



JURNAL LINGUISTIK Vol. 20 (2) Disember 2016 (001-012)

Perbandingan antara Korpus Nama Ikan Asal dengan Korpus Nama Ikan yang Distandardkan

Muhammad Nasri Rifin & Hasnah Mohamad

muhammadnasri9217@gmail.com, hasnah_m@upm.edu.my

Fakulti Bahasa Moden dan Komunikasi, Universiti Putra Malaysia

Abstrak

Perubahan yang berlaku pada sesuatu korpus disebabkan oleh proses penyeragaman atau penstandardan berbeza-beza mengikut jenis korpus, saiz korpus dan bahasa sesuatu korpus. Corak perubahan ini boleh dilihat berdasarkan perbandingan di antara korpus yang asal dengan korpus yang telah distandardkan. Dalam kajian ini, fokus diberikan kepada pengelasan dan perbandingan di antara kekerapan kata dalam korpus nama biasa ikan (KNBI) yang asal dengan kekerapan kata dalam KNBI yang telah distandardkan. Kekerapan kata yang dibandingkan dalam kajian ini terbahagi kepada aspek jumlah kekerapan *type* mengikut golongan kata, peratus kekerapan *type* mengikut golongan kata, jumlah kekerapan *token* mengikut golongan kata dan peratus kekerapan *token* mengikut golongan kata. Dalam meneliti perbandingan ini, dua korpus pilihan dan korpus tiada pertindihan telah dihasilkan. Kedua-dua korpus ini telah dibina daripada senarai nama ikan tempatan dan senarai nama ikan sahah dalam buku Ikan Laut Malaysia yang diselenggarakan oleh Yusri, Hamdan dan Rahman pada 2010. Setelah itu, kata dan kekerapan kata dalam kedua-dua korpus ini dikelaskan dan dibandingkan berdasarkan pendekatan *corpus-based* dan teori tatabahasa Melayu, iaitu *Tatabahasa Dewan*. Hasil kajian menunjukkan bahawa terdapat jumlah dan peratus kekerapan kata dalam KNBI asal yang lebih tinggi berbanding jumlah dan peratus kekerapan kata dalam KNBI yang telah distandardkan. Pada masa yang sama, hasil kajian ini juga menunjukkan bahawa terdapat jumlah dan peratus kekerapan kata dalam KNBI asal yang lebih rendah berbanding jumlah dan peratus kekerapan kata dalam KNBI yang telah distandardkan. Secara umumnya, hasil kajian ini telah dapat memperjelas kesan dan hasil daripada proses penyeragaman yang telah dilakukan.

Kata kunci: korpus asal, korpus yang telah distandardkan, nama biasa ikan, *type*, *token*

Abstract

The changes that occurred in a corpus due process of harmonization or standardization vary according to the type of corpus, the size of a corpus and language of corpus. The pattern of these changes can be seen by comparison between the original corpus and standardized corpus. In this study, focuses was given on the classification and comparison between the frequency of words in the original corpus of fish common names (KNBI) with the frequency of words in the standardized KNBI. Words frequency that were compared in this study is divided into the several aspect such as the total of type frequency by word class, the percentage of type frequency by word class, the total of token frequency by word class and the percentage of token frequency by word class. In this comparative study, two selected and non-redundant corpus was built. Both of the corpus was built from local fish name list and the standard fish name list in a book named Ikan Laut Malaysia organized by Yusri, Hamdan and Rahman on 2010. After that, words and words frequency in both of the corpus is classified and compared based on the corpus-based approach and Malay grammar theory, know as Tatabahasa Dewan. The results showed that the total and percentage of frequency of words in the original KNBI is higher than the total and percentage of frequency of words in KNBI that have been standardized. At the same time, the results also showed that the another total and percentage of frequency of words in the original KNBI is lower than the

total and percentage of frequency of words in KNBI that have been standardized. Generally, the results of this study was able to demonstrate and clarify the effects and results of the standardization that was conducted

Keywords: *original corpus, standardized corpus, common name of fish, type, token*

1. Pengenalan

Nama biasa merupakan frasa nama atau istilah yang lazim digunakan oleh masyarakat umum dalam merujuk sesuatu organisma hidup. Misalnya, nama ‘gombing ekor putih’ merupakan nama biasa ikan yang digunakan oleh masyarakat umum dalam merujuk kepada spesies *Chromis leucura* dan *Stegastes leucurus*. Nama biasa ini lazim digunakan oleh masyarakat umum kerana penggunaan nama ini dapat memberikan banyak kelebihan (Gibb, 2014). Meskipun begitu, dalam konteks tertentu, menurut Claborn (2010) penggunaan nama biasa ini juga dapat mewujudkan masalah seperti pertindihan, kekaburan serta perselisihan makna dan maklumat. Sehubungan itu, bagi mengatasi masalah dan kelemahan ini, usaha penyeragaman dan penstandardan nama-nama biasa ini, khususnya nama-nama biasa ikan laut di sekitar perairan Malaysia telah dilakukan.

