayah Indonesia yang memiliki bahasa daerah. *Kedua*, data berupa kosakata yang terkumpul dari berbagai bahasa daerah itu dikelompokkan berdasarkan unsur budaya (organisasi sosial, sistem mata pencaharian, sistem pengetahuan, teknologi, religi, kesenian, dan sistem bahasa). *Ketiga*, kosakata budaya yang telah dikelompokkan dari berbagai daerah itu diidentifikasi berdasarkan kesamaan makna, jika ditemukan dua atau lebih kosakata dari daerah yang berbeda, tetapi bereferen atau berkonsep sama, dipilih salah satunya yang paling populer. Keempat, jika kedua atau lebih kosakata itu memiliki tingkat kepopuleran yang sama, dipilih yang paling tepat berdasarkan sistem fonologis bahasa Indonesia. *Terakhir*, masukkan istilah budaya yang belum tercakup dalam KBBI sebagai entri baru.

Oleh karena itu, dalam penyusunan KBBI perlu dikedepankan aspek ketersebaran dari berbagai bahasa daerah dan keterwakilan semantis. Langkah itu dapat memperkaya kosakata bahasa Indonesia sekaligus menjadi cerminan budaya Indonesia yang heterogen.

Kata Kunci: lema budaya, ketersebaran, keterwakilan semantis

I. PENDAHULUAN

1.1. Latar Belakang

Indonesia mempunyai potensi jumlah bahasa daerah yang sangat besar. Penelitian Summer Institute of Linguistics (SIL) pada 2001 menyatakan bahwa bahasa daerah di seluruh Indonesia berjumlah 726 bahasa (Montolalu, et al. 2005). Sementara itu, jumlah bahasa daerah yang ada di Indonesia dan Papua Nugini sebesar 25% dari seluruh bahasa di dunia (Crystal 2003). Artinya, secara kuantitatif bahasa daerah di Indonesia mendominasi pada tataran global.

Kenyataan itu mencerminkan kekayaan dan keanekaragaman budaya daerah yang patut disyukuri. Rosidi pernah menyatakan bahwa bahasa dan kebudayaan daerah merupakan suatu realitas yang menjadi modal untuk meningkatkan derajat bangsa (Rosidi 1995). Meskipun demikian, bahasa daerah—juga budayanya—itu dapat punah. Kepunahan bahasa daerah itu dinyatakan oleh Tondo (2009) disebabkan oleh sepuluh faktor, yaitu faktor ekonomi, migrasi, perkawinan antaretnik, bencana alam dan musibah, kondisi masyarakat yang bilingual atau multilingual, globalisasi, pengaruh bahasa mayoritas, kurangnya penghargaan terhadap bahasa etnik sendiri, kurangnya intensitas komunikasi berbahasa daerah, dan dimensi sosial politik bahasa Indonesia. Kesepuluh penyebab kepunahan bahasa daerah itu sebenarnya dapat diantisipasi dengan penguatan bahasa daerah melalui pengintegrasian kosakata budaya daerah ke dalam Kamus Besar Bahasa Indonesia (KBBI).

Pengintegrasian tersebut tidak hanya berasal dari bahasa daerah yang memiliki jumlah penutur besar, tetapi juga dari bahasa daerah yang memiliki jumlah penutur terbatas. Dengan kata lain, penyusun KBBI perlu mempertimbangkan aspirasi budaya tiap-tiap masyarakat di Indonesia, tidak sebatas melihat budaya masyarakat mayoritas yang berada di kota besar.

Dengan mempertimbangkan budaya daerah dari seluruh masyarakat itu, KBBI akan memiliki kekayaan lema sekaligus mencerminkan kekayaan kebudayaan Indonesia yang termanifestasi dalam lema-lemanya.

1.2. Masalah

Sebagai dokumentasi bahasa, KBBI diharapkan memayungi bahasa-bahasa daerah di Indonesia. Namun kenyataannya, lema (entri) yang bersumber dari bahasa daerah pada KBBI Edisi IV pun justru sangat minim. Jumlah lema yang berasal dari kata dan ungkapan bahasa daerah pada kamus itu hanyalah 86 lema. Jumlah itu jauh lebih sedikit jika dibandingkan lema yang berasal dari kata dan ungkapan asing yang sebanyak 2.204 lema (Macaryus 2010). Ironis.

1.3. Tujuan

Pada hakikatnya, pemasukan lema yang berasal dari kata dan ungkapan daerah memberikan ruang gerak bagi budaya daerah untuk berkembang di Indonesia. Kosakata daerah yang dijadikan lema ke dalam KBBI mampu memperkenalkan budaya daerahnya kepada masyarakat Indonesia melalui pengguna KBBI sehingga nantinya kosakata bahasa Indonesia mampu mencerminkan keragaman kebudayaan Indonesia secara holistik.

Pada praktiknya tentu tidak semua kosakata daerah dijadikan lema KBBI. Kosakata daerah yang berkategori budayalah yang dijadikan lema KBBI karena kosakata itu berpeluang untuk dapat berkontribusi dalam memperkaya khazanah kebudayaan Indonesia.

1.4. Tinjauan Pustaka

Berbagai kritik, masukan, bahkan penelitian terkait dengan KBBI Edisi IV pernah dilakukan. Macaryus (2009) telah melontarkan kritikan sekaligus memberikan masukan yang sangat berharga terkait dengan KBBI Edisi IV. Berbagai kritikan itu mencakup ketidaktaatan dalam proses morfologis, inkonsistensi penentuan lema dan sublema, kerancuan bentuk baku dan tidak baku, ketidakkonsistenan dalam menyerap kata, serta ketidakakuratan data kabupaten/kota di Indonesia dan Hari Besar Nasional. Sementara itu, masukan yang ia berikan terkait dengan pendefinisian lema. Ia memandang bahwa penggunaan definisi dengan negator *tidak* sebaiknya dihindari, tetapi diletakkan pada alternatif terakhir setelah definisi yang lainnya. Selain itu, ia juga memberikan masukan yang berhubungan dengan pengembangan ungkapan dan peribahasa melalui pemanfaatan fenomena-fenomena kontekstual dan aktual. Cara itu dilakukan agar generasi muda memanfaatkan peribahasa sebagai dasar berpikir, bersikap, dan bertindak, sekaligus menjadi kontrol yang mengarahkan pada terbentuknya perilaku dan budaya tertentu.

Adapun penelitian yang berhubungan dengan KBBI edisi IV pernah dilakukan oleh Supriyanti (2012). Hasil penelitiannya menunjukkan adanya pengaruh stereotip dalam penyusunan definisi (termasuk contoh penggunaan lema) serta pada kemunculan lema dan sublema. Dalam penelitiannya juga ditemukan jenis oposisi gender dan pola pendefinisian lema yang beroposisi gender itu.

Pada makalah ini, penulis memandang bahwa ancaman kepunahan bahasa daerah perlu direspons secara serius karena punahnya bahasa daerah berarti berkurang pula kekayaan kebudayaan Indonesia. Salah satu upaya mengatasinya adalah dengan pengentrian sekaligus pengintegrasian kosakata budaya bahasa daerah itu ke dalam lema KBBI yang bermuara pada pengayaan kosakata bahasa Indonesia.

II. KERANGKA TEORI DAN METODE PENELITIAN

Idealnya, sebuah kamus besar memiliki lema yang lengkap. Lema itu berupa morfem dasar, baik yang bebas, seperti *batu*, *pergi*, dan *pulang*, maupun yang terikat, seperti

juang, henti, dan abai (Chaer 2007). Pada kamus ekabahasa, kelengkapan tersebut menggambarkan kekayaan kosakata bahasanya. Dalam konteks Indonesia, kamus besar yang standar adalah KBBI. KBBI memiliki lema yang lebih lengkap dibandingkan dengan kamus-kamus ekabahasa lain di Indonesia sehingga dapat dijadikan acuan dalam penyusunan kamus-kamus lain yang sifatnya terbatas. Sebagai perbandingan, KBBI edisi terakhir (Edisi IV) memuat lebih dari 90.000 lema, sedangkan Kamus Umum Bahasa Indonesia (KUBI) karya W.J.S. Poerwadarminta hanya memuat sekitar 22.000 lema (Sulaiman 2009).

Keunggulan KBBI dari segi jumlah lema tidak terlepas dari banyaknya ungkapan dan kosakata asing yang diserap di dalamnya. Memang Johnson, sebagaimana dikutip oleh Chaer (2007) pernah menyatakan bahwa fungsi sebuah kamus adalah untuk menjaga kemurnian bahasa. Namun, anggapan preskriptif itu tidak lagi sesuai dengan realitas kini dengan intensitas komunikasi yang tinggi antarmasyarakat tutur. Saat ini penyerapan kata atau kelompok kata dari bahasa asing merupakan suatu keniscayaan. Demikian pula penyerapan dari bahasa daerah ke dalam lema, terutama terkait dengan budaya.

2.1 Lemma Budaya

Dengan hanya berjumlah 86 lema yang berasal dari kosakata dan ungkapan daerah, sebagaimana telah disebutkan oleh penulis pada bagian awal, KBBI hanya menunjukkan kebudayaan Indonesia secara parsial. Untuk itu, pengentrian kosakata budaya daerah mutlak dilakukan. Dalam pengentriannya, penyusun perlu mengelompokkan kata-kata berkategori budaya yang akan dijadikan lema. Terkait dengan itu, beberapa ahli mempunyai pendapat yang beragam ikhwal kategori budaya. Kroeber dan Kluckhohn, sebagaimana dikutip oleh Georgas (2003) menyatakan bahwa

Culture consists of patterns, explicit and implicit and for behavior acquired and transmitted by symbols, constituting the distinctive achievements of human groups, including their embodiments in artefacts: the essential core of culture consists of traditional (i. e. historically derived and selected) ideas and especially their attached values; cultural system may on the one hand be considered as products of action, on the other as conditioning elements of further action”.

Kategori budaya itu bersifat umum sehingga belum aplikatif. Sementara itu, Newmark mengelompokkan kategori budaya menjadi lima, yaitu budaya ekologis, budaya material, budaya sosial, organisasi sosial, dan kiasan serta tradisi (Newmark 1988). Budaya ekologis mencakup flora, fauna, pegunungan, dataran, dsb.), sedangkan budaya material mencakup artefak, makanan, pakaian, rumah, kendaraan, dan alat transportasi serta komunikasi. Sementara itu, budaya sosial mencakup pekerjaan dan kesenangan. Berbeda dengan budaya sosial, organisasi sosial mencakup adat istiadat, kegiatan adat, prosedur, konsep, agama, dan estetika. Adapun kiasan merupakan bahasa tubuh yang mengandung makna dan tradisi merupakan kebiasaan khas yang sering dilakukan oleh masyarakat.

Senada dengan Newmark, Hoed menyatakan bahwa dilihat dari segi perwujudannya terdapat tujuh unsur budaya yakni organisasi sosial, sistem mata pencaharian, sistem pengetahuan, teknologi, religi, kesenian, dan sistem bahasa (Hoed 2006). Batasan Hoed itu lebih komprehensif sekaligus aplikatif sehingga dapat dijadikan acuan dalam penyusunan KBBI, terutama dalam pemasukan kata berkategori budaya daerah.

2.2 Metode Penelitian

Pada penelitian ini digunakan metode deskriptif kualitatif. Artinya, penulis menggambarkan kondisi KBBI Edisi IV dari segi lema dan sublema yang terkandung di dalamnya dari berbagai ulasan mengenai kamus itu. Selain itu, penulis menawarkan tahap penyusunan KBBI yang bersumber dari kosakata budaya daerah.

III. ANALISIS PENELITIAN

Kamus berperan penting dalam berbagai aspek kehidupan. Untuk itu, penyusun kamus dituntut memiliki wawasan yang luas, integritas yang tinggi, dan ketelitian. Ketelitian itu tercermin dari ketaatasannya dalam mengikuti tahap penyusunan kamus.

3.1. Tahap Penyusunan Kamus

Beberapa pakar menawarkan tahap penyusunan kamus. Secara umum, Ginzburg, et al. (1979) menyatakan bahwa tugas penyusun kamus mencakup tiga tahap, yaitu mengumpulkan data, menyeleksi dan menetapkan lema, dan menyusun tiap-tiap lema. Secara khusus, Chaer (2007) mengusulkan prosedur penyusunan kamus ke dalam sepuluh tahap, yaitu persiapan, pengumpulan data, pengabjadan, pengolahan data, penyusunan kartu naskah, pengetikan naskah, koreksi naskah, cetak coba, koreksi cetak coba, dan reproduksi naskah.

Pada tahap persiapan, penyusun menentukan tujuan kamus yang akan disusun: kamus ekabahasa, dwibahasa, atau aneka bahasa beserta sasaran dan ukuran kamus tersebut. Pada tahap ini, penyusun menyiapkan alat tulis yang diperlukan, menyediakan atau mempersiapkan korpus data, menyediakan sumber rujukan, dan melakukan studi kelayakan.

Pada tahap pengumpulan data, penyusun menuliskan setiap data, nama pengumpul data, dan sumbernya dalam satu kartu. Kemudian, pada tahap pengabjadan kartu data, penyusun menempatkan setiap data berdasarkan urutan abjad. Pengabjadan itu dilakukan secara vertikal dan horisontal. Secara vertikal artinya menempatkan kata dasar yang berawalan huruf *a* di atas kata yang berawalan huruf *b*, menempatkan kata yang berawalan huruf *b* di atas kata yang berawalan huruf *c*, dan seterusnya. Sementara itu, secara horisontal berarti jika ditemukan dua buah kata dasar yang memiliki huruf pertama yang sama, huruf kedua yang diurutkan berdasarkan abjad, demikian pula seterusnya.

Pada tahap berikutnya, penyusun mengolah data. Pada tahap ini penyusun melakukan empat pekerjaan sekaligus, yaitu menyeleksi ulang urutan abjad data, menandai suku kata, memberi makna, dan memberi label-label penjelas. Tahap selanjutnya, penyusun mulai menyusun kartu naskah. Pada tahap ini, penyusun memeriksa kartu-kartu yang telah diolah, lalu mengelompokkan lema dan sublema.

Adapun pada tahap pengetikan naskah, penyusun mengetik naskah kamus berdasarkan kartu naskah yang telah disusun secara alfabetis, diberi makna definisi, dan keterangan-keterangan lain. Setelah itu, naskah dikoreksi kata per kata, sekaligus disusun naskah kata pengantar yang berisi pedoman penggunaan kamus serta berbagai petunjuk terkait dengan kamus tersebut, lalu dicetak coba. Kemudian, cetakan itu dikoreksi untuk mendapatkan hasil yang sempurna. Setelah berbagai tahap tersebut dilalui, penyusun memproduksi kamus untuk dipublikasikan dan dipasarkan.

