

## **BAB II**

### **KAJIAN PUSTAKA**

#### **2.1 Semantik**

Semantik adalah bagian dari struktur bahasa yang berhubungan dengan makna ungkapan dan dengan struktur makna suatu wicara. Setiap jenis penelitian yang berhubungan dengan bahasa, apakah struktur kalimat, kosakata, ataupun bunyi-bunyi bahasa, pada hakikatnya tidak terlepas dari makna (Sutedi, 2014). Dalam kajian semantik Jepang, menurut Sutedi (2014) semantik (*imiron*) merupakan salah satu cabang linguistik (*gengogaku*) yang mengkaji tentang makna. Lebih lanjut, Chaer (2013) pula mengatakan bahwa semantik adalah salah satu dari tiga tataran analisis bahasa (fonologi, gramatikal dan semantik) yang fokus mempelajari makna atau arti dalam bahasa. Suwandi (2008) menjelaskan bahwa semantik menelaah lambang-lambang atau tanda-tanda yang menyatakan makna, hubungan makna yang satu dengan makna yang lain dan pengaruhnya terhadap manusia dan masyarakat. Terdapat beberapa objek kajian semantik dalam bahasa Jepang yang dapat dianalisis antara lain makna kata (*go no imi*), relasi makna (*go no imi kankei*) antara satu kata dengan kata yang lainnya, makna frase (*ku no imi*), dan makna kalimat (*bun no imi*).

#### **2.2 Sinonim**

Sinonim adalah dua kata atau lebih yang memiliki makna sama. Sinonim dalam bahasa Jepang disebut *ruigigo*. Sudjianto dan Dahidi (2014) mengatakan

bahwa *ruigigo* adalah beberapa kata yang memiliki bunyi ucapan yang berbeda namun memiliki makna yang sangat mirip. Selain itu, Eriyanti dkk. (2020) pula mengatakan bahwa sinonim secara sederhana dikenal dengan sebutan persamaan makna yang berarti meskipun dua kata memiliki arti yang berbeda, namun dari segi makna adalah sama. Seperti nomina *joutai*, *jitai* dan *jijou*, ketiganya memiliki makna yang mirip namun secara ucapan dan penulisannya berbeda. Menurut Puspita (2016), sinonim terbagi menjadi tiga jenis yaitu:

a. Sinonim mutlak

Sinonim ini dapat bertukar tempat dalam konteks kebahasaan apapun tanpa mengubah makna struktural dan makna leksikal dalam rangkaian kata, frasa maupun kalimat. Contohnya seperti kosmetik dengan alat kecantikan.

b. Sinonim semirip

Sinonim ini dapat bertukar tempat dalam konteks kebahasaan tertentu tanpa mengubah makna struktural dan makna leksikal dalam rangkaian kata, frasa maupun kalimat. Contohnya seperti lahiriah dengan jasmaniah.

c. Sinonim selingkung

Sinonim ini dapat saling menggantikan dalam konteks kalimat kalimat tertentu tanpa mengubah makna struktural dan leksikal. Contohnya seperti letih dengan lesu.

Palmer (1981) berpendapat bahwa perbedaan makna pada pasangan sinonim disebabkan oleh lima kemungkinan, yaitu (1) terdapat pengaruh dialek atau adanya kebiasaan setempat; (2) terdapat perbedaan dalam pemakaian kata; (3) adanya perbedaan pada nilai kata; (4) terdapat perbedaan berdasarkan kolokialnya;

dan (5) terdapat perbedaan karena hiponimi. Chaer (2013) menyatakan bahwa ketidaksamaan makna kata yang bersinonim dipengaruhi oleh enam faktor, yaitu (1) faktor waktu; (2) faktor tempat atau wilayah; (3) faktor keformalan; (4) faktor sosial; (5) faktor bidang kegiatan; dan (6) faktor nuansa makna. Berdasarkan pernyataan di atas, analisis perbedaan makna kata yang termasuk dalam pasangan sinonim harus didasarkan pada berbagai komponen pembeda makna tersebut.