Usaha ini dapat meningkatkan ketepatan makna dan maklumat, namun usaha ini juga boleh mendatangkan hasil dan kesan sampingan. Hasil dan kesan sampingan ini dapat diperjelas menerusi aspek-aspek linguistik yang dapat diterokai melalui data korpus bahasa. Menurut Rozaimah (2015), melalui data korpus bahasa, pelbagai aspek linguistik dapat diterokai. Antara aspek linguistik ini ialah aspek perbandingan antara jumlah kekerapan kata dan peratus kekerapan kata dalam koleksi pilihan atau korpus nama biasa ikan (KNBI) yang asal dengan KNBI yang telah distandardkan. Perbandingan jumlah kekerapan kata mengikut golongan kata antara KNBI yang asal dengan KNBI yang telah distandardkan dapat memperlihatkan perubahan dalam konteks saiz maklumat yang terkandung dalam KNBI bahasa Melayu. Sementara itu, perbandingan peratus kekerapan kata mengikut golongan kata antara KNBI yang asal dengan KNBI yang telah distandardkan pula dapat memperlihatkan perubahan dari sudut imbalan maklumat dalam KNBI bahasa Melayu.

Maklumat dan rujukan tentang kedua-dua aspek perubahan ini sangat penting dalam memberi penjelasan tentang hasil dan kesan sampingan daripada usaha penyeragaman dan penstandardan yang dilakukan, namun maklumat dan rujukan tentang aspek-aspek perubahan ini masih kurang dan belum cukup terperinci. Oleh itu, bagi mengisi lompang tersebut, kajian ini telah dirangka bagi mengelaskan dan membandingkan kekerapan kata dan peratus kekerapan kata mengikut golongan kata yang terdapat dalam KNBI yang asal dengan KNBI yang telah distandardkan. Pengelasan dan perbandingan kekerapan kata dan peratus kekerapan kata ini mampu menyediakan maklumat dan rujukan yang lebih banyak dan terperinci mengenai perubahan saiz maklumat dan perubahan imbalan yang dapat memperjelas kesan dan hasil daripada proses penyeragaman dan penstandardan yang telah dilakukan.

2. Kajian Lepas

Dalam konteks perbandingan di antara dua korpus yang sama tetapi berbeza dari sudut perkembangan mengikut masa, kajian Saily, Nurmi dan Sairio (2013) menunjukkan bahawa untuk sesuatu istilah, korpus yang terdahulu mempunyai jumlah *token* yang lebih tinggi berbanding jumlah *token* yang terdapat dalam korpus yang baharu. Hal ini telah dibuktikan oleh Saily, Nurmi dan Sairio (2013) dalam kajian mereka yang menunjukkan bahawa korpus gabungan *Corpus of Early English Correspondence* (CEEC) dengan *Corpus of Early English Correspondence Extension* (CEECE) tahun 1680 hingga 1719 mempunyai jumlah *token* ‘do’ yang lebih tinggi daripada jumlah *token* ‘do’ yang terdapat dalam korpus gabungan CEEC dengan CEECE tahun 1720 hingga 1759 dan korpus gabungan CEEC dengan CEECE tahun 1760 hingga 1800.

Pada masa yang sama, perbandingan di antara dua korpus yang sama tetapi berbeza dari sudut perkembangan mengikut masa dalam kajian Saily, Nurmi dan Sairio (2013) juga menunjukkan bahawa

korpus yang terdahulu mempunyai jumlah *token* yang lebih rendah berbanding jumlah *token* yang terdapat dalam korpus yang baru. Hal ini telah dibuktikan berdasarkan korpus gabungan *Corpus of Early English Correspondence* (CEEC) dengan *Corpus of Early English Correspondence Extension* (CEECE) tahun 1720 hingga 1759 yang mempunyai jumlah *token* 'do' yang lebih rendah daripada jumlah *token* 'do' dalam korpus gabungan CEEC dengan CEECE tahun 1760 hingga 1800.

Sementara itu, dalam konteks perbandingan di antara dua korpus yang sama jenis, kajian Johnson dan Ensslin (2006) menunjukkan meskipun sesuatu korpus mempunyai bilangan *token* yang lebih tinggi untuk sesuatu kumpulan istilah, namun korpus tersebut juga mempunyai bilangan *token* yang lebih rendah untuk sesuatu kumpulan istilah yang lain. Misalnya, meskipun korpus *The Times* mempunyai bilangan *token* (199 *token*) yang lebih tinggi berbanding bilangan *token* (177 *token*) dalam korpus *The Guardian* untuk kumpulan istilah 'linguistic', namun korpus *The Times* mempunyai bilangan *token* (3501 *token*) yang lebih rendah berbanding bilangan *token* (4381 *token*) dalam korpus *The Guardian* untuk kumpulan istilah 'language'. Seterusnya, perbandingan di antara korpus asal dengan korpus yang distandardkan berdasarkan kamus dalam kajian Stevenson dan Gaizauskas (2000) juga membuktikan bahawa korpus yang asal memiliki bilangan unit bahasa yang lebih tinggi daripada bilangan unit bahasa yang terdapat dalam korpus yang telah distandardkan. Hal dapat dilihat menerusi korpus *The Limes* asal yang mempunyai bilangan kata yang lebih tinggi, iaitu sebanyak 2157 *token* dalam frasa nama organisasi, 3947 *token* dalam frasa nama orang dan 1489 *token* dalam frasa nama tempat berbanding korpus *The Limes* standard yang mempunyai bilangan kata sebanyak 1978 *token* dalam frasa nama organisasi, 3769 *token* dalam frasa nama orang dan 1412 *token* dalam frasa nama tempat.