Tahapan yang ditawarkan oleh Ginzburg, et al. dan Chaer di atas bersifat normatif dan berlaku untuk penyusunan kamus secara umum. Oleh karena itu, penulis mengusulkan tahap penyusunan KBBI yang terkait dengan budaya dan dilandasi oleh semangat pelestarian bahasa-bahasa dan budaya daerah di Indonesia

3.2 Tahap Penyusunan Lema Budaya: Ketersebaran dan Keterwakilan Semantis

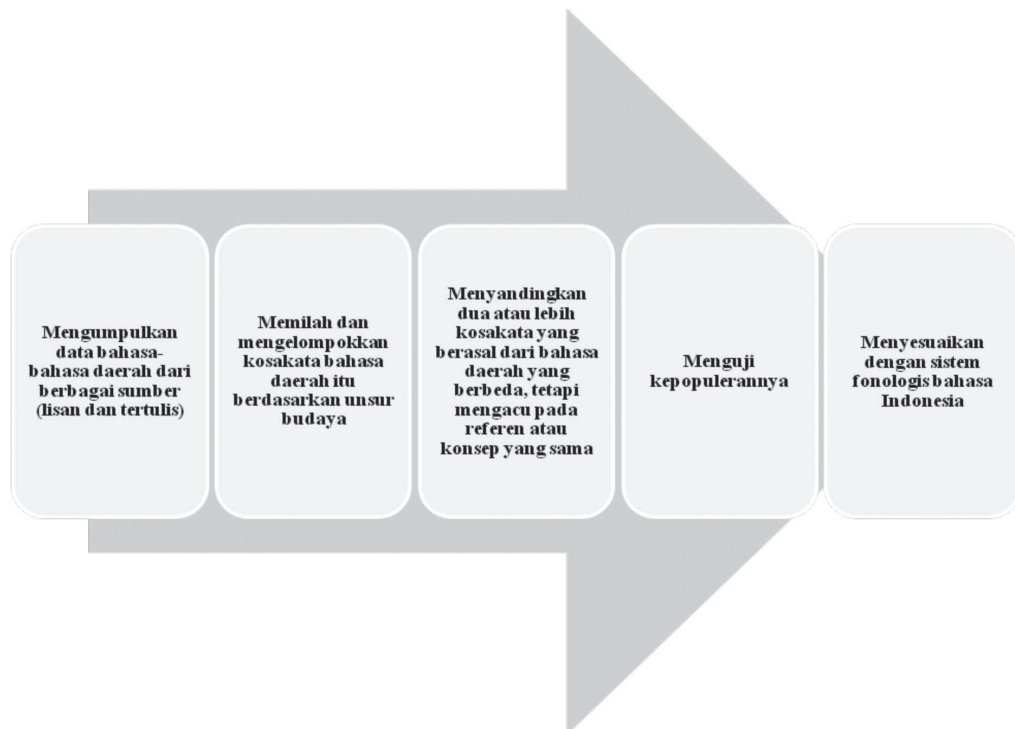
Sesuai dengan semangat Sumpah Pemuda yang meletakkan kedudukan bahasa Indonesia sebagai bahasa nasional sekaligus pemersatu, maka dalam penyusunan KBBI juga perlu dikedepankan prinsip kesetaraan sehingga kosakata bahasa Indonesia tidak hanya didominasi oleh bahasa daerah tertentu, tetapi juga diberikan peluang ketersebaran geografis. Untuk itu, perlu diletakkan kedudukan setiap bahasa daerah secara setara.

Dalam penentuan korpus, setiap bahasa daerah diberikan kemungkinan yang sama sebagai bakal lema KBBI. Meskipun demikian, tidak berarti semua kata budaya dari seluruh bahasa daerah dimasukkan ke dalam lema. Jika ditemukan dua kata budaya atau lebih dari bahasa daerah yang berbeda, tetapi mengacu pada satu referen yang sama, cukup dipilih salah satunya. Pemilihan itu didasarkan pada dua hal, yaitu keterwakilan semantis dan kesamaan fonologis dengan bahasa Indonesia. Keterwakilan semantis maksudnya dari dua atau lebih kosakata budaya itu dipilih salah satunya yang lebih representatif atau populer (tolok ukurnya adalah yang paling banyak digunakan di berbagai kesempatan: bahasa lisan ataupun tertulis).

Jika dua kata atau lebih itu sama-sama populer, pilihlah yang sesuai dengan kaidah fonologis bahasa Indonesia. Sebagai contoh: kata *peyeum* (bahasa Sunda) dan *tape* (bahasa Jawa) mengacu pada referen yang sama. Untuk menentukannya ke dalam lema KBBI, dipilih yang lebih populer. Namun jika keduanya sama populer, dipilih yang sesuai dengan kaidah fonologis bahasa Indonesia. Kata *peyeum* tidak sesuai dengan kaidah fonologis bahasa Indonesia karena dalam sistem bunyi bahasa Indonesia tidak dikenal diftong *eu*. Sementara itu, kata *tape* lebih sesuai dengan sistem fonologis bahasa Indonesia karena semua fonem /t/, /a/, /p/, dan /e/ dikenal dalam sistem bunyi bahasa

Indonesia. Meskipun demikian, pengentrian kosakata berkategori budaya yang berasal dari bahasa daerah tetap harus disertai dengan pelabelan seperti *Jw* (Jawa), *Sd* (Sunda), *Mn* (Minang), dsb., yang menunjukkan bahwa lema tersebut berasal dari bahasa daerah.

Dalam konteks pengentrian kosakata budaya daerah ke dalam KBBI, penulis menawarkan tahap penyusunnya yang digambarkan dalam diagram alur berikut.



Gambar 1. Tahap Pengentrian Kosakata Budaya ke dalam KBBI

Pada tahap awal, penyusun KBBI perlu mengumpulkan data berbagai bahasa daerah dari berbagai sumber: lisan dan tulisan. Data kebahasaan dari sumber lisan digunakan mengingat bahwa tidak semua masyarakat daerah di Indonesia memiliki budaya literasi yang baik sehingga bahasanya tidak terdokumentasi secara tulis. Sementara itu, sumber data dari tulisan ditujukan untuk bahasa daerah yang terdokumentasi secara tulis. Data kebahasaan yang dikumpulkan itu (lisan dan tulisan) dibatasi tidak lebih dari satu tahun saat pengumpulan data.

Tahap selanjutnya adalah pemilahan dan pengelompokkan kosakata berbagai bahasa daerah itu berdasarkan unsur budayanya. Pemilahan dan pengelompokkan itu mengacu pada ketujuh unsur budaya, sebagaimana yang ditawarkan oleh Hoed (2006), yakni organisasi sosial, sistem mata pencaharian, sistem pengetahuan, teknologi, religi, kesenian, dan sistem bahasa.

Tahap berikutnya adalah penyandingan dua atau lebih kosakata yang berasal dari bahasa daerah yang berbeda, tetapi mengacu pada referensi atau konsep yang sama. Cara itu dilakukan untuk memudahkan dalam memilih salah satu kosakata yang tepat untuk dijadikan lema agar nantinya tidak terjadi tumpang tindih dan kerancuan antarlema yang berakibat pada tidak efektif dan efisiennya KBBI.

Tahap selanjutnya adalah pengujian kepupuleran. Jika ditemukan dua atau lebih kosakata dari bahasa daerah yang berbeda, penyusun perlu memilih salah satunya berdasarkan kepopulerannya. Untuk itu, penyusun perlu mentransliterasi sumber kebahasaan dari bahasa lisan untuk memudahkan dalam membandingkan antarkosakata yang bereferen atau berkonsep tunggal itu.

Tahap terakhir yaitu penyesuaian dengan sistem fonologis bahasa Indonesia. Langkah ini diambil jika ditemukan bahwa terdapat dua atau lebih kosakata yang bereferen atau berkonsep tunggal dan memiliki tingkat kepopuleran yang sama. Dalam memilih salah satu kosakata untuk dijadikan lema perlu didasarkan pada kesesuaian dengan sistem fonologis bahasa Indonesia.

IV. SIMPULAN

KBBI termasuk kamus yang standar di Indonesia, tetapi masih didominasi oleh serapan bahasa asing. Kenyataan itu menunjukkan bahwa kosakata bahasa Indonesia justru tidak mencerminkan kebudayaan Indonesia. Untuk itu, perlu diambil langkah-langkah yang mengakomodasi kosakata daerah dengan memasukkannya ke dalam lema KBBI.

Pengentrian kosakata daerah ke dalam KBBI lebih diutamakan yang berkategori budaya mengingat bahwa budaya daerah di Indonesia sangat kaya. Dalam pengentriannya, diklasifikasikan terlebih dahulu kosakata berkategori budaya, yaitu organisasi sosial, sistem mata pencaharian, sistem pengetahuan, teknologi, religi, kesenian, dan sistem bahasa. Setelah diklasifikasikan berdasarkan kategori budaya, perlu dikedepankan aspek ketersebaran dan keterwakilan semantis.

Daftar Pustaka

- Chaer, Abdul. 2007. *Leksikologi dan Leksikografi Indonesia*. Jakarta: PT Rineka Cipta.
- Chaer, Abdul. 2013. *Fonologi Bahasa Indonesia*. Jakarta: PT Rineka Cipta.
- Crystal, David. 2003. *Language Death*. New York: Cambridge University Press.
- Georgas, James. 2003. "Cross-Cultural Psychology, Intelligence, and Cognitive Processes" dalam *Culture and Children's Intelligence: Cross-Cultural Analysis of WISC III*. California: Elsevier Science.
- Ginzburg, et al.. 1979. *A Course in Modern English Lexicology (Second Edition: Revised and Enlarged)*. Moskow: Vysšaja Škola.
- Hoed, Benny Hoedoro. 2006. *Penerjemahan dan Kebudayaan*. Jakarta: PT Dunia Pustaka Jaya.
- Macaryus, Sudartomo. 2010. "Kamus Besar Bahasa Indonesia Edisi IV: Beberapa Problem dan Kemungkinan Mengatasinya" dalam *Peneroka Hakikat Bahasa*. Yogyakarta: Universitas Sanata Dharma.
- Montolalu, Lucy Ruth, et al.. 2005. "Tipologi Bahasa dan Bahasa-Bahasa di Dunia" dalam *Pesona Bahasa: Langkah Awal Memahami Linguistik*. Jakarta: PT Gramedia Pustaka Utama.
- Newmark, Peter. 1988. *A Textbook of Translation*. New York: Prentice Hall.
- Parera, Jos Daniel. 1986. *Studi Linguistik Umum dan Historis Bandingan*. Jakarta: Erlangga.
- Rosidi, Ajip. 1995. *Sastra dan Budaya: Kedaerahan dalam Keindonesiaan*. Jakarta: PT Dunia Pustaka Jaya.
- Supriyanti, Nanik. *Praktik Leksikografi atas Nomina Persona Berorientasi Gender dalam Kamus Besar Bahasa Indonesia Edisi IV*. Tesis pada Program Pascasarjana Program Studi Linguistik Fakultas Ilmu Pengetahuan Budaya Universitas Indonesia. Depok: Tidak Diterbitkan.
- Tondo, Fanny Henry. 2009. "Kepunahan Bahasa-Bahasa Daerah: Faktor Penyebab dan Implikasi Etnolinguistik" dalam *Jurnal Masyarakat dan Budaya* Vol. 11 No. 2 Tahun 2009. Jakarta: P2KK-LIPI.
- Wierzbicka, Anna. 1997. *Understanding Cultures Through Their Key Words: English, Russian, Polish, German, and Japanese*. New York: Oxford University Press.

Mekanisme Pengolahan dan Penyajian Kosakata Budaya dalam Kamus Pemelajar BIPA

Atin Fitriana, Dien Rovita, Totok Suhardijanto

Universitas Indonesia

Abstrak

Kertas kerja ini menyajikan penanganan kosakata budaya dalam usaha penyusunan kamus BIPA yang hingga kini masih terus dilakukan di lingkungan Universitas Indonesia. Penjaringan data kosakata budaya dalam penelitian leksikografi ini dilakukan dengan menggunakan korpus BIPA yang disusun oleh Program BIPA LBI UI. Korpus elektronik BIPA yang berjumlah 1,5 juta kata tersusun dari kumpulan teks bahan ajar, percakapan guru dan siswa, soal tes, serta hasil karya pemelajar BIPA. Metode yang digunakan dalam penelitian ini merupakan gabungan dari metode dalam korpus linguistik, dalam hal penentuan dan pengolahan kosakatanya, dan metode leksikografi dalam hal penyajian entrinya.

Pada makalah ini, fokus perhatian diberikan pada bagaimana penyusunan dan pendefinisian kosakata budaya dalam kamus pemelajar BIPA. Kriteria penentuan kosakata budaya dalam makalah ini didasarkan pada pemakaian kata sesuai dengan konteksnya yang ada di dalam korpus BIPA. Penggunaan kosakata budaya sebagai bagian dari penggunaan bahasa dapat memperlihatkan ciri sebuah budaya masyarakat tertentu. Sebuah kata dapat dikatakan sebagai kosakata budaya ketika kata tersebut menjadi khas dan menunjukkan karakteristik pola kehidupan sebuah kelompok masyarakat tertentu. Tentu saja, penyajian kosakata tersebut dalam kamus BIPA tidak cukup hanya dengan pemberian definisi umum dan sinonimi, tetapi juga dengan pemberian contoh, gambar, dan deskripsi lain yang dapat membantu pemelajar BIPA memahaminya dengan lebih baik.

Kata kunci: kosakata budaya, korpus BIPA, pengolahan, penyusunan, kamus pemelajar BIPA

I. PENDAHULUAN

1. 1Latar Belakang

Kamus adalah istilah umum untuk bahan rujukan yang memuat perbendaharaan kata suatu bahasa yang disusun menurut tradisi abjad bahasa yang bersangkutan. Istilah ini kemudian meluas penggunaannya untuk menunjukkan bahan rujukan yang terbatas atas lebih teknis, misalnya kamus istilah, kamus singkatan, kamus kata-kata berirama, dsb, serta meluas menjadi kamus ensiklopedis, juga kamus monolingual bi- atau multilingual (Kridalaksana, 2003: xi).

Kamus BIPA merupakan rujukan berisi perbendaharaan kosakata dalam bahasa Indonesia berikut makna yang terkandung di dalamnya, khususnya kosakata bahasa Indonesia yang digunakan oleh pemelajar BIPA dalam pembelajaran bahasa Indonesia serta dalam kehidupan sehari-hari. Berdasarkan hal tersebut, kamus BIPA harus mengandung kosakata yang tepat, berkaitan dengan kebutuhan para pemelajar BIPA. Kosakata tersebut dapat berupa kosakata dasar dan kosakata budaya. Kosakata budaya yang menjadi pokok pembicaraan dalam makalah ini dibatasi sebagai kosakata yang dalam penjelasannya tidak dapat dilakukan dengan sekadar menggunakan definisi umum dan/atau sinonimi, tanpa memasukkan unsur-unsur konteks budaya setempat. Misalnya, kata *sekatén* yang dalam KBBI Edisi IV (2008) dijelaskan dengan “/sekatén/ n pasar malam (terutama di Yogyakarta dan Surakarta) yg diadakan tiap bulan Maulud (untuk merayakan Maulid Nabi Muhammad saw.).” Bagi pemelajar BIPA, tentu saja, hal ini tidak cukup membantu.

Secara garis besar, terdapat beberapa tantangan dalam penyusunan kamus BIPA. Beberapa tantangan tersebut antara lain (1) seberapa besar jumlah kosakata dalam korpus yang akan dimasukkan ke dalam kamus, (2) bagaimana penentuan kosakata yang akan dimasukkan dan tidak akan dimasukkan ke dalam kamus, (3) bagaimana susunan lema dan sublema kamus, (4) bagaimana pendefinisian kosakata yang tepat untuk pengguna kamus. Terkait dengan tantangan terakhir, kamus yang disusun akan digunakan untuk pemelajar BIPA mulai dari tingkat dasar hingga mahir, sehingga yang termasuk di dalam persoalan tersebut adalah bagaimana pendefinisian atau pemberian informasi berkaitan dengan kosakata budaya, ungkapan, dan peribahasa.

1.2 Masalah Penelitian

Saat ini pengajaran bahasa Indonesia sebagai bahasa asing cukup berkembang. Oleh karena itu kebutuhan terhadap materi ajar sangat besar. Kamus, sebagai sebuah buku yang berisi penjelasan mengenai kosakata dalam sebuah bahasa, juga sangat diperlukan dalam kegiatan pembelajaran. Oleh karena itu, kami merasa bahwa penyusunan kamus BIPA juga sangat diperlukan.

Penelitian ini merupakan salah satu kegiatan yang dilakukan dalam usaha penyusunan kamus yang masih terus dilakukan. Dalam penelitian ini, kami berfokus pada kosakata budaya yang akan dimasukkan dalam kamus BIPA, khususnya bagaimana pengolahan kosakata budaya yang diambil dari korpus BIPA yang telah dikumpulkan sebelumnya, berdasarkan kriteria tertentu, dan bagaimana penyajian kosakata yang telah ditentukan berdasarkan kriteria tertentu tersebut di dalam kamus, khususnya dengan melihat pendefinisian katanya.