### **2.3 Sintaksis**

Sintaksis merupakan ilmu yang mempelajari struktur antar kata, atau pembahasan kata-kata dalam struktur satuan linguistik yang lebih besar dari kata, yaitu frase, klausa, dan kalimat (Chaer, 2014). Sintaksis memiliki sistematis berupa kategori, fungsi, dan peran. Struktur seperti predikat, subjek, objek, keterangan, dan sebagainya merupakan fungsi sintaksis. Fungsi sintaksis dapat menunjukkan unsur-unsur yang terdapat dalam suatu kalimat karena dapat mendeskripsikan konstruksi atau struktur pada suatu kalimat (klausa) dengan rinci. Verba, nomina, adjektiva, adverbial, dan sebagainya termasuk kategori sintaksis. Kategori sintaksis dapat digunakan untuk mendeskripsikan kategori apa yang dapat digunakan mengisi setiap fungsi dalam kerangka kalimat yang berupa argumen dan predikat dalam kalimat yang dianalisis (Sutedi, 2014).

### **2.4 Kolokasi**

Definisi kolokasi menurut Baker (1997) adalah kecenderungan sejumlah kata untuk bergabung secara teratur dalam suatu bahasa, tetapi kata yang mana dapat

berkolokasi dengan kata apa tidak ada hubungannya secara logis. Yule (2017) pula mengatakan bahwa kolokasi adalah sekelompok kata yang muncul secara bersamaan. Mongkolchai (2008) mendefinisikan kolokasi sebagai kemunculan bersama pada unit leksikal dalam suatu kombinasi dengan perbedaan frekuensi serta penerimaannya. Penggabungan kata dapat terjadi secara tidak sengaja atau berdasarkan pola, pertama disebut sebagai prinsip pilihan terbuka dan yang kedua sebagai prinsip idiom (Sinclair, 1991). Dengan kata lain, kolokasi adalah gabungan dari kata atau frasa dengan kata atau frasa lain dimana terdengar wajar oleh penutur bahasa tersebut. Contohnya seperti saat menyebut air bersih yang dapat diminum, apakah disebut air putih, air bening, air tawar atau air mineral. Tentunya setiap bahasa memiliki kolokasi termasuk bahasa Jepang. Seperti kata *manga* yang banyak disandingkan diawal kata dengan *shoujo*, *koma*, *shounen* dan lain-lain bila sandingannya dengan nomina. Lain hal dengan *shousetsu* yang banyak disandingkan dengan kata *tantei*, *suiri*, *chouhen* dan lain-lain. Meski kata *manga* dan *shousetsu* sama-sama tergolong buku bacaan fiksi, namun kata sandingan yang sering digunakannya berbeda. Kolokasi memiliki peranan penting dalam proses pembelajaran bahasa Jepang. Dengan mempelajari kolokasi, pembelajar dapat membedakan kombinasi kata yang cocok untuk suatu frasa.

Dalam penelitian bahasa Inggris, berbagai rentang telah dianalisis, seperti tiga kata “*content word*” pada kiri-kanan, empat kata kiri-kanan, dan dua kata yang berdekatan (Hori, 2009), tetapi secara umum diterima bahwa kolokasi yang penting lebih banyak ditemukan dalam rentang empat kata (Stubbs, 2002). Di sisi lain, dalam penelitian di Jepang, pedoman ini tidak selalu ditetapkan. Dalam pencarian

kata kolokasi, rentang kolokasi yang lebih luas digunakan untuk menyelidiki lingkungan semantik umum tempat kata tersebut muncul, sementara rentang yang lebih sempit digunakan untuk menyelidiki lingkungan sintaksis lokal tempat kata tersebut muncul, seperti sebelum dan sesudah kata tersebut (Shin'ichiro, 2012).