Kajian Collin dan Hakala (2011) yang melibatkan perbandingan antara *Corpus of Early English Correspondence* (CEEC) dan korpus *Early Modern English Medical Texts* (EMEMT) yang asal dengan CEEC dan korpus EMEMT yang telah distandardkan menunjukkan bahawa korpus asal memiliki peratus bilangan *token* yang lebih tinggi daripada peratus bilangan *token* yang terdapat dalam korpus yang telah distandardkan. Hal ini telah terbukti berdasarkan peratus bilangan *token* yang terdapat dalam CEEC yang asal, iaitu sebanyak 18.2% dan korpus EMEMT yang asal, iaitu sebanyak 11.5 % yang lebih tinggi daripada peratus bilangan *token* yang terdapat dalam CEEC yang telah distandardkan, iaitu sebanyak 6.7% dan EMEMT yang telah distandardkan, iaitu sebanyak 4.2%.

Perbandingan antara CEEC dan korpus EMEMT yang asal dengan CEEC dan korpus EMEMT yang telah distandardkan dalam kajian Collin dan Hakala (2011) juga menunjukkan bahawa korpus yang asal memiliki peratus bilangan *type* yang lebih tinggi daripada peratus bilangan *type* yang terdapat dalam korpus yang telah distandardkan. Hal ini telah terbukti berdasarkan peratus bilangan *type* yang terdapat dalam CEEC asal, iaitu sebanyak 56.7% dan korpus EMEMT asal, iaitu sebanyak 26.4 % yang lebih tinggi daripada peratus bilangan *type* yang terdapat dalam CEEC yang telah diseragamkan, iaitu sebanyak 33.8% dan korpus EMEMT yang standard, iaitu sebanyak 12.2%.

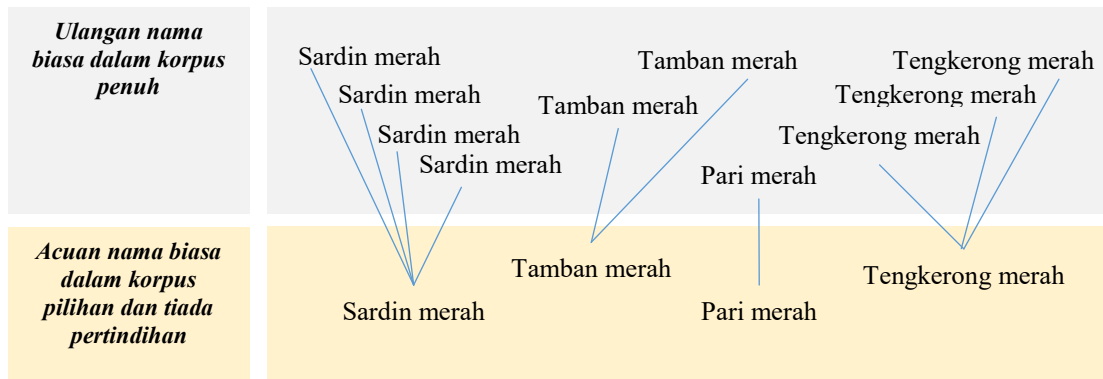
Kajian Singnoi (2011) yang berfokus kepada perbandingan kekerapan *token* dalam korpus nama biasa tumbuhan Thailand mengikut golongan kata menunjukkan bahawa peratus *token* kata nama pada teras nama adalah lebih tinggi daripada peratus *token* kata nama pada penerang nama. Hal ini terbukti berdasarkan analisis yang menunjukkan bahawa peratus *token* kata nama pada teras nama biasa tumbuhan Thailand ialah sebanyak 63%, sedangkan peratus *token* kata nama pada penerang nama biasa tumbuhan Thailand ialah sebanyak 15%.

Sementara itu, kajian Kos (2011) mengenai perbandingan di antara korpus penerang nama biasa burung British yang asal dengan korpus penerang nama biasa burung British yang standard menunjukkan bahawa bilangan *type* kata penerang dalam korpus asal, iaitu sebanyak 130 *type* adalah lebih rendah berbanding bilangan *type* kata penerang dalam korpus yang telah distandardkan, iaitu sebanyak 291 *type*. Kajian Kos (2011) ini juga menunjukkan bahawa bilangan *token* kata penerang dalam korpus asal, iaitu sebanyak 174 *token* adalah lebih rendah berbanding bilangan *token* kata penerang dalam korpus yang telah distandardkan, iaitu sebanyak 424 *token*.

Kajian-kajian lepas ini secara keseluruhannya dapat menunjukkan dapatan hasil perbandingan di antara dua korpus. Kajian-kajian ini juga dapat menunjukkan bahawa banyak aspek yang boleh dinilai dalam perbandingan kekerapan kata di antara dua korpus. Antaranya ialah aspek jumlah kekerapan *type* mengikut golongan kata, jumlah kekerapan *token* mengikut golongan kata, peratus kekerapan *type* mengikut golongan kata dan peratus kekerapan *token* mengikut golongan kata.