1.3 Tujuan Penelitian

Berdasarkan perumusan masalah di atas, makalah ini bertujuan untuk memberikan eksplanasi mengenai mekanisme pengolahan kosakata budaya berdasarkan korpus BIPA serta penyajian kosakata budaya tersebut di dalam kamus, khususnya penyajian definisi kosakata budaya tersebut di dalam kamus.

1.4 Tinjauan Pustaka

1.4.1 Penelitian Lexsikografi

Dora Amalia (2014) dalam disertasinya membahas pendefinisian dan model entri verba dalam kamus pemelajar bahasa Indonesia. Dalam disertasi yang berjudul “Formulasi Pendefinisian dan Model Pengentrian Verba dalam Kamus Pemelajar Bahasa Indonesia”, Amalia (2014:2—3) menjelaskan bahwa dalam pengajaran BIPA penggunaan kamus tidak terintegrasi dengan baik dan tidak mendapat prioritas dalam pelajaran BIPA. Selain itu Amalia juga menambahkan bahwa kamus yang biasa digunakan oleh pemelajar BIPA adalah kamus dwibahasa yang biasanya berbentuk kamus elektronik.

Dalam penyusunan sebuah kamus, bagian mikrostruktur kamus yang paling penting adalah entri, dan dalam menyusun entri pada kamus ekabahasa, definisi merupakan bagian yang terpenting (Amalia, 2014:6-7). Oleh karena itu tujuan dari penelitian ini adalah untuk membuat formulasi pendefinisian dan membuat model pengentrian verba dalam kamus untuk pemelajar bahasa Indonesia.

Penelitian ini menggunakan empat landasan teoretis. Teori yang pertama adalah teori modern fungsi leksikografis. Teori tersebut digunakan dalam perancangan dan penyusunan kamus dan pentingnya pengguna kamus dalam penyusunan kamus. Teori kedua yaitu teori leksikografi pedagogis. Teori ketiga yaitu teori semantik rangka. Teori keempat adalah teori tentang kolokasi yang digunakan dalam memilih dan menentukan informasi leksikografis. Metode yang digunakan dalam penelitian ini adalah metode campuran yaitu gabungan antara metode kuantitatif dan kualitatif.

Di dalam penelitian ini, peneliti juga melakukan evaluasi kamus pemelajar bahasa Indonesia. Evaluasi kamus pemelajar tersebut, yaitu kamus pemelajar *A Learner's Comprehensive Dictionary of Indonesia* (LCDI) dan *The Learner's Dictionary of Today's Indonesian* (LDTI). Kesimpulan yang diperoleh dari hasil evaluasi yang dilakukan

terhadap kedua kamus tersebut, yaitu LCDI lebih unggul dalam hal jumlah dan jenis lema. Selain itu, LCDI lebih memudahkan pengguna karena untuk tiap abjad menggunakan halaman baru. Sementara LDTI unggul pada aspek kebergunaan (dalam penyusunan entri, terdapat banyak penjelasan tanpa menggunakan banyak tanda atau kode). LDTI menjadikan bentuk derivasi sebagai lema tersendiri, sehingga memudahkan pengguna dalam mencari kata tanpa harus mencari kata dasarnya terlebih dahulu. Selain itu, LDTI juga menggunakan definisi yang lebih kontekstual dibandingkan LCDI.

Pada identifikasi karakteristik pengguna kamus pemelajar bahasa Indonesia, penelitian dilakukan dengan cara menetapkan profil pengguna dan riset kamus. Berdasarkan hasil wawancara, terdapat profil pengguna, yaitu tipe pengguna (dewasa, pemelajar bahasa, tingkat kemahiran menengah sampai mahir, dan kamus untuk keperluan belajar), tipe penggunaan (untuk keperluan acuan dan belajar bahasa), dan tingkat keterampilan pengguna (mempunyai pengetahuan kebahasaan yang memadai).

Berdasarkan penelitian yang telah dilakukan, yaitu berdasarkan hasil wawancara, terdapat dua jenis definisi yang paling banyak dipilih oleh partisipan dalam konteks reseptif, yaitu definisi analitis dan definisi kontekstual. Definisi analitis adalah definisi yang lazim dalam kamus, namun umumnya partisipan lebih menyukai jenis definisi yang berisi informasi sintaksis dalam bentuk pola valensi. Terdapat dua hal yang berkaitan dengan formulasi pendefinisian, yaitu pemilihan lema dan kosakata pendefinisian. Dalam penelitian yang dilakukan, pemilihan lema dan kosakata pendefinisian dilakukan menggunakan perhitungan frekuensi dan distribusi pemakaian kata di dalam korpus. Semakin tinggi frekuensi penggunaan kata, maka semakin penting kata tersebut dijadikan sebagai lema atau semakin besar kata tersebut dipahami dan layak digunakan sebagai kosakata pendefinisian.

Dalam penelitian ini, Amalia (2014) mengusulkan definisi singkat, yaitu berupa genus atau sinonim. Selain itu, untuk mempermudah mendapatkan informasi, penelitian ini mengusulkan untuk menuliskan definisi singkat dengan jenis huruf yang berbeda, yaitu menggunakan huruf kapital bercetak tebal yang ukurannya lebih kecil satu ukuran dari jenis huruf dalam definisi. Definisi singkat dapat berupa satu atau dua kata, diberi nomor agar memudahkan pencarian, dan disusun berurutan ke bawah sesuai makna yang paling banyak ditemukan di dalam korpus. Sementara itu, peneliti mengusulkan definisi kontekstual sebagai formulasi pendefinisian untuk konteks produk. Definisi kontekstual tersebut disajikan dengan pola **jika... berarti** dalam kalimat lengkap. Selain itu, definisi kontekstual juga dilengkapi dengan valensi. Adapun pola entri yang disajikan dalam penelitian ini, yaitu [**lema/sublema**] [/lafal/] [*kelas kata*] [*label*] [*runs on*] [**nomor polisem**] [**DEFINISI SINGKAT**] [definisi lengkap] [(**pola valensi**)] [*contoh kalimat*].

Berdasarkan penelitian yang dilakukan, model pengentrian verba dalam kamus harus meliputi definisi singkat dan ilustrasi. Untuk fungsi produktif, pengentrian verba didukung dengan informasi tentang lafal, informasi sintaksis (kelas kata dan pola valensi), contoh kalimat, kolokasi, dan catatan penggunaan. Adapun pola entri

verba dalam kamus tersebut direalisasikan dalam enam kategori verba, yang meliputi verba transitif berobjek, verba transitif berobjek dan berpelengkap, verba semitransitif, verba intransitif tak berpelengkap, verba intransitif berpelengkap, dan verba intransitif berpelengkap nomina dengan preposisi tetap.

1.4.2 Penelitian Leksikografi dan Korpus

Penelitian leksikografi dengan metode korpus dilakukan oleh David Lindemann (2013). Dalam artikel yang berjudul “*Bilingual Lexigraphy and Corpus Methods. The Example of German-Basque as Language Pair*”, Lindemann (2013) melakukan penelitian leksikografi berbasis korpus dalam bahasa Jerman-Basque dan mendeskripsikan rancangan penelitian untuk kamus elektronik Jerman-Basque. Kamus tersebut merupakan kamus bahasa Basque bagi pemelajar bahasa Jerman.

Pada penelitian leksikografi dengan menggunakan korpus khususnya pada dua bahasa, terdapat beberapa tahapan penelitian yang dilakukan. Langkah pertama yang dilakukan adalah dengan memasang bahasa yang memiliki populasi yang tinggi dengan populasi yang rendah. Bahasa Basque merupakan bahasa dengan populasi penggunaan yang rendah jika dibandingkan dengan bahasa Jerman. Populasi bahasa yang lebih kecil terhadap pasangan bahasa yang ditentukan dengan metode data paralel dapat dijadikan serta disusun untuk korpus terjemahan. Langkah kedua, yaitu dengan menyusun korpus bahasa Jerman-Basque yang baru. Terdapat sekitar 2 juta token tiap bahasa. Korpus tersebut disimpan dalam format TMX dan dimasukkan ke dalam peranti lunak *SketchEngine*. Pada bagian bahasa Jerman dilakukan lematisasi dan pembubuhan kelas kata (*POS tagging*) dengan menggunakan *TreeTagger*. Sementara itu, bahasa Basque diolah pada tahap tokenisasi saja. Langkah selanjutnya adalah mencari keterangan pada daftar lema berdasarkan korpus. Pada bahasa Jerman, daftar frekuensi lema pada korpora yang besar tersedia di bawah izin publik. Pada alur kerja, peneliti membandingkan DeReWo-40.000 daftar dengan meninjau kembali lema pada tiga kamus, menghapus, menggantikan, atau menambah lema secara manual. Langkah yang sama juga dilakukan dengan bahasa Basque, namun dengan beberapa catatan khusus.

Langkah berikutnya, yaitu membuat naskah kamus semi otomatis. Peneliti mencoba dua perlengkapan pada metode statistik. Hasil yang paling baik dihasilkan melalui GermanNet/EuroWordNet dan Basque WordNet/MCR. Langkah pertama dengan mencoba menggabungkan 16.000 unit leksikal dalam bahasa Basque dan 10.200 unit leksikal dalam bahasa Jerman menjadi 8.500 kata. Selanjutnya, dilakukan penyuntingan entri pada kamus. Pada tahap ini, penyuntingan dilakukan dengan bantuan peralatan (*tools*) di dalam komputer dan juga kemampuan dari pembuat kamus. Langkah terakhir, yaitu dengan melihat konkordansi pada korpus sebagai bagian dari halaman hasil pada bagian pencarian entri di dalam kamus. Dengan menggunakan konkordansi, tidak hanya dapat melihat secara leksikal tetapi juga secara sintaksis.

Pada pembuatan kamus elektronik bahasa Jerman-Basque, peneliti memperhatikan apa yang terjadi di negara Basque, yaitu bahasa Basque merupakan bahasa yang digunakan dalam pelajaran. Bagi bahasa Jerman yang merupakan bahasa asing di sana, buku teks didesain dan ditulis dalam bahasa Jerman dan bahasa Basque tanpa diterjemahkan ke dalam bahasa Spanyol atau Prancis. Guru bahasa Jerman di Basque mengajar menggunakan kamus monolingual, atau menggunakan kamus bahasa Spanyol-Jerman karena kamus tersebut lebih banyak memuat hal-hal berkaitan dengan makro dan mikrostruktur kamus dibandingkan kamus saku bahasa Jerman-Basque yang telah tersedia. Untuk mengisi kekosongan tersebut, peneliti membuat rancangan penelitian pembuatan kamus elektronik bahasa Jerman-Basque yang baru.

Pada bagian makrostruktur, peneliti mulai dengan mengambil *DeReWo*-40.000 daftar kata dengan membandingkan tiga kamus bahasa Jerman, dan mendapatkan sampel sebanyak 4500 lema. Pengetahuan linguistik pembuat kamus, pengetahuan cara pembuatan kamus, korpus baru bahasa Jerman-Basque, dan beberapa kamus yang ada digunakan sebagai sumber untuk menyunting entri kamus. Pada *eudelex*, lema dalam bahasa Jerman ditampilkan pertama sebagai entitas sintaktis, polisemi disajikan pada level kedua beserta terjemahan bahasa Basque dengan padanan peta arti kata dalam bahasa Jerman. Sebagai fitur tambahan, peneliti memperkenalkan konten versi permulaan tampilan web *eudelex* dari *Wiktionary* dan *Wikipedia* bahasa Jerman. Peneliti menjelaskan juga mengenai struktur entri dalam kamus berdasarkan TEI, yaitu terdiri atas informasi sintaksis dan semantik. Selain itu, polisemi juga hadir di dalam struktur entri kamus sebagai bagian dari sintaksis seperti “nomina”, “adjektiva”, “verba transitif”, “verba intransitif”, dan sebagainya.

II. KERANGKA TEORI

Sebagai landasan dasar dalam penelitian ini, pada bagian ini akan dijelaskan beberapa konsep yang dianut dalam tulisan ini. Konsep-konsep tersebut terkait dengan perihal kosakata budaya, serta kamus dan kerja perkamusan.

2.1 Kosakata Budaya

Sebagai landasan dalam penentuan kosakata yang akan dimasukkan ke dalam kamus, digunakan klasifikasi unsur-unsur kebudayaan menurut Koentjaraningrat (1979) dan Duranti (2009). Budaya merupakan sesuatu yang dipelajari dan diturunkan dari generasi ke generasi, mulai dari tingkah laku, interaksi sosial, hingga komunikasi antarmanusia, dan menjadi karakteristik sebuah kelompok masyarakat tertentu (Duranti, 2009).

Duranti (2009) menjelaskan makna budaya berdasarkan 6 hal, yaitu budaya sebagai perbedaan dari alam, budaya sebagai pengetahuan, budaya sebagai komunikasi, budaya sebagai sistem media/ alat, budaya sebagai sistem praktik, budaya sebagai sistem partisipasi.

Pada penjelasan budaya (*culture*) sebagai sesuatu yang berbeda dari alam (*nature*), Duranti (2009:25) menjelaskan pandangan Hegel, bahwa budaya merupakan sebuah proses pemisahan dari alam. Budaya berarti kemampuan untuk melangkah keluar dari sebuah batasan dan mengambil sudut pandang lain. Proses tersebut menjadikan pengetahuan tidak hanya sebatas dari pengetahuan sendiri tapi juga didapatkan dari pengetahuan luar. Budaya sebagai pengetahuan, Duranti (2009:27-28) menjelaskan budaya berdasarkan bahasa. Menurutnya, untuk mengetahui budaya sama seperti mengetahui sebuah bahasa, dan untuk menjelaskan budaya sama seperti menjelaskan bahasa. Selain itu, Duranti juga menjelaskan bahwa budaya sebagai distribusi pengetahuan sosial yang menitikberatkan bahwa proses pemahaman oleh masing-masing individu tidak sama dan tidak semua orang memiliki akses yang sama untuk informasi atau menggunakan teknik yang sama dalam mencapai tujuan.

Budaya sebagai komunikasi berarti melihat budaya sebagai sebuah sistem tanda (Duranti, 2009:33). Budaya merupakan sebuah representasi dunia, sebuah cara untuk memahami realitas dalam menjelaskan cerita, dongeng, deskripsi, teori, proverba, produk artistik dan pertunjukan. Budaya sebagai sebuah sistem media/ alat berkaitan dengan kebutuhan manusia dalam mengontrol lingkungan, seperti hubungan manusia dengan makanan (garpu) dan manusia dengan cuaca (payung) (Duranti, 2009:39). Budaya sebagai sebuah sistem praktik, Duranti (2009:45) menjelaskan bahwa sebuah bahasa merupakan kumpulan dari praktik yang tidak hanya berhubungan dengan sistem kata dan gramatikal, tetapi juga kekuatan simbolik dari sebuah cara berkomunikasi, dengan beberapa sistem klasifikasi, tujuan dan bentuk referensi, leksikon, dan metafora. Budaya sebagai sebuah sistem partisipasi berkaitan dengan budaya sebagai sistem praktik dan didasarkan pada asumsi bahwa setiap tindakan di dunia, termasuk komunikasi verbal sebagai sifat sosial, kolektif, dan kualitas partisipatori. Sebagai contoh, bahasa merupakan bagian berbagai tindakan berbicara yang merupakan sebuah tindakan dari partisipasi pada sebuah komunitas pengguna sebuah bahasa (Duranti, 2009:46).