## 2.5 Korpus

Korpus merupakan kumpulan teks autentik, baik teks tulis maupun transkrip percakapan dalam jumlah besar yang disimpan secara elektronik. Menurut Setiawan (2017) korpus merupakan kumpulan tulisan yang ditulis oleh seseorang baik berupa *hard copy* dan *soft copy*. Sementara menurut Isam dkk (2014), Korpus merupakan kumpulan data dari berbagai genre yang menunjukkan bahasa sebenarnya itu digunakan, disimpan secara sistematis dalam bentuk digital yang dapat diolah untuk memperoleh berbagai jawaban mengenai kebahasaan seperti struktur bahasa dan tata bahasa. Linguistik korpus menurut Biber & Reppen (2015) yaitu pendekatan penelitian yang memfasilitasi penyelidikan empiris terhadap variasi dan penggunaan bahasa, yang menghasilkan temuan penelitian yang memiliki kemampuan generalisasi dan validitas yang jauh lebih besar daripada yang seharusnya. Penelitian berbasis korpus bertujuan untuk mendapatkan deskripsi bahasa empiris baru dari sebuah papan kosong dengan menggunakan korpus atau teks itu sendiri sebagai titik awal dan bukan melalui model, teori atau hipotesis yang sudah ada sebelumnya (Shin'ichiro, 2012).

Korpus yang digunakan dalam penelitian ini adalah NINJAL-LWP untuk BCCWJ (disingkat NLB), *Balanced Corpus of Contemporary Written Japanese* (BCCWJ) yang dibangun oleh *National Institute for Japanese Language and*

*Linguistics* (NINJAL). Korpus ini adalah sistem pencarian *online* yang dikembangkan bersama oleh Institut Nasional untuk Bahasa dan Linguistik Jepang dan Institut Linguistik Lago. Baru-baru ini, berbagai alat korpus telah dikembangkan dan membantu mengidentifikasi beragam fitur linguistik seperti kolokasi, pola tata bahasa, maupun kata kunci (Leńko-Szymańska & Boulton, 2015). Korpus sudah sering digunakan dalam berbagai macam area penelitian linguistik termasuk penerjemahan, linguistik forensik, analisis wacana, dan pengajaran. Dalam ranah pendidikan, korpus bisa menjadi sumber yang menyediakan data deskriptif tentang bagaimana bahasa digunakan yang dapat dimanfaatkan oleh pengajar atau pelajar sebagai media untuk mencari bukti atau konfirmasi intuisi mereka tentang pengetahuan sebuah bahasa. Salah satu penggunaan korpus dalam dunia pendidikan yaitu untuk mempelajari padanan kata atau sinonim.

### **2.5.1 Skor MI**

Skor MI atau kandungan informasi timbal balik adalah indikator yang menjelaskan seberapa banyak informasi yang dimiliki sebuah kata tentang pasangan kemunculannya. Informasi timbal balik memiliki skor tinggi bila pasangan *co-occurrence* bersama ditentukan secara otomatis saat kata itu muncul. Secara konseptual, kandungan informasi timbal balik adalah rasio yang diinduksi secara logaritmik dari frekuensi yang diukur dan frekuensi yang diharapkan dari dua kata. Skor MI dihitung oleh sistem dengan rumusnya adalah sebagai berikut (Shin'ichiro, 2008)

$$I = \log_2 \frac{\text{frekuensi kemunculan bersama} \times \text{total kata pada korpus}}{\text{frekuensi kata pusat} \times \text{kemunculan frekuensi kata}}$$

### 2.5.2 Koefisien LD

Koefisien LD juga digunakan secara luas dalam teori informasi dan sebagai algoritma untuk mengekstraksi kata terkait secara otomatis dari situs web. Jumlah total kata dalam korpus tidak diperlukan untuk penghitungan koefisien LD, dan kekuatan kolokasi dua kata hanya diukur dengan hubungan antara frekuensi kata pusat dan frekuensi kata yang muncul bersamaan (*User Manual* pada korpus NLB). Koefisien LD adalah skor yang diperoleh dengan membagi frekuensi kemunculan bersama dengan jumlah frekuensi kata pusat dan frekuensi kata yang muncul bersamaan dan mengalikannya dengan dua. Skor LD dihitung oleh sistem dengan rumusnya adalah sebagai berikut (Shin'ichiro, 2008).