3. Metodologi

Korpus asal dan korpus standard yang dibandingkan dalam kajian ini adalah jenis korpus pilihan (*selected corpus*) dan korpus tanpa pertindihan (*non-redundant corpus*). Korpus pilihan mengandungi potongan-potongan unit tatabahasa yang dipilih berdasarkan sesuatu fungsi atau tujuan, sementara korpus tiada pertindihan pula mengandungi potongan-potongan unit tatabahasa yang tidak bertindih antara satu sama lain. Contoh bagi korpus pilihan dan korpus tanpa pertindihan ini ialah korpus istilah Kejuruteraan dan Sains dan Teknologi (S&T), DBP yang digunakan dalam kajian Junaini et. al. (2017) dan korpus NAMEDAT yang dibangunkan oleh Freire dan Filho (2009) bagi meneliti pengelasan kata dalam acuan nama ikan. Korpus pilihan dan korpus tanpa pertindihan ini berbeza daripada korpus teks penuh, misalnya korpus Dewan Bahasa dan Pustaka, di bawah subkorpus akhbar (Berita Harian dan Harakah), Buku (DB3), dan Majalah Bukan Ilmiah yang telah digunakan dalam kajian Hishamunid dan Norsimah (2012). Perbezaan antara korpus penuh dengan korpus yang digunakan dalam kajian ini boleh dilihat dengan lebih jelas berdasarkan rajah 1.1.

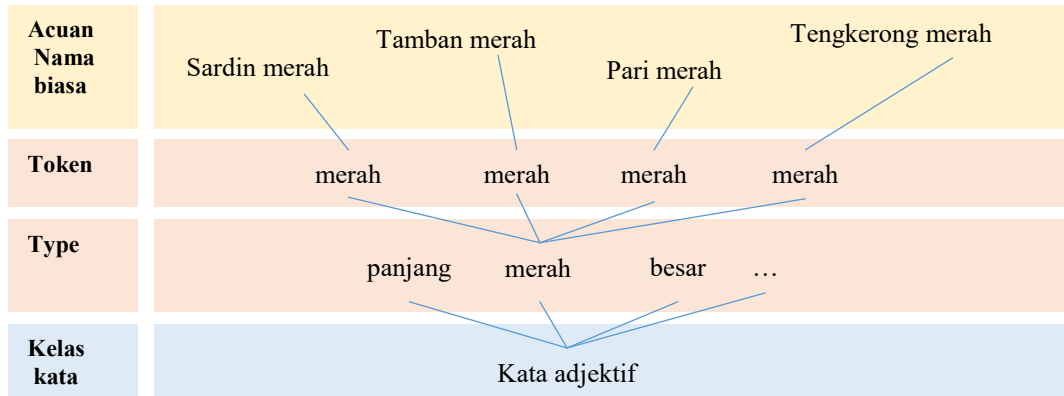


Rajah 1.1. Bentuk ulangan nama dan acuan nama

Jelasnya, korpus asal dan korpus standard yang digunakan dalam kajian ini mengandungi senarai istilah atau frasa-frasa nama biasa ikan yang setiap daripadanya terdiri daripada satu wakil atau acuan sahaja. Penggunaan korpus yang mempunyai ciri-ciri korpus pilihan dan korpus tanpa pertindihan dalam kajian ini bertujuan untuk melihat corak dan kekerapan penggunaan dalam pembinaan acuan istilah atau acuan nama biasa ikan, bukan melihat corak dan kekerapan penggunaan kata pada setiap satu istilah atau frasa nama biasa ikan dalam satu kumpulan teks.

Sehubungan itu, korpus asal dalam kajian ini telah dibina daripada indeks nama tempatan dan korpus yang telah distandardkan pula dibina daripada indeks nama sahah yang terdapat dalam buku panduan Ikan Laut Malaysia yang diselenggarakan oleh Yusri, Hamdan dan Rahman (2010) di bawah Dewan Bahasa dan Pustaka. Pembinaan kedua-dua korpus ini telah berjaya menghimpunkan sebanyak 1223 istilah atau frasa nama biasa ikan yang asal dan sebanyak 1629 istilah atau frasa nama biasa ikan yang telah distandardkan. Dalam mencapai objektif kajian ini, pengkaji telah menggunakan pendekatan *corpus-based* dan *Tatabahasa Dewan* sebagai sandaran. Berdasarkan prinsip dalam pendekatan *corpus-based*, hubungan antara data dengan teori diutamakan dalam kajian ini. Oleh itu, pengelasan data atau kata dalam kajian ini telah dilakukan berdasarkan teori tatabahasa sedia ada, iaitu *Tatabahasa Dewan*. Menurut *Tatabahasa Dewan*

(Nik Safiah et. al, 2014), kata-kata dalam bahasa Melayu dapat digolongkan kepada empat golongan utama, iaitu golongan kata nama, kata kerja, kata adjektif dan kata tugas. Oleh itu, kata-kata yang terdapat dalam kedua-dua korpus yang dibandingkan dalam kajian ini dikelaskan kepada golongan kata nama, kata kerja, kata adjektif dan kata tugas.