Sementara itu, Koentjaraningrat (1979:203-204) membagi unsur kebudayaan menjadi tujuh, yaitu bahasa, kesenian, sistem religi, sistem teknologi, sistem mata pencaharian, organisasi sosial, dan sistem ilmu pengetahuan. Ketujuh unsur kebudayaan tersebut merupakan unsur kebudayaan yang universal, yaitu ada di setiap budaya. Tiap unsur kebudayaan tersebut dapat dibagi lagi menjadi beberapa bagian yang lebih terperinci. Bahasa dapat meliputi bahasa tulis dan bahasa lisan. Kesenian dapat meliputi seni musik, seni tari, seni sastra, dsb. Sistem religi dapat meliputi hal-hal yang berkaitan dengan kepercayaan dan upacara keagamaan. Sistem teknologi dapat meliputi alat-alat yang digunakan untuk memenuhi kebutuhan sehari-hari, seperti alat makan, senjata, pakaian. Sistem mata pencaharian dapat meliputi pekerjaan, seperti bertani, bercocok tanam, dan berdagang. Organisasi sosial dapat meliputi hal hal yang berkaitan dengan kekerabatan, perkumpulan, sistem kenegaraan. Sementara itu, sistem ilmu pengetahuan dapat meliputi hal-hal yang berkaitan dengan pengetahuan lingkungan (flora dan fauna) dan pengetahuan tentang manusia.

Berdasarkan penjelasan mengenai unsur-unsur budaya tersebut, sebuah budaya dapat direpresentasikan melalui berbagai aspek, salah satunya melalui bahasa. Meskipun unsur-unsur kebudayaan tersebut bersifat universal, namun terdapat ciri yang membedakan antara sebuah budaya masyarakat tertentu dengan budaya masyarakat lainnya. Sapir – Whorf dalam Jourdan dan Tuite (2006:35) menjelaskan bahwa struktur semantik pada berbagai bahasa yang berpegang pada cara penutur bahasa tersebut berpikir dan bertindak sesungguhnya tidak dapat dibandingkan satu sama lain. Bahasa, cara berpikir, dan budaya saling terhubung, sehingga masing-masing bahasa dapat memiliki asosiasi dengan sudut pandangnya tersendiri. Hal tersebut karena baik bahasa, budaya, serta cara pikir masyarakat pada kebudayaan tertentu memiliki ciri khas yang membedakannya dengan kebudayaan lain. Sebagai contoh, Jourdan dan Tuite (2006: 35-36) menjelaskan bahwa terdapat perbedaan kosakata warna pada masing-masing bahasa. Perbedaan kosakata warna tersebut dapat diperoleh dari penutur bahasa yang berbeda. Berdasarkan penjelasan Sapir-Whorf tersebut, penggunaan kosakata budaya sebagai bagian dari penggunaan bahasa dapat memperlihatkan ciri sebuah budaya masyarakat tertentu. Sebuah kata dapat dikatakan sebagai kosakata budaya ketika kata tersebut menjadi khas dan menunjukkan karakteristik pola kehidupan sebuah kelompok masyarakat. Adapun penggunaan kosakata budaya tersebut menjadi penting dalam menunjukkan identitas pengguna sebuah bahasa.

2.2. Kamus dan Kerja Perkamusan

Dalam leksikografi atau pekerjaan perkamusan, ada beberapa langkah yang harus ditempuh yang mencakup penentuan siapa pengguna kamus, jenis kamus, serta penyajian isi kamus. Untuk memahami langkah-langkah dalam penyusunan kamus tersebut, pada bagian ini akan dijelaskan mengenai pengguna kamus, jenis kamus, mikrostruktur dan makrostruktur kamus, serta pendefinisian kamus.

Berdasarkan penggunaannya, pada umumnya, kamus digunakan sebagai referensi untuk mencari makna sebuah kata, khususnya untuk kebutuhan menulis. Untuk kebutuhan berbicara dan menyimak, kamus juga digunakan untuk memeriksa pelafalan. Selain itu, kamus juga dapat digunakan untuk melihat struktur gramatika sebuah bahasa, etimologi sebuah kata, atau informasi mengenai bunyi sebuah kata (Bogaards, 2003:26—27).

Hal yang juga menjadi bahan pertimbangan di dalam menyusun sebuah kamus adalah pengguna kamus. Terdapat beberapa jenis pengguna kamus. Berdasarkan usianya, pengguna kamus dapat dibedakan atas orang dewasa, anak-anak, atau remaja. Pengguna kamus juga dapat dibedakan atas penutur jati sebuah bahasa atau pemelajar sebuah bahasa. Pengguna kamus juga dapat dibedakan atas pengguna bahasa secara umum atau secara khusus berdasarkan bidang-bidang tertentu (Atkins dan Rundell, 2008: 28—29).

Kamus dapat dibedakan atas beberapa jenis. Atkins dan Rundell (2008: 24—25) mengelompokkan kamus berdasarkan beberapa hal sebagai berikut.

- (1) Berdasarkan jumlah bahasanya, kamus dapat dibedakan atas kamus monolingual, kamus bilingual, dan kamus multilingual.
- (2) Berdasarkan cakupan isinya, kamus dapat dibedakan atas kamus umum, kamus ensiklopedis, kamus khusus (misalnya kamus istilah hukum, kamus istilah olahraga cricket, kamus istilah keperawatan, dll.), serta kamus yang berisi pemakaian bahasa secara khusus (misalnya kamus kolokasi, peribahasa, idiom, dan lain-lain).
- (3) Berdasarkan ukurannya, kamus dapat dibedakan atas kamus standar, kamus ringkas, kamus saku.
- (4) Berdasarkan medianya, kamus dapat dibedakan atas kamus dalam bentuk cetak, kamus dalam bentuk elektronik, dan kamus dalam bentuk *web-based*.
- (5) Berdasarkan pengorganisasiannya, kamus dapat dibedakan atas kamus yang disusun dengan urutan kata lalu artinya dan kamus yang disusun dengan urutan kata, arti, kemudian kata.
- (6) Berdasarkan pengguna kamusnya, kamus dapat dibedakan atas kamus yang digunakan oleh pengguna kamus yang menggunakan bahasa yang sama, kamus yang digunakan oleh dua kelompok yang menggunakan dua bahasa yang berbeda, serta kamus yang digunakan oleh pengguna kamus dari seluruh dunia.
- (7) Berdasarkan kemahiran penggunaannya, kamus dapat dibedakan atas kamus yang digunakan oleh linguistis atau profesional, kamus yang digunakan oleh orang dewasa yang terpelajar, kamus yang digunakan oleh pelajar sekolah, kamus yang digunakan oleh anak-anak, dan kamus yang digunakan oleh orang yang mempelajari sebuah bahasa.
- (8) Berdasarkan tujuan penggunaan kamus, kamus dapat dibedakan atas kamus untuk tujuan *decoding*, yaitu kamus digunakan untuk memahami makna sebuah kata dan digunakan untuk menerjemahkan sebuah teks dari bahasa asing ke bahasa sasaran dan kamus juga digunakan untuk tujuan *encoding*, yaitu kamus digunakan untuk memeriksa penggunaan kata yang kurang tepat, untuk menerjemahkan sebuah teks ke bahasa sasaran, kemudian ke bahasa asing lain, serta untuk mengajar sebuah bahasa.

Kridalaksana (2003: xiii--xiv) menyatakan bahwa mikrostruktur kamus berkaitan dengan susunan lema dalam entri kamus, termasuk artikel dalam ensiklopedia. Mikrostruktur kamus terdiri atas bagian inti dan bagian pelengkap. Pada bagian inti terdapat lema atau kata kepala, sedangkan pada bagian pelengkap terdapat kelas kata, sublema dengan informasinya, contoh, derivasi dan infleksi, lafal, variasi historis sosial dan geografis, gaya bahasa, konteks penggunaan bahasa, sinonim dan antonim, serta etimologi. Sementara itu, makrostruktur kamus berkaitan dengan susunan entri di dalam kamus, yakni yang berkaitan dengan susunan abjad, rujuk silang, kata majemuk dan gabungan leksem yang idiomatis atau bukan, tipografi, gambar atau diagram sebagai ilustrasi, serta petunjuk penggunaan kamus.

Di dalam mikrostruktur kamus, terdapat lema beserta informasinya, khususnya definisi. Di samping lema, definisi merupakan bagian penting di dalam kamus. Atkins dan Rundell (2008: 407) membedakan dua hal dalam pendefinisian kosakata, yaitu *reference* atau *decoding* dan *productive* atau *encoding*. Penyusunan definisi kosakata yang dilakukan untuk tujuan *decoding* adalah jika pengguna kamus yang akan disusun melihat definisi tersebut karena ia menemukan sebuah kata asing dan perlu mengetahui artinya. Sementara itu, penyusunan definisi kosakata yang dilakukan untuk *encoding* adalah jika pengguna kamus yang disusun ingin menulis atau mengatakan sesuatu dengan kosakata tersebut, dan ini termasuk *encoding* dari maksud yang ada dalam pikirannya dengan cara yang natural, sesuai, dan efektif. Berdasarkan penjelasan tersebut dapat dilihat bahwa jika pengguna kamus hanya ingin mengetahui arti sebuah kosakata yang asing untuknya, pendefinisian sebuah kosakata dapat dilakukan dengan lebih sederhana. Sebaliknya, jika pengguna ingin mengetahui arti sebuah kosakata untuk digunakan, baik dalam lisan maupun tulisan, pendefinisian sebuah kosakata tidak hanya memperhatikan konteks pemakaian kosakata, tetapi lebih jauh lagi juga memperhatikan kaitannya dengan semantik, kolokasi, sosiolinguistik, dan pragmatik.

Berkaitan dengan pendefinisian, Atkins dan Rundell (2008: 413—422) menjelaskan beberapa jenis pendefinisian dalam kamus, antara lain definisi dengan model tradisional, prototipe, dan sinonim. Definisi dengan model tradisional mengandung dua unsur definisi, yaitu kata superordinat dan informasi tambahan. Kata superordinat merupakan bagian dalam definisi berupa kategori semantis, sedangkan informasi tambahan merupakan bagian dalam definisi yang berisi penjelasan yang bersifat membedakan satu kata dengan kata lain yang berasal dari kategori yang sama (kohiponimnya). Definisi model tradisional ini juga disebut sebagai definisi berupa genus dan *differentia* (genus dan pembeda). Atkins dan Rundell memberikan contoh kata convertible berikut ini.

car with a folding or detachable roof

Dalam definisi tersebut, car adalah genus, sedangkan pembedanya *convertible* dengan mobil lainnya adalah ‘with a folding or detachable roof’ (Atkins, 2008: 414). Selain itu, Atkins dan Rundell (2008: 417—418) juga menjelaskan pendefinisian dengan model prototipe. Pendefinisian dengan prototipe memperlihatkan definisi yang berisi konsep prototipe sebuah kata. Atkins dan Rundell memberikan contoh definisi kata *bus* berikut ini.

a large motor vehicle motor vehicle carrying passengers by road, typically one serving the public on fixed route and for a fare (ODE-2 2003)

Bagian pertama definisi di atas merupakan informasi yang penting tentang kata *bus*, sedangkan bagian berikutnya merupakan penjelasan yang berisi ciri bus pada umumnya. Pada kenyataannya terkadang terdapat bus yang tidak diperuntukkan bagi masyarakat

umum dan tidak dipungut biaya, misalnya bus sekolah, bus yang mengangkut penumpang dari bandara, dari terminal di bandara ke pesawat, dan bus yang mengangkut pegawai. Definisi kata *bus* di atas tidak mengandung informasi mengenai hal itu semua, tetapi definisi tersebut mengandung informasi yang cukup yang memungkinkan pengguna kamus untuk memahami variasi-variasi lainnya dari prototipe, khususnya prototipe kata *bus*.

Model pendefinisian yang lain yang dijelaskan oleh Atkins dan Rundell adalah definisi dengan model sinonim. Pendefinisian dengan model sinonim dapat efektif ketika sinonim yang diberikan di dalam definisi disertai informasi yang memberikan penjelasan tentang batasan-batasan tertentu dalam persamaan kata tersebut. Akan tetapi, di lain pihak, pendefinisian dapat menjadi kurang efektif karena tidak ada dua kata yang bermakna sama persis. Pendefinisian dengan model sinonim dapat digunakan jika definendum dan sinonim yang digunakan sama persis secara semantis, yang dalam beberapa hal dapat dibedakan antara lain dari sudut registernya, distribusi regional, atau sikap penutur (Atkins dan Rundell, 2008: 420—421).

III.METODE

Penjaringan data kosakata budaya dalam penelitian ini dilakukan dengan menggunakan korpus BIPA yang disusun oleh Program BIPA LBI UI (Suhardijanto dan Diautami 2015). Korpus BIPA yang hingga kini telah mencapai 1,5 juta kata disusun dari teks materi ajar, tulisan mahasiswa, transkripsi tuturan guru dan mahasiswa, dan dokumen lain terkait pembelajaran BIPA, tidak hanya yang ada di lingkungan BIPA UI, tetapi juga di tempat lainnya.

Dalam korpus 1, 5 juta kata yang telah disusun ditemukan beberapa kosakata budaya, baik yang asli Indonesia maupun yang serapan, antara lain yang berkaitan dengan makanan (*bajigur*), kata sapaan (*abang*), kesenian daerah (*angklung*), salam/ungkapan (*assalamualaikum*), adat (*sekaten*) dan busana (*jilbab*).

Dalam pemilihan dan penentuan kosakata, juga digunakan metode korpus seperti n-gram, kolokasi dan konkordansi untuk menyortir dan mengecek pemakaian kosakata tersebut di dalam teks (lebih lanjut lihat Linquist 2010 dan Chen 2011). Penggunaan n-gram dengan format bigram atau trigram dilakukan untuk menemukan kata yang biasanya berkombinasi dengan kosakata budaya. Selanjutnya, untuk mengetahui dalam konteks apa kosakata budaya digunakan dalam penelitian ini dimanfaatkan metode kolokasi. Teknik akan memberikan informasi mengenai konteks pemakaian kosakata budaya yang bersangkutan. Sementara itu, untuk mengecek pemakaian kosakata budaya di dalam korpus, digunakan metode konkordansi atau KWIC (*key word in context*). Ketiga metode tersebut diperlukan untuk membantu merumuskan definisi yang tepat dan sesuai dengan tujuan kamus BIPA. Selain itu, pengecekan pemakaian kosakata melalui teknik konkordansi atau juga dapat bermanfaat untuk menemukan contoh pemakaian yang dapat digunakan dalam pengerjaan mikrostruktur kamus.

3.1 Pengolahan Data

Dalam penentuan kosakata budaya, digunakan beberapa teknik analisis korpus (Linguist 2010, Chen 2011, dan McEnery). Untuk menjangkau kosakata tersebut, digunakan korpus BIPA yang bersumber dari materi ajar dan hasil pekerjaan siswa, baik lisan maupun tulis, serta tuturan guru. Dalam korpus yang berjumlah 1,5 juta kata itu ditemukan beberapa kosakata budaya, baik yang asli Indonesia maupun yang serapan, antara lain yang berkaitan dengan makanan (*bajigur*), kata sapaan (*abang*), kesenian daerah (*angklung*), salam/ungkapan (*assalamualaikum*), adat (*sekaten*) dan busana (*jilbab*).

Setelah kosakata budaya terpilih, disusunlah kriteria untuk menentukan mana kosakata yang pantas dimasukkan sebagai entri di dalam kamus BIPA dan mana yang tidak. Kriteria tersebut

1. *Sesuai dengan tujuan kamus.* Kamus yang akan disusun adalah kamus untuk tujuan pedagogis. Oleh karena itu, kosakata budaya yang dimasukkan sebagai entri di dalam kamus BIPA adalah kosakata yang bermanfaat dan dapat digunakan oleh pemelajar BIPA, baik dalam proses belajar-mengajar dan juga untuk bersosialisasi dengan masyarakat Indonesia.
2. *Bukan merupakan kosakata yang belum diserap dalam bahasa Indonesia.* Sebagai acuan, digunakan KBBI. Kosakata budaya yang akan dimasukkan sebagai entri di dalam kamus merupakan kosakata yang sudah diserap ke dalam bahasa Indonesia, baik kosakata asing maupun kosakata yang berasal dari bahasa daerah.
3. *Berdasarkan frekuensi dan distribusi pemakaian kosakata di dalam korpus.* Adapun terdapat beberapa kriteria dalam pemakaian kosakata di dalam korpus, yaitu frekuensi kosakata lebih dari 3, distribusi pemakaian kosakata minimal digunakan oleh 3 orang siswa baik tulis maupun lisan, distribusi pemakaian kosakata minimal digunakan lebih dari 3 dalam materi ajar.