$$D = 2 \times \frac{\text{frekuensi kemunculan bersama}}{\text{frekuensi kata pusat} \times \text{kemunculan frekuensi kata}}$$

Seperti yang dapat dilihat dari rumusnya, koefisien LD adalah indikator yang sederhana namun sangat berguna dan valid. Kiyomi dan Masao (2004), yang membandingkan validitas sembilan jenis nilai statistik dalam deteksi kata fitur, melaporkan bahwa akurasi koefisien LD adalah yang terbaik.

### 2.5.3 *Concordancer*

Istilah *concordancer* pada awalnya merujuk pada kumpulan contoh kalimat berbasis kertas yang digunakan oleh penulis tertentu, tetapi saat ini istilah tersebut merujuk pada perangkat lunak yang secara otomatis membuat indeks contoh dari korpus (Shin'ichiro, 2012). Sementara konkordansi adalah daftar kata-kata yang ada pada sebuah teks yang biasanya disertai kutipan dimana kata itu berasal. Konkordansi telah dikembangkan dalam pengembangan linguistik korpus. McEnery & Hardie (2012) menjelaskan generasi pertama adalah konkordansi komputer besar yang dikembangkan pada awal linguistik korpus yaitu akhir tahun 1970-an, dan generasi kedua konkordansi komputer pribadi yang dikembangkan pada akhir tahun 1980-an. Konkordansi komputer pribadi yang dikembangkan pada akhir tahun 1980-an disebut sebagai generasi kedua, kemudian konkordansi multi-fungsi yang dikembangkan sejak akhir tahun 1990-an sebagai generasi ketiga, dan antarmuka pencarian online yang saat ini sedang dikembangkan termasuk generasi keempat. Menurut Shin'ichiro (2012), ada dua jenis konkordansi yang digunakan saat ini, yaitu konkordansi fungsi tunggal yang mengkhususkan diri pada pencarian konkordansi untuk mengekstrak contoh, dan konkordansi multi-fungsi yang mengemas berbagai jenis pencarian yang umum digunakan dalam bahasa Inggris dan yang pertama dalam bahasa Jepang.

Konkordansi adalah alat bantu dasar bagi siapa saja yang bekerja dengan korpus. Alat ini menunjukkan kepada pengguna apa yang ada di

dalam korpus. Alat ini membawa pengguna ke *raw data*, yang mendasari analisis apa pun. Mencapai ke sana dari *word sketch* hanyalah salah satu cara untuk mencapai konkordansi (Killgarriff dkk, 2014). Tabel frekuensi pada konkordansi tersebut dapat digunakan dalam berbagai cara sebagai kompresi informasi kolokasi yang efisien, tetapi karena tabel tersebut dibuat secara otomatis, tabel tersebut sering kali berisi kata-kata yang hanya muncul pada posisi tersebut secara acak, terutama pada *co-occurrence* atau kemunculan bersama yang jauh dari kata pusat. Karena alasan ini, dalam linguistik korpus, analisis kemunculan bersama umumnya dilakukan dalam kombinasi dengan analisis konkordansi, yang memungkinkan pengamatan langsung terhadap data tekstual, bukan secara independen. Secara khusus, analisis kemunculan bersama digunakan untuk mengidentifikasi kata-kata penting yang muncul bersama dan untuk mengekstrak garis konkordansi yang berisi kata utama dan kata yang muncul bersama yang bersangkutan pada saat yang sama, untuk memeriksa apakah kemunculan bersama kedua kata tersebut benar-benar berhubungan atau tidak. Perlu dicatat bahwa tabel frekuensi adalah sarana, bukan tujuan, dari interpretasi korpus (Shin'ichiro, 2012).