Rajah 1.2. Bentuk acuan nama biasa, token, type dan kelas kata

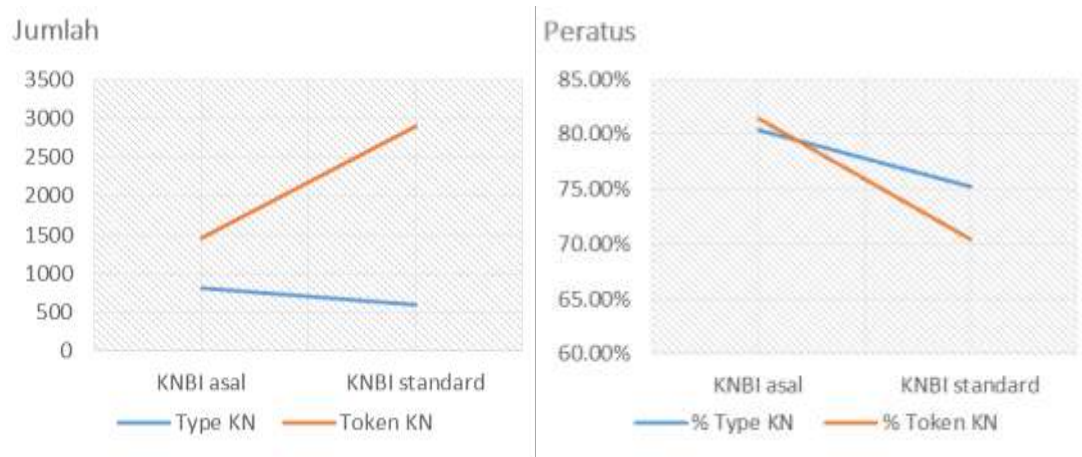
Dalam usaha membuat pengelasan kata secara sistematik, kaedah pengkodan digunakan. Setelah pengelasan dilakukan, kekerapan kata atau lebih khusus lagi kekerapan *type* dan kekerapan *token* dalam nama biasa ikan bagi kedua-dua korpus yang diteliti ditentukan dengan menggunakan perisian *Word Smith* secara interaktif. Istilah *type* ini merujuk kepada jenis kata berbeza, manakala istilah *token* pula merujuk kepada pengulangan kata (Imran, 2011). Hal ini boleh dirujuk berdasarkan rajah 1.2. Seterusnya, untuk membandingkan jumlah kekerapan dan peratus kekerapan *type* dan *token* dalam korpus asal dengan *type* dan *token* dalam korpus yang telah distandardkan, kaedah statistik deskriptif telah digunakan. Melalui kaedah ini, perbandingan data iaitu *type* dan *token* dilakukan dengan menggunakan digit dan grafik. Dalam kajian ini, digit yang digunakan terdiri daripada angka yang merujuk kepada jumlah dan peratus kekerapan, manakala grafik yang digunakan terdiri daripada graf garis.

4. Analisis dan Perbincangan

Bahagian ini memaparkan hasil pengelasan kata dan perbandingan kekerapan kata antara KNBI yang asal dengan KNBI yang telah distandardkan. Secara khusus, hasil pengelasan dan perbandingan yang dipaparkan dalam bahagian ini tertumpu kepada aspek jumlah kekerapan *type*, peratus kekerapan *type*, jumlah kekerapan *token* dan peratus kekerapan *token* mengikut golongan kata, seperti dalam rajah 2.1, rajah 2.2, rajah 2.3 dan rajah 2.4.

4.1 Kekerapan Kata Nama

Data kajian ini menunjukkan bahawa terdapat pelbagai *type* dan *token* kata nama dalam KNBI asal dan KNBI yang telah distandardkan. Antara *type* kata nama ini ialah *type* 'merak' pada nama biasa ikan 'lidah merak', *type* 'pasir' pada nama biasa ikan 'buntal pasir bersih' dan *type* 'daun' pada nama biasa ikan 'bombin daun'. *Token* kata nama pula, misalnya *token* 'daun' pada nama biasa ikan 'lidah daun tirus', 'banang bulat daun' dan 'daun belang'.



Rajah 2.1. Perbandingan kekerapan kata nama antara KNBI asal dengan KNBI standard.

Berdasarkan pengelasan *type* yang telah dilakukan, didapati jumlah kekerapan *type* kata nama dalam KNBI asal ialah 810 *type*, manakala jumlah kekerapan *type* kata nama dalam KNBI yang telah distandardkan ialah 604 *type*. Perbandingan di antara kedua-dua jumlah ini menunjukkan bahawa jumlah kekerapan *type* kata nama dalam KNBI asal adalah lebih tinggi berbanding jumlah kekerapan *type* kata nama dalam KNBI yang telah distandardkan. Perbandingan di antara kedua-dua jumlah ini juga menunjukkan bahawa proses penyelarasan nama biasa yang dilakukan menyebabkan jumlah kekerapan *type* kata nama dalam KNBI bahasa Melayu menurun sebanyak 206 *type*.

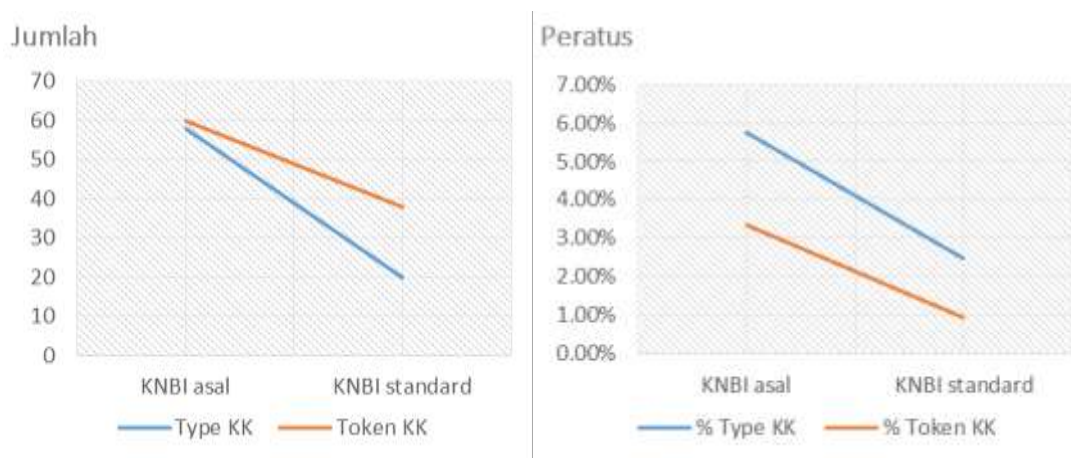
Pada masa yang sama, didapati peratus kekerapan *type* kata nama dalam KNBI asal ialah 80.36%. Peratus kekerapan *type* kata nama dalam KNBI yang telah distandardkan pula ialah 75.22%. Perbandingan kedua-dua nilai peratus ini menunjukkan bahawa peratus kekerapan *type* kata nama dalam KNBI asal adalah lebih tinggi berbanding peratus kekerapan *type* kata nama dalam KNBI yang telah distandardkan. Selain itu, perbandingan di antara kedua-dua nilai peratus ini menunjukkan bahawa peratus kekerapan *type* kata nama dalam KNBI bahasa Melayu menurun sebanyak 5.14% disebabkan oleh proses penyelarasan nama biasa yang telah dilakukan. Walau bagaimanapun, penurunan peratus kekerapan *type* kata nama ini kurang drastik berbanding penurunan peratus kekerapan *token* kata nama dalam KNBI bahasa Melayu.