Berikut ini adalah beberapa kosakata budaya yang ditemukan di dalam korpus beserta frekuensi penggunaannya.

1. Makanan: *gudeg* (4), *bajigur* (7), *rendang* (27)
2. Kata Sapaan: *abang* (6), *mas* (129), *mbak* (167)
3. Kesenian Daerah: *wayang* (20), *angklung* (25), *gamelan* (80)
4. Salam/ Ungkapan: *Insya Allah* (4), *Assalamualaikum* (20), *Alhamdulillah* (33)
5. Adat: *Ngaben* (3), *Sekaten* (9), *gunungan* (8)
6. Busana: *sarung* (7), *kebaya* (8), *jilbab* (135)

3.2 Penyusunan Kamus

Dalam penyusunan kamus, terdapat beberapa tahap yang harus ditempuh, yaitu (1) penentuan pengguna kamus; (2) penentuan jenis kamus (apakah kamus yang akan

disusun termasuk dalam kamus monolingual atau bilingual, kamus umum atau kamus khusus, kamus besar atau kamus terbatas, dan lain-lain), kamus untuk umum atau untuk pelajar, untuk orang dewasa atau anak-anak; (3) penentuan sumber kosakata yang kemudian akan disusun menjadi korpus; (4) penentuan mikrostruktur kamus, termasuk di dalamnya mencakup penetapan kosakata yang akan dimasukkan sebagai lema di dalam kamus dan penetapan pendefinisian kamus; serta (5) penetapan makrostruktur kamus.

Penelitian ini berfokus pada kamus yang sedang disusun, yaitu kamus ekabahasa yang ditujukan untuk pemelajar BIPA sebagai pengguna kamus. Dalam kaitan penyusunan kamus, kamus pemelajar BIPA ini menggunakan korpus BIPA sebagai sumber kosakata. Setelah dilakukan pengolahan terhadap kosakata-kosakata dalam korpus tersebut dan ditetapkan beberapa kosakata yang masuk, langkah berikutnya adalah penyajian kosakata tersebut di dalam kamus. Salah satu langkah penting dalam penyajian kamus tersebut adalah penentuan cara pendefinisian yang tepat untuk sebuah lema, khususnya lema berupa kosakata budaya. Berkaitan dengan pengolahan dan penyajian kamus BIPA ditemukan beberapa permasalahan. Beberapa permasalahan yang muncul dalam pengolahan dan penyajian kosakata tersebut juga akan dijelaskan dalam makalah ini.

IV. ANALISIS PENGOLAHAN DAN PENYAJIAN KOSAKATA DALAM KAMUS PEMELAJAR BIPA

Pada bagian ini akan dijelaskan analisis pengolahan, penentuan, dan penyajian kosakata budaya dalam kamus pemelajar BIPA yang sedang disusun. Karena ada beberapa topik yang menjadi fokus dalam pembahasan ini, bagian ini dibagi menjadi tiga subbagian yang akan membicarakan pengolahan, penyajian, serta permasalahan yang dijumpai dalam pengerjaan kosakata dalam kegiatan penyusunan kamus BIPA.

4.1 Pengolahan Kosakata dalam Kamus Pemelajar BIPA

Sebagaimana telah disebutkan pada bagian sebelumnya, penyusunan kamus BIPA bertujuan pedagogis karena ditujukan untuk membantu para pemelajar bahasa Indonesia. Meskipun demikian, pada tahap awal, penyusunan difokuskan pada pembuatan kamus ekabahasa BIPA yang akan menjadi dasar pengembangan kamus dwibahasa atau multilingual BIPA. Oleh karena itu, kosakata budaya yang dimasukkan sebagai kandidat entri pada tahap ini jelas merupakan kosakata yang bermanfaat dan mendukung kegiatan belajar para pemelajar bahasa Indonesia sebagai bahasa asing. Hal ini segera dapat dipenuhi karena kata yang digunakan sebagai entri dalam kamus BIPA semuanya berasal dari korpus BIPA yang sebagian besar merupakan kompilasi teks hasil karya pemelajar dan teks materi ajar BIPA. Dengan demikian, dapat diasumsikan bahwa sebagian besar data yang ada di dalam korpus BIPA merupakan kosakata yang terkait dengan topik-topik seputar pembelajaran bahasa Indonesia sebagai bahasa asing. Selain

kriteria paedagogis, kosakata yang bermanfaat dalam bersosialisasi dengan masyarakat Indonesia pun dipertimbangkan sebagai kandidat entri pada kamus pemelajar BIPA.

Untuk menyortir data, digunakan ambang frekuensi 3 sebagai batas ambang (*threshold*) sehingga kosakata yang hanya muncul sekali atau dua kali dalam korpus BIPA. Dengan kriteria ini diperoleh, misalnya, beberapa kosakata seperti *angklung* (25), *sarung* (7), dan *gado-gado* (11). Sementara itu, kata seperti *angkringan* (2), *langgar* (1), dan *sesajen* (1) tidak akan masuk menjadi kandidat entri. Kata *angkringan* atau *langgar*, selain berfrekuensi rendah, sebenarnya juga penggunaannya sangat terbatas. Jadi, kata ini menjadi kosakata yang sangat spesifik karena *angkringan* merujuk pada jenis penjaja makanan yang hanya ada di Yogyakarta, sementara *langgar* juga merupakan jenis tempat ibadah kecil bagi muslim di pedesaan, khususnya di Jawa.

Kemudian, kriteria kedua adalah kandidat entri merupakan kosakata bahasa Indonesia yang sudah diserap dan dicatat di dalam KBBI edisi keempat (2008). Dengan demikian, kosakata yang belum diserap jelas bukan merupakan kandidat entri dalam kamus pemelajar BIPA yang sedang disusun. Sebagian besar kosakata budaya merupakan kosakata yang sangat erat kaitannya dengan kebudayaan setempat. Oleh karena itu, sebagian besar kosakata budaya merupakan kosakata asing maupun kosakata yang berasal dari bahasa daerah. Dalam hal ini, digunakan KBBI sebagai dasar penentuan. Dengan demikian, kata seperti *anjeun* (2) jelas tidak akan masuk menjadi kandidat entri.

Selain frekuensi sebagai dasar penentuan kosakata layak atau tidak menjadi kandidat entri, dalam penelitian ini juga digunakan kriteria distribusi pemakaian kosakata di dalam korpus. Jadi, distribusi pemakaian kosakata minimal digunakan dalam dua atau lebih teks. Dengan demikian, jika ditemukan kosakata yang berfrekuensi tiga, namun hanya muncul pada satu naskah teks, kosakata tersebut tidak dipilih sebagai kandidat entri karena persebarannya yang sangat spesifik dan terbatas.

4.2 Penyajian Kosakata dalam Kamus Pemelajar BIPA

Pada bagian kerangka teori telah dijelaskan beberapa jenis kamus. Berdasarkan jumlah bahasa sasarannya, kamus BIPA termasuk dalam kamus ekabahasa. Berdasarkan jangkauan isinya, kamus ini termasuk dalam kamus terbatas, dalam hal ini terbatas pada kosakata dalam bahasa Indonesia yang digunakan oleh pemelajar BIPA. Sementara itu, berdasarkan penggunaannya, kamus BIPA digunakan oleh orang yang mempelajari bahasa Indonesia sebagai bahasa asing.

Seperti yang sudah dijelaskan di atas, kamus BIPA ini disusun berdasarkan korpus BIPA. Setelah dilakukan penyeleksian terhadap kosakata di dalam korpus tersebut, terpilih beberapa kosakata yang masuk sebagai entri kamus. Kosakata tersebut dapat berupa kosakata dasar dan kosakata budaya. Salah satu kegiatan yang dilakukan berikutnya adalah memberikan definisi terhadap kosakata yang masuk sebagai lema. Dalam memberikan definisi terhadap kosakata tersebut terdapat beberapa hal

yang harus diperhatikan, antara lain (1) definisi seperti apa yang dibutuhkan untuk memahami kosakata budaya tersebut secara komprehensif; (2) informasi lain apa lagi yang dibutuhkan untuk memahami kosakata budaya tersebut (misalnya informasi mengenai etimologi dan perkembangan makna sebuah kosakata); (3) bagaimana pembentukan kosakata budaya tersebut (berkaitan dengan pembentukan kata berupa afiksasi, reduplikasi, komposisi, abreviasi, dan lain-lain).

Dalam penyajian kosakata tersebut di dalam kamus BIPA ini, tujuan penyusunan kamus dan pengguna kamus menjadi dasar penentuan bagaimana sebuah kosakata didefinisikan dan diberikan penjelasan lain. Sasaran pengguna kamus ini adalah pemelajar BIPA, yaitu orang yang mempelajari bahasa Indonesia dan menggunakan bahasa Indonesia sebagai bahasa asing. Kamus ini diharapkan dapat membantu pengguna kamus, yaitu pemelajar BIPA, dalam menelusuri kosakata bahasa Indonesia, khususnya bahasa Indonesia yang digunakan oleh pemelajar BIPA dalam kegiatan mempelajari bahasa Indonesia serta dalam kehidupan sehari-hari.

Di dalam kamus BIPA, kosakata budaya membutuhkan penjelasan atau definisi berdasarkan konteks budayanya, yang terkadang membutuhkan penjelasan berupa definisi yang lebih komprehensif. Berdasarkan penjelasan di atas, pengguna kamus BIPA dapat melakukan penelusuran kosakata dalam bahasa Indonesia dengan tujuan untuk mencari makna sebuah kosakata yang asing untuknya dan lebih jauh lagi untuk menerapkan kosakata tersebut ke dalam lisan atau tulisan. Oleh karena itu, dibutuhkan beberapa cara dalam memberikan definisi terhadap sebuah kosakata sehingga makna dari kosakata tersebut sesuai dengan konteks pemakaiannya.

Berdasarkan korpus yang telah disusun, terdapat sejumlah kosakata yang dapat digolongkan ke dalam kosakata budaya. Kosakata budaya tersebut dapat berupa kosakata salam dan ungkapan (*assalamualaikum*, *alhamdulillah*), kata sapaan (*abang*, *mas*, *mbak*), makanan (*bajigur*, *gudeg*, *rendang*), kesenian (*wayang*, *angklung*, *gamelan*), adat (*ngaben*, *sekaten*, *gunungan*), busana (*sarung*, *kebaya*, *jilbab*), dan lain-lain. Agar kosakata tersebut dapat dipahami dengan tepat oleh pengguna kamus, harus ditetapkan bagaimana cara menyusun definisi yang tepat juga.

Selain sebagai pendokumentasian kosakata dalam sebuah bahasa, kamus juga berfungsi memberikan informasi mengenai makna kosakata-kosakata tersebut. Oleh karena itu, definisi kosakata di dalam kamus merupakan hal yang penting. Berdasarkan batasan mengenai pendefinisian yang telah dijelaskan pada bagian kerangka teori, berikut akan dijelaskan definisi beberapa kosakata budaya sebagai contoh.

Berdasarkan korpus yang digunakan sebagai sumber informasi untuk melihat batasan mengenai kosakata budaya di atas, beberapa kosakata dapat didefinisikan dengan model tradisional, antara lain kata kosakata budaya berupa salam/ungkapan, kosakata kesenian, kosakata pakaian, kosakata adat, dan kosakata makanan. Berikut contoh pendefinisian yang dapat diberikan terhadap kosakata-kosakata budaya tersebut.

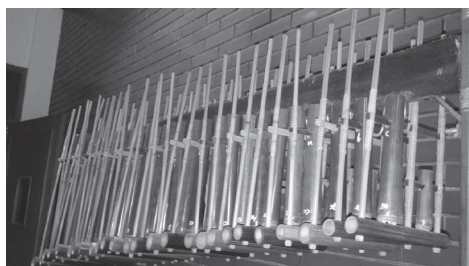
1. Definisi kata *Assalammualaikum*

assalamualaikum	salam yang diucapkan oleh umat muslim ketika bertemu dengan umat muslim lain, ketika membuka sebuah percakapan (melalui telepon, surat, dll.), ketika berkunjung ke tempat atau rumah seseorang, ketika membuka sebuah acara atau pertemuan: <i>Assalammualaikum wr.wb. Melalui surat ini, saya bermaksud mengundang Bapak/Ibu dalam acara sebagai berikut.</i>
-----------------	---

Pada definisi kata *assalammualaikum* di atas, salam merupakan genus, sedangkan informasi lain, yaitu ‘diucapkan oleh umat muslim ketika bertemu dengan umat muslim lain, ketika membuka sebuah percakapan (melalui telepon, surat, dll.), ketika berkunjung ke tempat atau rumah seseorang, ketika membuka sebuah acara atau pertemuan’ merupakan pembeda kata *Assalammualaikum* dari kosakata salam lainnya. Definisi kata *assalamualaikum* di atas juga dilengkapi dengan contoh pemakaian. Berdasarkan korpus yang digunakan, salah satu pemakaian kata tersebut, muncul dalam kalimat salam pembuka sebuah surat.

2. Definisi kata *angklung*

angklung	alat musik tradisional Indonesia yang berasal dari Jawa Barat, yang terbuat dari bambu yang dimainkan dengan cara digoyangkan.
----------	--



Sumber: <http://update-profile.blogspot.co.id/2011/11/sejarah-alat-musik-angklung.html?m=1>

Kata *angklung* juga dapat didefinisikan dengan model tradisional. Kata *alat musik tradisional* pada definisi di atas merupakan genus dan ‘berasal dari Jawa Barat, yang terbuat dari bambu yang dimainkan dengan cara digoyangkan’ merupakan pembeda *angklung* dari alat musik tradisional lain yang juga berasal dari Jawa Barat, misalnya *calung*, atau alat musik tradisional dari daerah lain, misalnya *gamelan*, yang dimainkan dengan cara dipukul. Selain itu, jika pada definisi kata *assalammualaikum* diberikan contoh pemakaian sebagai pelengkap definisi, pada definisi kata *angklung* di atas diberikan dua gambar, yaitu gambar alat musik *angklung* dan gambar orang yang sedang memainkan *angklung*. Kamus ini merupakan kamus pemelajar BIPA. Informasi berupa gambar di atas menjadi informasi yang sangat berguna untuk memberikan gambaran yang lebih lengkap kepada pemelajar BIPA mengenai alat musik *angklung*.

3. Definisi kata *sekatén*

sekatén	upacara tradisional yang dilakukan masyarakat Daerah Istimewa Yogyakarta yang diselenggarakan dari tanggal 5 sampai 11 bulau Rabi’ul Awal, atau dalam kalender Jawa disebut bulan Maulud, dalam rangka memperingati hari kelahiran Nabi Muhammad saw.
Tahapan Upacara Sekaten:	
(1) Pada tanggal 5 Rabi’ul Awal, pukul 16.00—23.00, dibunyikan gamelan sebagai tanda dimulainya upacara Sekaten. Setelah itu, pada pukul 23.00 gamelan dipindahkan ke pagongan yang ada halaman Masjid Agung.	
(2) Setelah itu, di pagongan, gamelan dibunyikan pada siang dan malam hari, kecuali pada waktu salat dan pada hari Jumat.	
(3) Tahap selanjutnya adalah pembacaan riwayat kelahiran Nabi Muhammad saw di serambi Masjid Agung pada tanggal 11 Rabi’ul Awal, pada pukul 20.00—23.00.	
(4) Tahap terakhir adalah pengembalian gamelan dari Masjid Agung ke keraton, pada tanggal 11 Rabi’ul Awal pukul 23.00.	
(5) Pada tanggal 12 Rabi’ul Awal dilakukan penutupan upacara Sekaten dengan menyelenggarakan upacara Garebeg Maulud.	