Seterusnya, pengelasan *token* dalam kajian pula ini mendapati bahawa jumlah kekerapan *token* kata nama dalam KNBI asal ialah 1461 *token*, sementara jumlah kekerapan *token* kata nama dalam KNBI yang telah distandardkan ialah 2892 *token*. Perbandingan di antara kedua-dua jumlah ini menunjukkan bahawa jumlah kekerapan *token* kata nama dalam KNBI asal adalah lebih rendah berbanding jumlah kekerapan *token* kata nama dalam KNBI yang telah distandardkan. Hasil perbandingan daripada kedua-dua jumlah ini juga menunjukkan meskipun proses penyelarasan nama biasa yang dilakukan telah menyebabkan jumlah kekerapan *type* kata nama dalam KNBI bahasa Melayu menurun, namun proses ini telah menyumbang kepada peningkatan jumlah kekerapan *token* kata nama dalam KNBI bahasa Melayu sebanyak 1431 *token*. Hasil analisis dalam kajian ini juga mendapati bahawa peratus kekerapan *token* kata nama dalam KNBI asal ialah 81.48%, manakala peratus kekerapan *token* kata nama dalam KNBI yang telah distandardkan ialah 70.45%. Perbandingan di antara dua nilai peratus ini menunjukkan bahawa walaupun jumlah kekerapan *token* kata nama dalam KNBI asal adalah lebih rendah berbanding jumlah kekerapan *token* kata nama dalam KNBI yang telah distandardkan, namun peratus kekerapan *token* kata nama dalam KNBI asal adalah lebih tinggi berbanding peratus kekerapan *token* kata nama dalam KNBI yang telah distandardkan. Hal ini juga dapat menunjukkan bahawa meskipun proses penyelarasan nama biasa yang dilakukan telah menyumbang kepada peningkatan jumlah kekerapan *token* kata nama dalam KNBI bahasa Melayu, namun proses ini telah menyebabkan peratus kekerapan *token* kata nama dalam KNBI bahasa Melayu menurun sebanyak 11.03%.

Menurut Lailatul (2014), kata nama ialah kata yang menamakan sesuatu. Oleh itu, penurunan jumlah kekerapan *type*, penurunan peratus kekerapan *type* dan penurunan peratus kekerapan *token* yang terdiri daripada kata nama dalam KNBI bahasa Melayu masing-masing dapat memperjelas bahawa proses penyelarasan nama biasa yang dilakukan telah mendatangkan kesan pengecilan saiz variasi maklumat, pengurangan imbangan variasi maklumat dan pengurangan imbangan taburan maklumat tentang nama serta penamaan berkaitan spesies-spesies ikan yang dirujuk dalam KNBI bahasa Melayu. Walau bagaimanapun, proses penyelarasan nama biasa yang telah dilakukan ini turut menyumbang kepada penambahan saiz taburan maklumat tentang nama atau penamaan berkaitan dengan spesies-spesies ikan yang terkandung dalam KNBI bahasa Melayu. Hal ini dapat diperjelaskan berdasarkan peningkatan jumlah kekerapan *token* kata nama dalam KNBI bahasa Melayu.

4.2 Kekerapan Kata Kerja

KNBI asal dan KNBI standard yang dibandingkan dalam kajian ini mengandungi pelbagai *type* dan *token* kata kerja. Antaranya *type* 'lompat' yang terdapat pada nama biasa ikan 'yu lompat', *type* 'sodok' pada nama biasa ikan 'duri sodok' dan *type* 'buru' pada nama biasa ikan 'seriding buru'. *Token* kata kerja pula, misalnya *token* 'terbang' yang terdapat pada nama biasa ikan 'terbang luncur', 'belalang terbang' dan 'curut terbang'.



Rajah 2.2. Perbandingan kekerapan kata kerja antara KNBI asal dengan KNBI standard.

Berdasarkan pengelasan *type* dalam kajian ini, didapati bahawa jumlah kekerapan *type* kata kerja dalam KNBI ialah 58 *type*, sementara jumlah kekerapan *type* kata kerja dalam KNBI yang telah distandardkan ialah 20 *type*. Perbandingan di antara kedua-dua jumlah ini menunjukkan bahawa jumlah kekerapan *type* kata kerja dalam KNBI asal adalah lebih tinggi berbanding jumlah kekerapan *type* kata kerja dalam KNBI yang telah distandardkan. Hal ini juga dapat menunjukkan bahawa proses penyelarasan nama biasa yang dilakukan telah menyebabkan jumlah kekerapan *type* kata kerja dalam KNBI bahasa Melayu menurun sebanyak 38 *type*. Kadar penurunan ini lebih tinggi daripada kadar penurunan jumlah kekerapan *token* kata kerja dalam KNBI bahasa Melayu. Hal ini kerana semasa proses penyelarasan dilakukan, jumlah *type* kata kerja yang disingkirkan lebih tinggi daripada jumlah *type* kata kerja baharu yang ditambah dalam KNBI bahasa Melayu dan beberapa *type* yang kekal dalam KNBI bahasa Melayu mengalami peningkatan jumlah *token* yang sangat tinggi. Misalnya, *type* kata kerja 'terbang' yang terdiri daripada dua token dalam KNBI asal meningkat kepada 16 token dalam KNBI yang telah distandardkan.

Kajian ini juga mendapati bahawa peratus kekerapan *type* kata kerja dalam KNBI asal ialah 5.75%, manakala peratus kekerapan *type* kata kerja dalam KNBI yang telah distandardkan ialah 2.49%. Perbandingan di antara kedua-dua nilai peratus ini menunjukkan bahawa peratus kekerapan *type* kata kerja dalam KNBI asal adalah lebih tinggi berbanding peratus kekerapan *type* kata kerja dalam KNBI yang telah

distandardkan. Selisih di antara kedua-dua peratus ini juga dapat menunjukkan bahawa proses penyelarasan nama biasa yang dilakukan telah menyebabkan peratus kekerapan *type* kata kerja dalam KNBI bahasa Melayu menurun sebanyak 3.26%. Kadar penurunan peratus kekerapan *type* kata kerja ini juga lebih tinggi daripada kadar penurunan peratus kekerapan *token* kata kerja dalam KNBI bahasa Melayu sebagaimana kadar penurunan jumlah kekerapan *type* kata kerja lebih tinggi daripada kadar penurunan jumlah kekerapan *token* kata kerja dalam KNBI bahasa Melayu.

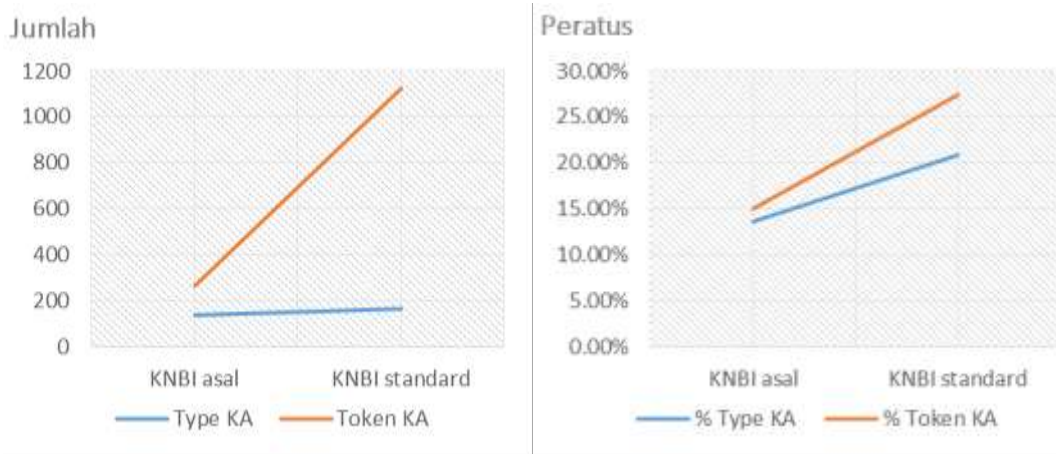
Pengelasan *token* dalam kajian ini mendapati bahawa jumlah kekerapan *token* kata kerja dalam KNBI asal ialah 60 *token*. Jumlah kekerapan *token* kata kerja dalam KNBI yang telah distandardkan pula ialah 38 *token*. Perbandingan di antara kedua-dua jumlah *token* ini menunjukkan bahawa jumlah kekerapan *token* kata kerja dalam KNBI asal adalah lebih tinggi berbanding jumlah kekerapan *token* kata kerja dalam KNBI yang telah distandardkan. Hal ini selari dengan dapatan Saily, Nurmi dan Sairio (2013) yang menunjukkan bahawa jumlah *token* bagi sesuatu himpunan kata kerja dalam korpus asal lebih tinggi daripada jumlah *token* bagi sesuatu himpunan kata kerja dalam korpus yang baharu. Selain itu, beza di antara kedua-dua jumlah kekerapan *token* ini dapat menunjukkan bahawa proses penyelarasan nama biasa yang dilakukan telah menyebabkan jumlah kekerapan *token* kata kerja dalam KNBI bahasa Melayu menurun sebanyak 22 *token*. Walau bagaimanapun, kadar penurunan jumlah kekerapan *token* kata kerja ini sangat kecil dan lebih kecil berbanding dua per 65 daripada kadar penurunan jumlah kekerapan *token* kata nama dalam KNBI bahasa Melayu.