Kata *sekatén* di atas juga didefinisikan dengan cara tradisional. Genus kata ini adalah *upacara tradisional*. Untuk membedakan dengan upacara tradisional yang lain, diberikan pembeda, yaitu ‘dilakukan masyarakat Daerah Istimewa Yogyakarta yang diselenggarakan dari tanggal 5 sampai 11 bulau Rabi’ul Awal, atau dalam kalender Jawa disebut bulan Maulud, dalam rangka memperingati hari kelahiran Nabi Muhammad saw.’. Selain itu, untuk memberikan gambaran yang lebih jelas lagi mengenai kata *sekatén*, definisi dapat ditambahkan dengan informasi mengenai tahap-tahap upacara. Informasi tambahan ini bermanfaat untuk memberikan penjelasan yang lebih komprehensif serta untuk menambahkan informasi mengenai budaya Indonesia, khususnya mengenai upacara tradisional.

4. Definisi kata *rendang*

rendang	makanan tradisional Indonesia yang berasal dari Sumatera Barat, yang terbuat dari daging, santan, dan dibumbui dengan cabai, bawang putih, bawang merah, ketumbar, kemiri, kunyit, jahe, langkuas, daun kunyit, daun jeruk, daun salam, merica, gula, dan garam.
---------	--

Resep Rendang

Bahan-bahan:

- 1 kg daging sapi
- 1 liter santan dari 2 buah kelapa parut
- 15 siung bawang putih
- 15 siung bawang merah
- 10 buah cabai merah, buang bijinya
- 5 buah kemiri sangrai
- 1 cm jahe
- 1 cm kunyit
- 1 ruas langkuas
- 1 sendok teh merica
- 1 sendok teh ketumbar
- 3 lembar daun jeruk
- 3 lembar daun salam
- 1 lembar daun kunyit
- 2 sendok teh gula
- 2 sendok teh garam



Cara membuat:

Pertama-tama, iris daging. Setelah itu, haluskan bumbu. Tumis bumbu sampai berbau harum kemudian masukkan daging. Aduk hingga semua bumbu merata dan daging berubah warna. Setelah itu masukkan santan encer. Kemudian, masukkan garam dan gula, lalu aduk-aduk sampai santan menyusut setengahnya. Setelah itu, masukkan santan kental, lalu terus aduk sampai daging menjadi empuk dan berwarna kecoklatan.

Sumber: <https://cookpad.com/id/cari/rendang>

Kosakata *rendang* juga didefinisikan dengan pendefinisian tradisional, yaitu dengan kata *makanan tradisional* yang merupakan genus dan ‘berasal dari Sumatera Barat, yang terbuat dari daging, santan, dan dibumbui dengan cabai, bawang putih, bawang merah, ketumbar, kemiri, kunyit, jahe, langkuas, daun kunyit, daun jeruk, daun salam, merica, gula, dan garam’ yang merupakan pembeda kata *rendang* dari makanan tradisional Indonesia lainnya, khususnya yang berasal dari Sumatera Barat, misalnya *sate padang*, atau makanan yang terbuat dari daging, misalnya semur. Beberapa makanan tradisional dari Sumatera Barat terbuat dari bahan daging sapi, misalnya *dendeng batokok*. Untuk lebih memberikan pembedaan antara rendang dengan makanan tradisional lain, khususnya yang berasal dari Sumatera Barat, definisi di atas dilengkapi dengan resep dan gambar. Hal ini tentu saja untuk memberikan pemahaman yang lebih komprehensif kepada pemelajarBIPA, di samping, secara tidak langsung memperkenalkan budaya Indonesia, khususnya makanan tradisional.

Selain pendefinisian secara tradisional, beberapa contoh data berupa kosakata budaya yang diambil dalam penelitian ini dapat didefinisikan dengan sinonimnya. Dalam data yang diambil dari korpus BIPA terdapat kosakata berupa kata sapaan, yaitu kata *abang*, *mas*, dan *mbak* dan kosakata pakaian, yaitu kata *kebaya*, *sarung*, dan *jilbab*. Kosakata *mbak* dan *jilbab* dapat didefinisikan dengan definisi sinonim. Berikut contoh pendefinisian kata *mbak* dan *jilbab*.

5. Definisi kata *mbak*

mbak	(1) kakak perempuan; (2) kata sapaan untuk wanita yang usianya lebih tua di daerah Jawa; (3) kata sapaan untuk wanita yang lebih muda, biasanya dalam situasi belum mengenal wanita itu atau dalam situasi yang berkaitan dengan pelayanan jasa di toko, restoran, dll.
------	---

Definisi (1) pada kata *mbak* merupakan definisi sinonim. Kata *mbak* bersinonim dengan kata *kakak perempuan*. Sementara itu, definisi (2) dan (3) merupakan definisi tradisional. Genus pada definisi (2) dan (3) sama, yaitu *kata sapaan*, tetapi pembeda definisi (2) dan (3) berbeda. Definisi (2) memperlihatkan pendefinisian dari sudut asal daerah kata *mbak*, yaitu ‘wanita yang usianya lebih tua di daerah Jawa’, sedangkan definisi (3) berkaitan dengan situasi pemakaiannya, yaitu ‘wanita yang lebih muda, biasanya dalam situasi belum mengenal wanita itu atau dalam situasi yang berkaitan dengan pelayanan jasa di toko, restoran, dll’.

6. Definisi kata *jilbab*

jilbab	(1) kerudung; (2) penutup kepala yang berbentuk lebar yang digunakan oleh wanita muslim untuk menutupi kepala, leher, sampai ke dada.
--------	---

Seperti definisi kata *mbak*, definisi kata *jilbab* juga dapat dilakukan dengan pendefinisian sinonim (definisi 1) dan cara tradisional (definisi 2). Kata *jilbab* bersinonim dengan kata *kerudung*. Untuk melengkapi informasi perbedaan antara kerudung dan jilbab, diberikan definisi secara tradisional, yaitu dengan memberikan penjelasan berupa genus *penutup kepala* dan pembedanya ‘berbentuk lebar yang digunakan oleh wanita muslim untuk menutupi kepala, leher, sampai ke dada’.

Dengan demikian, dapat dilihat bahwa pendefinisian kosakata budaya dapat dilakukan hanya dengan satu cara, misalnya definisi tradisional saja, seperti definisi pada beberapa kata sebelumnya, atau dengan dua cara pendefinisi, yaitu sinonim dan tradisional, seperti pada definisi kata *mbak* dan *jilbab*.

Di dalam data yang diambil dari korpus terdapat kosakata budaya pakaian, yaitu kata *sarung*, *jilbab*, dan *kebaya*. Kata *jilbab*, seperti dijelaskan sebelumnya, dapat didefinisikan dengan model sinonim (dan termasuk di dalamnya definisi tradisional).

Di pihak lain, kata *kebaya* dapat didefinisikan dengan model prototipe. Berikut contoh pendefinisian kata *kebaya*.

7. Definisi kata kebaya

kebaya	pakaian tradisional untuk wanita yang digunakan di bagian atas tubuh, berlengan panjang, digunakan dengan kain panjang, dan biasanya digunakan dalam kegiatan tradisi budaya atau dalam menghadiri acara-acara formal.
--------	--

Definis kata kebaya di atas merupakan prototipe dari kebaya pada umumnya. Definisi tersebut mengandung pendefinisian tradisional, yang terdiri atas genus *pakaian tradisional* dan pembeda ‘untuk wanita yang digunakan di bagian atas tubuh, berlengan panjang, digunakan dengan kain panjang, dan biasanya digunakan dalam kegiatan tradisi budaya atau dalam menghadiri acara-acara formal’. Seperti yang sudah dijelaskan sebelumnya, dalam definisi prototipe, sebuah kata didefinisikan dengan informasi umum kata tersebut. Dalam definisi di atas, kata *kebaya* didefinisikan dengan ‘pakaian tradisional untuk wanita yang digunakan di bagian atas tubuh, berlengan panjang, digunakan dengan kain panjang’. Dalam perkembangannya, dalam kehidupan sehari-hari, kata *kebaya* juga dapat mengacu pada pakaian tradisional untuk wanita yang digunakan pada bagian atas tubuh, berlengan panjang atau pendek, dan dapat digunakan dengan kain panjang atau rok, bahkan dengan celana panjang. Selain itu, dalam definisi di atas, kata *kebaya* mengacu pada ‘pakaian yang biasanya digunakan dalam kegiatan tradisi budaya atau dalam menghadiri acara-acara formal’. Dalam perkembangannya, saat ini *kebaya* dapat digunakan dalam berbagai kesempatan, baik formal maupun nonformal. Definisi kata *kebaya* di atas tidak mengandung informasi mengenai hal itu yang berkaitan dengan perkembangan pemakaian kata tersebut, tetapi definisi tersebut mengandung informasi yang cukup yang memungkinkan pengguna kamus untuk memahami variasi-variasi lainnya dari prototipe, khususnya prototipe kata *kebaya*.

Berdasarkan contoh pendefinisian pada beberapa kosakata budaya yang diambil dari korpus BIPA di atas dapat dilihat bahwa beberapa kelompok kosakata budaya, yaitu kosakata salam/ungkapan, kesenian (alat musik), adat (upacara adat), dan makanan, dapat didefinisikan dengan pendefinisian tradisional, yaitu berdasarkan genus dan pembedanya. Selain itu, untuk memberikan pemahaman yang lebih komprehensif kepada pemelajar BIPA, beberapa definisi dilengkapi dengan contoh pemakaian, dengan gambar, dengan urutan prosesi, serta dengan resep masakan. Selain untuk memberikan pemahaman, informasi berupa contoh pemakaian, gambar, urutan prosesi, dan resep tersebut, secara tidak langsung juga bertujuan memperkenalkan budaya Indonesia.

Beberapa kelompok kosakata lain, yaitu kosakata kata sapaan, khususnya pada kata *mbak*, dan kosakata pakaian, khususnya pada kata *jilbab*, dapat didefinisikan dengan memberikan sinonimnya. Selain informasi berupa sinonim atau persamaan kata, definisi pada kosakata tersebut dapat dilengkapi dengan definisi tradisional.

Beberapa kosakata budaya juga dapat didefinisikan dengan melihat prototipenya. Pada contoh di atas, dapat dilihat definisi kosakata kebaya berdasarkan prototipenya. Tidak tertutup kemungkinan terdapat kosakata budaya lain yang juga dapat didefinisikan berdasarkan prototipenya.

4.3 Permasalahan yang Muncul dalam Pengolahan dan Penyajian Kosakata Budaya dalam Kamus BIPA

Berdasarkan pembahasan di atas, dalam proses menyusun kamus BIPA, mulai dari menyusun korpus sampai pemberian definisi, muncul beberapa permasalahan, berkaitan dengan pengolahan dan penyajian kosakata budaya.

Dalam pengolahan kosakata permasalahan muncul dalam menentukan batasan mana kosakata budaya yang harus masuk dan mana yang tidak. Persoalan itu tidak hanya bersifat teknis, tetapi juga dari segi teoretis. Karena semua data kosakata budaya dalam penyusunan kamus BIPA ini berasal dari sumber yang sama, yakni korpus BIPA, tentu saja kosakata budaya yang tidak muncul di dalam korpus tidak disertakan sebagai entri kamus. Pada Bagian 3.1 disebutkan bahwa salah satu kriteria pemasukan entri adalah frekuensi dan distribusi. Jika kosakata ditemukan muncul lebih dari tiga kali dan digunakan dalam tiga file berbeda, kosakata tersebut menjadi kandidat untuk dimasukkan ke dalam entri kamus. Permasalahan muncul manakala ada kosakata budaya yang sebenarnya cukup populer, namun ternyata hanya digunakan di dalam satu file. Menurut kriteria yang digunakan pada Bagian 3.1, data kosakata tersebut tidak dapat dimasukkan sebagai entri. Permasalahan lain menyangkut kriteria semantis-kultural dalam penentuan kosakata budaya yang harus masuk ke dalam kamus. Kadang-kadang ditemukan kosakata budaya yang lazim digunakan di dalam masyarakat, namun karena belum menjadi entri dalam KBBI, kosakata tersebut tidak dapat dimasukkan sebagai entri, misalnya *ngabuburit*. Kriteria yang digunakan di sini memang mendasarkan penyeleksian hanya pada entri yang ada di dalam KBBI.

Selain itu, muncul juga permasalahan dalam penyajian kosakata budaya di dalam kamus BIPA, khususnya dalam pendefinisian kosakata. Penyusunan definisi pada kamus ekabahasa seringkali menimbulkan permasalahan, terutama dalam pemilihan kata yang menjadi padanan atau definisi kosakata. Kosakata dalam padanan atau definisi tersebut harus dapat merepresentasikan kosakata tersebut sesuai konteks pemakaian. Akan tetapi, pemakai kamus BIPA ini berasal dari berbagai tingkat, dengan tingkat pemahaman struktur dan kosakata yang berbeda. Oleh karena itu, diperlukan cara yang tepat agar sebuah kosakata, khususnya kosakata budaya, dapat dijelaskan secara komprehensif dan dapat dipahami oleh pengguna kamus. Jika sebuah kosakata budaya kurang dapat dipahami hanya berdasarkan definisi tradisional, sinonim, atau prototipe, seperti contoh definisi yang telah dipaparkan di atas, dapat dicari cara lain untuk memberikan penjelasan terhadap kosakata tersebut. Misalnya dengan memberikan informasi berupa konteks pemakaian (dalam tulisan atau percakapan lisan), dengan memberikan penjelasan berupa gambar atau ilustrasi, atau dengan cara-cara lainnya.

V. PENUTUP

Makalah ini menyajikan mekanisme pengolahan dan penyajian kosakata budaya, yaitu dengan menjelaskan bagaimana kriteria pengolahan kosakata budaya berdasarkan korpus BIPA serta menjelaskan bagaimana penyajian kamus BIPA, khususnya dengan melihat bagaimana pendefinisian kosakata budaya tersebut. Salah satu kebaruan dalam penyusunan kamus ini adalah digunakannya Korpus BIPA yang disusun dari data penggunaan bahasa secara riil baik dalam bentuk materi ajar, transkripsi tuturan siswa dan guru BIPA, tulisan siswa, serta dokumen yang berhubungan dengan BIPA.

Pada makalah ini telah dijelaskan batasan mengenai apa yang dimaksud dengan kosakata budaya serta kriteria yang digunakan untuk menentukan apakah kosakata budaya tertentu dimasukkan sebagai entri kamus atau tidak. Selain itu, juga telah disebutkan juga beberapa permasalahan yang terkait dengan pengolahan kosakata, khususnya dalam hal penentuan entri, dan penyajian kosakata dalam bentuk entri kamus, khususnya strategi penyajian definisinya. Dalam proses pengolahan dan penyajian tersebut ditemukan beberapa permasalahan.

Penyusunan kamus BIPA ini masih terus berjalan sehingga penyesuaian dapat saja dilakukan dalam perjalanan waktu untuk tujuan penyempurnaan. Saat ini, penyusunan lebih diprioritaskan untuk pembuatan kamus ekabahasa BIPA sebagai dasar. Untuk pengembangan selanjutnya, akan diutamakan penyusunan kamus dwibahasa BIPA untuk bahasa Inggris-bahasa Indonesia dan sebaliknya. Pada masa depan, bukan tidak mungkin akan disusun pula kamus dwibahasa BIPA untuk pemelajar dari Jepang, Korea, dan negara ASEAN karena pada saat ini merekalah yang mendominasi peserta ajar BIPA di beberapa institusi penyelenggaraan BIPA.

Daftar Pustaka

- Amalia, D. (2014). *Formulasi Pendefinisian dan Model Pengentrian Verba dalam Kamus Pemelajar Bahasa Indonesia*. Disertasi pada Program Doktor Ilmu Linguistik Universitas Indonesia.
- Atkins, B.T. and Michael Rundell. (2008). *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press.
- Bogaards, Paul. (2003). "Uses and Users of Dictionaries" dalam *A Practical Guide to Lexicography*. Piet van Sterkenburg (Ed.). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Chaer, Abdul. (2007). *Leksikologi dan Leksikografi Indonesia*. Jakarta: Rhineka Cipta.
- Cheng, W. (2011). *Exploring Corpus Linguistics: Language in Action (Routledge Introductions to Applied Linguistics)*. New York: Routledge.

- Cruse, D. A. (1992). *Meaning in Language, an Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.
- Duranti, Alessandro. (2009). *Linguistics Anthropology*. Cambridge: Cambridge University Press.
- Jourdan, Christine dan Tuite, Kevin. (2006). *Language, Culure, and Society*. Cambridge: Cambridge University Press.
- Koentjaraningrat. (1979). *Pengantar Ilmu Antropologi*. Jakarta: Aksara Baru.
- Kridalaksana, H. (2003). "Sambutan Ilmiah Kepala Pusat Leksikologi dan Leksikografi Fakultas Ilmu Pengetahuan Budaya Universitas Indonesia." Dalam Lilie Suratminto dan Munawar Holil (ed.). *Rintisan Kajian Leksikologi dan Leksikografi*. Depok: Fakultas Ilmu Pengetahuan Budaya, Universitas Indonesia.
- Lindemann, D. (2013). Bilingual Lexicography and Corpus Methods. The Example of German-Basque as Language Pair. *Procedia-Social and Behavioral Sciences*, 95, 249-257.
- Lindquist, H. (2009). *Corpus Linguistics and the Description of English*. Edinburg: Edinburg University Press.
- McEnery, T., dan Hardie, H. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Purwaningsih. Ernawati. (2014, Januari 13). *Upacara Tradisional Sekaten*. 5 Juli 2016. <http://bpadjogja.info/article/site/view/id/594/t/upacara-tradisional-sekaten>
- Suhardijanto, T. dan Diautami, A. (2015). Frekuensi dan Perbandingan Kosakata Dasar: Informasi dari Korpus Elektronik BIPA Universitas Indonesia. Makalah disajikan dalam Konferensi Internasional Pengajaran BIPA, Hotel Harris Sunroad, Denpasar, Bali, 30 Oktober-3 November 2015.
- Zgusta, L. (1971). *Manual of Lexicography*. The Hague, Paris: Mouton.

Aspek-Aspek dalam Penanganan Kata-Kata Budaya dalam Kamus Besar Bahasa Indonesia: Nama-Nama Makanan Indonesia

Defina dan Renny Soelistiyowati

Dosen Bahasa Indonesia IPB

defina@ipb.ac.id; fina_faaisal@yahoo.co.id; najwa_gendut@yahoo.com

Abstrak

Kamus Besar Bahasa Indonesia (KBBI) merupakan rujukan bagi masyarakat Indonesia untuk mengetahui makna sebuah kata. Bahkan, KBBI juga menjadi rujukan bagi warga negara asing yang mempelajari bahasa Indonesia (BIPA). Saat ini, KBBI pun akan dikembangkan dengan memuat lema-lema atau kata-kata yang ada di Indonesia, termasuk kata-kata dalam bahasa daerah. Salah satu kelompok kata atau lema yang bisa dimasukkan itu adalah lema-lema atau kata-kata kebudayaan, terutama kebudayaan daerah di Indonesia. Begitu banyak aspek budaya yang dapat dimuat dalam KBBI, namun dalam penulisan ini, aspek budaya yang dibahas hanyalah makanan atau kuliner Indonesia. Untuk itu, tujuan penulisan ini adalah 1) mengkaji penjelasan istilah makanan Indonesia yang dimasukkan ke dalam KBBI, 2) mengidentifikasi nama-nama makanan yang ada dalam KBBI. Metode dalam penulisan ini adalah metode deskriptif analitis. Teknik pengumpulan data dengan memilih 20 kosakata secara acak kata-kata atau kosakata makanan dalam KBBI edisi ke-3 dan ke-4. Hasil dari tujuan pertama adalah penjelasan kosakata makanan Indonesia masih kurang, yakni dari segi a) asal daerah makanan itu, b) proses pembuatan, c) alat dan bahan yang belum rinci, d) ciri khas rasa makanan itu, e) nama makanan itu di kalangan masyarakat. Hasil tujuan kedua adalah masih banyak nama-nama makanan daerah yang belum dimuat dalam KBBI. Kesimpulannya adalah 1) nama-nama makanan Indonesia yang ada dalam KBBI sebagai salah satu unsur budaya masih perlu penambahan penjelasan, dan 2) masih banyak nama-nama makanan daerah yang sudah banyak dikenal belum dimuat dalam KBBI.

Kata kunci: aspek budaya, KBBI, nama-nama makanan

I. PENDAHULUAN

1.1. Latar Belakang

Saat ini, bahasa Indonesia sedang digalakkan menjadi bagian dari bahasa-bahasa internasional. Penggalakan ini ditegaskan kembali pada Kongres Bahasa X dengan tema “Penguatan Bahasa Indonesia di Dunia Internasional”. Sehubungan dengan hal itu, tentu banyak hal-hal yang harus dipersiapkan, salah satunya adalah semakin melengkapi *Kamus Besar Bahasa Indonesia*, sebagai rujukan dalam pembelajaran bahasa dan budaya Indonesia. Pengajar, pegiat, dan peserta BIPA tentu membutuhkan referensi untuk mengetahui arti sebuah kata. Mereka akan berpedoman pada KBBI untuk memperkenalkan kosakata dalam bahasa Indonesia, baik kosakata formal maupun informal. Bahkan, ketika mengajarkan budaya Indonesia kepada peserta ajar BIPA tingkat pemula, pengajar BIPA akan membuka KBBI untuk mempernalkan dana kain/tenunan Indonesia, rumah adat, alat musik, seni tari, termasuk nama-nama makanan Indonesia, dan nilai-nilai/norma.

Terkait dengan hal itu, pada bulan April 2016 lalu, Badan Pengembangan dan Pembinaan Bahasa, Kemdikbud, telah menyelenggarakan *workshop* terkait dengan Kamus Besar Bahasa Indonesia (KBBI). Di *workshop* itu diperoleh informasi bahwa KBBI Edisi ke-5 segera akan diterbitkan. KBBI Edisi ke-5 ini dicanangkan akan memuat dua ratus ribu lema.

Dari hasil *workshop* itu ada rencana untuk memasukkan aspek-aspek budaya daerah ke dalam KBBI. Aspek-aspek budaya itu dapat berupa nilai-nilai, kebiasaan, dan karya seni, termasuk nama makanan-makanan masyarakat di setiap daerah di Indonesia. Meskipun selama ini dalam KBBI edisi sebelumnya sudah dimasukkan aspek-aspek budaya itu, akan lebih baik lagi jika aspek-aspek budaya itu perlu ditinjau ulang, baik dari segi penambahan lema maupun dari segi penyempurnaan makna.

Penelitian tentang leksikografi sudah banyak dilakukan. Salah satu artikel tentang leksikografi ditulis oleh Alberts (2011). Dalam abstraknya, diungkapkan bahwa leksikograf adalah mendokumentasikan kata-kata dalam kosakata bahasa umum. Selanjutnya, terkait dengan KBBI sebagai kamus yang memuat kosakata bahasa umum, penelitian tentang KBBI sudah banyak dilakukan, di antaranya oleh Supriyanti (2012). Pada kesimpulannya, Supriyanti (2012:153) mengatakan bahwa secara umum kamus hanya menggambarkan konsep yang berkembang di masyarakat. Permasalahan gender memengaruhi praktik leksikografi. Artinya, lema-lema yang ada dalam kamus adalah lema-lema yang ada di masyarakat. Namun, dari penelitian ini belum diketahui dengan pasti: apakah semua lema atau konsep yang ada di masyarakat sudah dimasukan dalam KBBI? Sebab, penelitian Supriyanti lebih difokuskan pada kosep gender. Bagaimana dengan konsep nomina untuk nama-nama makanan sebagai bagian dari budaya Indonesia? Apakah semua jenis makanan Indonesia sudah dirangkum dalam KBBI. Sebagai kamus besar, apakah KBBI sudah mendeskripsikan dengan lengkap kosakata jenis makanan itu?

1.2. Rumusan Masalah

Sehubungan dengan hal itu, dalam tulisan ini ada dua rumusan masalah. Kedua rumusan masalah itu adalah sebagai berikut.

1. Bagaimana penjelasan istilah makanan Indonesia yang dimasukan ke dalam KBBI?
2. Apakah semua nama makanan ciri daerah sudah dimuat dalam KBBI?

1.3. Tujuan

Berdasarkan rumusan masalah itu, tujuan penulisan ini juga ada dua. Kedua tujuan itu adalah sebagai berikut: 1) mengkaji penjelasan istilah makanan Indonesia yang dimasukkan ke dalam KBBI, dan 2) mengetahui beberapa nama makanan daerah yang ada dalam KBBI.

II. KERANGKA TEORI DAN METODE

Pada bagian kedua ini akan dijelaskan dua hal. *Pertama*, akan dijelaskan kerangka teori yang akan digunakan dalam penelitian ini. *Kedua*, akan dijelaskan metode yang digunakan dalam penelitian.

2.1. Kerangka Teori

Ada dua metode untuk mendeskripsikan kata-kata menurut Halliday *et.al.* (2004:3-4). Kedua metode itu adalah menulis kamus dan menulis *thesaurus*. Tesarus memuat kata-kata yang memiliki makna sama lalu dikelompokkan. Sebaliknya, kamus memuat kata-kata yang diatur dengan sederhana sehingga mudah ditemukan, seperti diatur berdasarkan alpabet.

Fungsi kamus menurut Chaer (2007:184) adalah menghimpun semua kosakata suatu bahasa. Kosakata di sini berarti kosakata yang hidup di tengah masyarakat sebagai konsep-konsep budaya masyarakat bahasa tersebut. Kamus ini pun banyak jenisnya, berdasarkan ukurannya ada namanya kamus besar.

Seperti yang telah diketahui, kamus disusun dari korpus data. Korpus data ini menurut Chaer (2007:214) adalah hal yang substansi bahasa sumber. Lebih lanjut dikatakan bahwa jika bahasa sumber tidak memiliki ragam bahasa tulisan, korpus data dapat diperoleh melalui perekaman bahasa tersebut dari penuturnya lalu hasil rekaman ditranskrip ke dalam bentuk bahasa tulisan, mulai dari ejaan fonetis sampai ejaan ortografis. Sebaliknya, jika bahasa sumber memiliki ragam bahasa tulisan, korpus data diambil dari naskah, buku, koran, majalah dan terbitan lain.

Konsep korpus yang dikemukakan oleh Chaer juga sudah dikemukakan sebelumnya oleh Charteris-Black (2004:31). Korpus menurut Charteris-Black adalah sebagai berikut.

A corpus is any large collection of text that arise from natural language use; in linguistic context, it is in contrast to other types of text that were invented specifically for illustrating a point about language. The notion of attested language is very important in corpus linguistics and implies that data are not invented for the benefit of a model but rather that the model emerges from large and representative sample of language. Other than this, then are no constraints on corpus composition nor are there any constraints on corpus size.

Dari kutipan tersebut, dapat dipahami bahwa sebuah korpus adalah setiap koleksi teks yang muncul dari penggunaan bahasa alami. Tidak ada batasan untuk komposisi korpus, begitu pun dengan ukurannya. Hal ini juga sesuai dengan pendapat Stubbs (2001) dalam Charteris-Black (2004:31). Stubbs mengungkapkan bahwa korpus dapat memberikan koleksi berbagai tipe teks. Korpus dapat dirancang untuk berbagai tipe teks yang lebih besar, termasuk tulisan, pengucapan, formal, nonformal, fiksi, nonfiksi, bahasa yang dihasilkan anak-anak dan orang dewasa dan teks dari periode yang berbeda.

Dengan demikian, KBBI sebagai salah satu jenis kamus berdasarkan ukurannya, awalnya disusun dari korpus dan tidak dibatasi dalam ukurannya. Dalam KBBI dapat dimuat bahasa tulisan dan bahasa lisan, baik formal maupun nonformal serta baik yang dihasilkan oleh orang dewasa maupun anak-anak.

2.2. Metode

Metode dalam penulisan ini adalah metode deskriptif analitis. Teknik pengumpulan data dengan memilih secara acak kata-kata/kosakata makanan dalam KBBI Edisi ke-3 dan ke-4. Untuk tujuan pertama, jumlah kata atau istilah makanan Indonesia yang dipilih 20 kata secara acak, yakni 1) *kalio*, 2) *tekwan*, 3) *rawon*, 4) *rending*, 5) *palai*, 6) *nasi*, 7) *lemang*, 8) *gudeg*, 9) *pempek*, 10) *tahu sumedang*, 11) *pecel*, 12) *asinan*, 13) *soto*, 14) *tahu*, 15) *lepat*, 16) *bikang*, 17) *satai*, 18) *rujak*, 19) *dodol*, dan 20) *ketoprak*. Dari beberapa kata akan dibandingkan dengan yang ada di Wikipedia.org. Semua kata itu dianalisis penjelasannya. Selanjutnya, untuk tujuan ke-2, penulis memberikan beberapa kata atau istilah makanan yang penulis peroleh secara lisan dalam kehidupan sehari-hari lalu mengecek keberadaan kata-kata tersebut di KBB Edisi ke-4 dan membandingkannya dengan Wikipedia sebagai salah satu media rujukan untuk mencari sebuah kata.

III. PEMBAHASAN

Sesuai dengan tujuan, pada bagian pembahasan ini ada dua hal yang akan dibahas. Pertama adalah menguraikan penjelasan istilah nama makanan Indonesia. Kedua adalah mengidentifikasi keberadaan beberapa nama makanan di Indonesia dalam KBBI.

3.1 Penjelasan istilah makanan Indonesia

Seperti yang diketahui, budaya daerah adalah cikal bakal budaya nasional. Artinya, makanan daerah sebagai budaya daerah adalah representatif dari budaya Indonesia. Sesuai dengan konsep kamus yang dikemukakan oleh Chaer bahwa kamus merangkum semua kosakata, dalam hal ini kosakata yang ada bahasa Indonesia. Kosakata itu adalah perwujudan dari budaya Indonesia.

Sesuai dengan pendapat Chaer bahwa kamus besar lengkap dengan ejaan, makna, lafal, etimologi dan kelaziman penggunaan, dalam pembahasan ini lebih difokuskan pada penjelasan berupa makna kata yang memuat etimologinya. Hanya saja, untuk kosakata makanan, etimologi bisa diperluas kepada asal daerah makanan itu, proses pembuatan, alat dan bahan yang belum rinci, ciri khas rasa makanan itu, nama makanan itu di kalangan masyarakat.

Pertama adalah penjelasan makna kata yang memuat etimologi asal daerah makanan itu. Di dalam KBBI, dapat dilihat pada bagian “Petunjuk pemakaian kamus”, yakni halaman xxxiii, nomor 4.c.2) bahasa daerah bahwa suatu lema bahasa ada diberi pengkodean sebagai bahasa daerah, seperti *Mmk* yang artinya bahwa kata itu berasal dari bahasa Mimika, *Mk* (Minangkabau), *Dy* (Dayak) dan sebagainya. Setelah diperhatikan, untuk lema-lema makanan, ada beberapa lema yang diberi kode asal daerahnya, namun ada juga yang tidak. Contohnya adalah kata *palai* (h. 1003) diberi kode *Mk*: “*n* pepes ikan kecil-kecil” dan kata *kalio* (h. 609) juga diberi kode *Mk* “*n* masakan dr daging dsb yg diberi santan dan bumbu, dimasakan hampir kering; rendang basah”. Artinya, kata *palai* dan kata *kalio* itu berasal dari Minangkabau atau digunakan dalam masyarakat Minangkabau. Contoh lainnya adalah *tekwan*. Penjelasan untuk kata *tekwan* (h. 1423) adalah “*Plb n* masakan berkuah dng isi pempek berbentuk kecil-kecil, irisan bengkuang bentuk korek api, jamur kuping, sedap malam, dsb ditaburi abon ebi, seledri, dan bawang goreng”. Artinya, kode *Plb* menunjukkan bahwa kata itu berasal dari Palembang.

Namun, kata rendang (h. 1163) tidak ada penjelasan masakan Minang atau Padang. Padahal, tidak lazim lagi di kalangan masyarakat Indonesia bahwa rendang itu adalah masakan yang berasal dari suku Minangkabau. Contoh lainnya adalah *rawon*. Kata *rawon* (h. 1149) diberi penjelasan sebagai berikut: “masakan (lauk) berkuah dibuat dr irisan daging dengan bumbu utamanya keluak, ditambah rempah2 lain”. Kata ini tidak diberi penjelasan asal daerahnya. Tidak semua suku di Indonesia mengenal makanan ini, padahal makanan ini berasal dari Surabaya.

Sementara itu, ada beberapa lema makanan tidak diberi kode asal daerah, tetapi dalam maknanya ada tambahan penjelasan asal daerahnya. Contohnya adalah kata *gudeg pempek*, dan *ketoprak*. Kata *gudeg* (h. 463) diberi penjelasan sebagai berikut: “masakan yg dibuat dr buah nangka muda diberi bumbu bersantan (masakan khas yogya)”. Contoh lainnya adalah pempek. Kata *pempek* (h. 1043) dijelaskan “*n* penganan khas daerah Palembang dr adonan tepung terigu dan ikan, dimakan dng kuah yg bercuka”. Begitu pun dengan kata *ketoprak* (h.691) diberi penjelasan: “*n* makanan (khas Jakarta)

terdiri atas ketupat, tahu goreng, kerupuk, dan taoge yg dibubuhi bumbu kacang yang berkecap.” Jadi, ada penjelasan asal makanan tersebut, yakni Yogyakarta, Palembang, dan Jakarta.

Kedua adalah penjelasan makna kata yang memuat etimologi proses pembuatan makanan itu. Dalam memberikan makna, ada kata-kata makanan itu yang diberikan penjelasan proses pembuatannya secara singkat, namun ada juga yang tidak ada proses pembuatannya. Contoh kata yang ada penjelasan proses pembuatannya adalah kata *rendang*. Kata *rendang* diberi penjelasan sebagai berikut: “daging yang digulai dengan santan sampai kuahnya kering sama sekali, yang tinggal hanya potongan daging dengan bumbu”. Jadi, ada proses pembuatannya yakni daging digulai dengan santan sampai kering. Contoh kata yang tidak ada penjelasan proses pembuatannya adalah kata *palai*. Kata *palai* diberi penjelasan sebagai berikut: “pepes ikan kecil-kecil”. Semestinya, penjelasan untuk palai dapat ditambahkan prosesnya, yakni kelapa yang sudah diparut digiling dengan bumbu lalu dimasukan ikan kecil-kecil dan dibungkus dengan daun pisang, setelah itu dibakar dengan bara api.

Ketiga adalah penjelasan makna kata yang memuat etimologi alat dan bahan pembuatan makanan itu. Pada penjelasan kata atau nama makanan, masih ada penjelasan alat dan bahannya yang belum rinci. Contohnya adalah kata *lemang*. Penjelasan kata *lemang* (KBBI Edisi ke-3, 2005:655) adalah sebagai berikut: “makan dari pulut diberi santan, dipanggang dl tabung”. Dari penjelasan itu dapat dilihat bahwa ada kata *tabung*. Pada KBBI Edisi ke-4 (h. 808) kata *lemang* diberi penjelasan : “penganan yg dibuat dr ketan diberi santan, dipanggang dl buluh bambu. Untuk kata *tabung* sebagai alat, kata ini tidak diberi penjelasan lagi, yakni bahan tabung itu apa? Namun, pada Edisi k-4 diberi penjelasan bambu. Namun, bambu itu banyak jenisnya dan tidak semua bambu dapat digunakan untuk *tabung lemang*. Jika dibuka kata *talang*, ada 5 nomor penjelasan, *talang* no 4 (h. 1383) diberi penjelasan “n, buluh tipis yg banyak mengandung air, biasanya untuk tabung memasak lemang”. Jika dibuka lema *nasi*, akan ada kata *nasi lemang* (h.953) dengan penjelasan “beras ketan dengan santan yang dimasak dl tabung bambu”. Jadi, penjelasan bambu ada pada lema *talang*, bukan pada lema *lemang*.

Keempat adalah penjelasan makna kata yang memuat etimologi rasa makanan itu. Sebagai rujukan, KBBI akan digunakan untuk memahami makna kata-kata, terutama kata-kata umum karena sesuai dengan pendapat Chaer (2007: 199) KBBI termasuk kamus deskriptif atau kamus pemerai yang lengkap dengan ejaan, makna, lafal, etimologi, dan kelaziman penggunaan. Akan tetapi, dari penjelasan jenis makanan dirasa masih ada yang kurang dijelaskan, yakni rasa makanan tersebut. Sementara itu, dirasa perlu ditambahkan penjelasan rasa makanan itu sebagai salah satu ciri khas rasa makanan itu. Contohnya adalah *gudeg* (h. 463): “masakan yg dibuat dr buah nangka muda diberi bumbu bersantan (masakan khas yogya), bisa ditambah penjelasan rasa, yakni manis. Begitu juga dengan makanan *pempek*, bisa diberi penjelasan manis asam dan *rendang* bisa diberi penjelasan pedas.

Penambahan rasa ini dapat dilakukan karena pada edisi ke-4 sudah ada kata makanan yang diberi penjelasan rasanya. Pada kata *tahu sumedang* (h. 1377), misalnya, diberi penjelasan rasa: “tahu yg berasal dr Sumedang, apabila digoreng, terdapat rongga di dalamnya, rasanya **gurih**.”

Kelima adalah penjelasan makna kata yang memuat etimologi nama makanan yang dikenal di kalangan masyarakat. Semestinya, di KBBI dapat ditambahkan nama makanan itu di kalangan masyarakat. Misalnya, *pecel*, makanan *pecel* (h.1034) dijelaskan sebagai berikut “*n* makanan yg terdiri atas sayuran rebus, spt kacang panjang, bayam, taoge yg disiram dng kuah sambal kacang dsb”. Sementara itu, makanan *pecel* yang berkembang di masyarakat lain adalah *pecel lele*. Pada makanan *pecel lele*, tidak ditemukan sayuran yang disebutkan dalam kamus untuk kata *pecel*. Contoh lainnya adalah asinan (h.93): “sayuran atau buh-buahan yg diawetkan dl cuka yg diberi garam, gula, dan rempah-rempah”. Di kalangan masyarakat juga dikenal nama asinan, bahkan dikenal nama asinan Bogor. Begitu juga dengan makanan soto, di KBBI (h. 1332) diberi penjelasan “masakan yang kuahnya dimasak tersendiri dan rangkaian isinya antara lain daging, kentang, bawang goreng yg dimasukkan kemudian, pada waktu akan dihidangkan”. Dari penjelasan tersebut tidak ada penjelasan jenis-jenis soto, seperti *soto Makasar*, *soto Padang*, *soto Surabaya*, dan *soto Lamongan*. Padahal, sudah diketahui bahwa soto Makasar dengan soto Makasar dan soto lainnya berbeda bahannya sehingga berbeda rasanya. Sebaliknya, jika dibuka kata *tahu*, nasi, *rujak*. Kata *tahu* (h. 1377), akan ditemukan kata *tahu sumedang*: “tahu yg berasal dr Sumedang, apabila digoreng, terdapat rongga di dalamnya, rasanya gurih.” Kata *nasi* (h.953) memiliki banyak jenisnya, di antaranya ada *nasi aking*, *nasi anjing*, *nasi beriani*, *nasi detus*, *nasi gandul*, *nasi golong*, *nasi goreng*, *nasi gurih*, *nasi jaha*, *nasi jambalang*, *nasi kebuli*, *nasi kerak*, *nasi kuning*, *nasi kunyit*, *nasi leman*, *nasi lengat*, *nasi lengko*, dan *nasi liwet*. Selain itu, ada kata *rujak* (h. 1187), dari lema *rujak* ada jenis-jenis *rujak*, yakni *rujak bebek*, *rujak buah dingin*, *rujak cingur*, *rujak colek*, *rujak gobet*, *rujak manis*, *rujak serut*, dan *rujak tumbuk*. *Rujak cingur*, misalnya, ini adalah makanan dari Surabaya. Dari penjelasan di dalam KBBI, dapat dilihat bahwa bahan-bahan *rujak cingur* ini berbeda dengan bahan-bahan *rujak* lain, begitu pun dengan prosesnya juga berbeda-beda sehingga rasanya akan berbeda pula.

Begitu pun dengan nama makanan *lepat* (h. 818), yakni “penganan yang dibuat dr ketan dan kelapa parut serta diberi garam dibungkus dengan daun kelapa muda berbentuk silinder dan direbus. Padahal, ada *lepat* yang disebut *lapek* (bahasa Minangkabau) dengan bahannya bisa dari pisang ditambah kelapa dan ketan lalu dibungkus dengan daun pisang kemudian dikukus atau direbus. Selain itu, juga dikenal nama *lepat* atau *lapek Bugis*. Bahkan, makanan *bikang*, di masyarakat dikenal nama *bika Ambon* dan *satai* dikenal *sate* seperti *sate Padang*, *Sate Ayam Ponorogo*. Penganan *dodol* (h.337), juga ada dikenal *dodol Garut*, tetapi di KBBI tidak ada, yang ada hanya *dodol*. Sebaliknya, jika dibuka Wikipedia.org akan terdapat banyak penjelasan tentang soto, dodol dan sate,

termasuk dodol Garut dan dodol Kandangan (dari Kalimantan Selatan). Di Wikipedia.org ada *soto Jepara*, *soto Semarang*, *soto Kudus*, *soto Betawi*, *soto Padang*, *soto Bandung*, *soto Sokaraja*, *soto Banjar*, *soto Medan*, dan *coto Makassar*.

3.2 Identifikasi keberadaan beberapa nama makanan di Indonesia dalam KBBI

Sehubungan dengan rencana pengembangan KBBI Edisi ke-5, dapat dimasukan nama-nama makanan yang ada di masyarakat Indonesia. Masih ada nama-nama makanan suatu daerah atau suku yang sudah dikenal oleh masyarakat daerah atau suku lainnya bahkan nama makanan berupa penganan. Contoh penganan yang dikonsumsi masyarakat suatu daerah adalah *cireng*. Penganan ini adalah penganan masyarakat Sunda dan sudah dikenal di kalangan masyarakat Indonesia, terutama di Jabodetabek. Sangat banyak penganan yang dapat dimasukkan ke dalam KBBI. Beberapa contoh penganan yang dapat dimasukan ke dalam KBBI adalah *nastar*, *lanting*, *kembang loyang*, *combro*. Untuk kata *kembang*, *kembang loyang* yang ada di KBBI (h.662) “*ark* bubur yg terbuat dr tepung beras bercampur gula”, sedangkan yang ada di masyarakat adalah penganan yang terbuat dari tepung beras dan diberi gula lalu dicetak dengan cetakan pipih sehingga saat digoreng hasilnya lebih tipis dan garing.

Jika dilihat Wikipedia.org., kata *nastar*, *lanting*, *kembang loyang*, *combro* ada. Kata *nastar*, misalnya, diberi penjelasan: “Nastar adalah sejenis kue kering dari adonan tepung terigu, mentega dan telur yang diisi dengan selai buah nanas. Asal katanya dari bahasa Belanda ananas.” Kata *lanting* diberi penjelasan: “Lanting Makanan Khas Kebumen”. Kata *kembang Loyang*, diberi penjelasan “...memiliki rasa yang renyah dan gurih. Kue kembang goyang juga dikenal dengan nama kue kembang loyang atau kue loyang di Sumatera.” Kata *combro* diberi penjelasan sebagai berikut: “combro atau kadang disebut comro atau gemet merupakan makanan khas dari Jawa Barat. Combro terbuat dari parutan singkong yang dibentuk bulat yang bagian...”

IV. PENUTUP

Sesuai dengan tujuan penulisan, kesimpulannya dalam tulisan ini ada dua. *Pertama*, nama-nama makanan Indonesia yang ada dalam KBBI sebagai salah satu unsur budaya masih perlu penambahan penjelasan. *Kedua*, masih banyak nama makanan daerah yang sudah banyak dikenal belum dimuat dalam KBBI, termasuk penganan sehari-hari.

Daftar Pustaka

- Alberts, Mariëtta, 2011. "Lexicography versus Terminography." *Lexikos* 11. <https://doaj.org/article/1177b967d9424ad0a83d22fad14ac441>(diunduh 16 Juni 2016)
- Chaer, Abdul. 2007. *Leksikologi&Leksikografi Indonesia*. Jakarta: Rineka Cipta.
- Charteris-Black, Jonathan. 2004. *Corpus Approaches to Critical Metaphor Analysis*. United Kingdom: Palgrave Macmillan Ltd.
- [Depdiknas] Departemen Pendidikan Nasional. 2008. *Kamus Besar Bahasa Indonesia*. Edisi ke-4. Jakarta: Gramedia.
- [Depdiknas] Departemen Pendidikan Nasional. 2005. *Kamus Besar Bahasa Indonesia*. Edisi ke-3. Cet. Ke-3. Jakarta: Balai Pustaka.
- Halliday, MAK, et.al. 2004. *Lexicology and Corpus Linguistics an Introduction*. New York: Continuum.
- Supriyanti, Nanik. 2012. Praktik leksikografi atas nomina person berorientasi gender dalam Kamus Besa Bahasa Indonesia Edisi IV. Tesis. Depok: UI. <https://www.lib.ui.ac.id%2Ffile%3Ffile%3Ddigital%2F20313850-T31267-Praktik%2520leksikografi.pdf> (diunduh 16 Juni 2016)
- <https://id.wikipedia.org/w/index.php?title=Istimewa:Pencarian&profile=default&fulltext=Search&search>

RUMUSAN
SEMINAR LEKSIKOGRAFI INDONESIA
Jakarta, 26—29 Juli 2016

Dengan rahmat Tuhan Yang Maha Esa, setelah mendengarkan dan memperhatikan sambutan Kepala Badan Pengembangan dan Pembinaan Bahasa, paparan pembicara utama dan pemakalah pleno, serta tanya jawab dan diskusi yang berkembang selama sidang, Seminar merekomendasikan hal-hal, sebagai berikut.

1. Bahasa-bahasa daerah di Indonesia perlu didokumentasikan dalam bentuk kamus.
2. Badan Pengembangan dan Pembinaan Bahasa perlu menyediakan pedoman penyusunan kamus untuk memudahkan dan mempercepat proses dokumentasi.
3. Proses penyusunan kamus dapat melibatkan pihak-pihak di luar Badan Pengembangan dan Pembinaan Bahasa.
4. Badan Pengembangan dan Pembinaan Bahasa melalui Balai/Kantor Bahasa di tiap-tiap provinsi memberikan fasilitasi ke berbagai pihak dalam proses penyusunan kamus.
5. Asosiasi leksikograf di Indonesia perlu didirikan untuk mengakomodasi aspirasi masyarakat pencinta kamus.
6. Seminar leksikografi menjadi agenda tahunan yang dilaksanakan oleh Badan Pengembangan dan Pembinaan Bahasa.

Jakarta, 29 Juli 2016

Tim Perumus

- Ketua : Dr. Wahya, M.Hum. (Universitas Padjadjaran)
- Anggota :
1. Dr. Dora Amalia (Badan Bahasa)
 2. Deny Arnos Kwary, Ph.D. (Universitas Airlangga)
 3. Totok Suhardijanto, M.Hum., Ph.D. (Universitas Indonesia)
 4. Nany Setyono Kurnia, Ph.D. (Unika Atma Jaya)
 5. Prihantoro (Universitas Diponegoro)
 6. Achmad Atho'illah, S.S. (Pondok Pesantren Nurul Huda)
 7. Azhari Dasman Darnis, M.Hum. (Badan Bahasa)
 8. Adi Budiwiyanto, M.Hum. (Badan Bahasa)
 9. Asep Rahmat Hidayat, M.Hum. (Balai Bahasa Jawa Barat)