Hasil kajian ini mendapati bahawa peratus kekerapan *token* kata kerja dalam KNBI asal ialah 3.35%, manakala peratus kekerapan *token* kata kerja dalam KNBI yang telah distandardkan ialah 0.93%. Perbandingan di antara kedua-dua nilai peratus ini menunjukkan bahawa peratus kekerapan *token* kata kerja dalam KNBI asal adalah lebih tinggi berbanding peratus kekerapan *token* kata kerja dalam KNBI yang telah distandardkan. Hasil perbandingan di antara kedua-dua nilai peratus ini juga menunjukkan bahawa proses penyelarasan nama biasa yang dilakukan telah menyebabkan peratus kekerapan *token* kata kerja dalam KNBI bahasa Melayu menurun sebanyak 2.42%. Penurunan peratus kekerapan *token* kata kerja ini agak kurang drastik berbanding penurunan peratus kekerapan *token* kata nama dalam KNBI bahasa Melayu.

Menurut Nugraha (2015), kata kerja ini ialah kata yang menyatakan tindakan dan perlakuan subjek. Berdasarkan pernyataan ini, penurunan jumlah kekerapan *type* dan penurunan peratus kekerapan *type* yang terdiri daripada kata kerja ini dapat menjelaskan bahawa proses penyeragaman yang dilakukan telah menyebabkan pengecilan saiz variasi maklumat dan pengurangan imbalan variasi maklumat tentang tindakan dan perlakuan spesies ikan atau perkara yang berkaitan dengan spesies ikan yang dirujuk dalam KNBI bahasa Melayu. Pada masa sama, penurunan jumlah kekerapan *token* dan penurunan peratus kekerapan *token* yang terdiri daripada kata kerja ini dapat menjelaskan bahawa proses penyeragaman yang dilakukan telah menyebabkan pengecilan saiz taburan maklumat dan pengurangan imbalan taburan maklumat tentang tindakan dan perlakuan spesies ikan atau perkara berkaitan dengan spesies ikan yang dirujuk dalam KNBI bahasa Melayu.

4.3 Kekerapan Kata Adjektif

Terdapat pelbagai *type* dan *token* kata adjektif dalam KNBI asal dan KNBI standard yang dibandingkan dalam kajian ini. Antara *type* kata adjektif ini ialah *type* ‘jarang’ pada nama biasa ikan ‘yu gergaji jarang’, *type* ‘keras’ pada nama biasa ikan ‘yu muncung keras’ dan *type* ‘biru’ pada nama biasa ikan ‘gombing biru’. Sementara itu, contoh *token* kata adjektif pula ialah seperti *token* ‘biru’ pada nama biasa ikan ‘gombing biru’, ‘jerung biru’ dan ‘kerapu biru bintik’.



Rajah 2.3. Perbandingan kekerapan kata adjektif antara KNBI asal dengan KNBI standard.

Pengelasan *type* dalam kajian ini mendapati bahawa jumlah kekerapan *type* kata adjektif dalam KNBI asal ialah 138 *type*, sementara jumlah kekerapan *type* kata adjektif dalam KNBI yang telah distandardkan ialah 168 *type*. Perbandingan di antara jumlah-jumlah ini menunjukkan bahawa jumlah kekerapan *type* kata adjektif dalam KNBI asal adalah lebih rendah berbanding jumlah kekerapan *type* kata adjektif dalam KNBI yang telah distandardkan. Beza di antara kedua-dua jumlah kekerapan *type* ini juga menunjukkan bahawa meskipun proses penyelarasan nama biasa yang dilakukan telah menyebabkan penurunan jumlah kekerapan *type* kata nama dan kata kerja dalam KNBI bahasa Melayu, namun proses ini telah menyumbang kepada peningkatan jumlah kekerapan *type* kata adjektif dalam KNBI bahasa Melayu, iaitu sebanyak 30 *type*.

Kajian ini mendapati bahawa peratus kekerapan *type* kata adjektif dalam KNBI asal ialah 13.69 %, manakala peratus kekerapan *type* kata adjektif dalam KNBI yang telah distandardkan ialah 20.92%. Perbandingan di antara kedua-dua nilai peratus ini menunjukkan bahawa peratus kekerapan *type* kata adjektif dalam KNBI asal adalah lebih rendah berbanding peratus kekerapan *type* kata adjektif dalam KNBI yang telah distandardkan. Selain itu, perbandingan di antara kedua-dua nilai peratus ini menunjukkan bahawa proses penyelarasan nama biasa yang dilakukan telah menyumbang kepada peningkatan peratus kekerapan *type* kata adjektif dalam KNBI bahasa Melayu, iaitu sebanyak 7.23% meskipun proses ini menyebabkan penurunan peratus kekerapan *type* kata nama dan kata kerja dalam KNBI bahasa Melayu.

Pengelasan *token* dalam kajian ini mendapati bahawa jumlah kekerapan *token* kata adjektif dalam KNBI asal ialah 270 *token*, sedangkan jumlah kekerapan *token* kata adjektif dalam KNBI yang telah distandardkan ialah 1125 *token*. Perbandingan di antara kedua-dua jumlah ini menunjukkan bahawa jumlah kekerapan *token* kata adjektif dalam KNBI asal adalah jauh lebih rendah berbanding jumlah kekerapan *token* kata adjektif dalam KNBI yang telah distandardkan. Selisih di antara kedua-dua jumlah kekerapan *token* ini juga menunjukkan bahawa proses penyelarasan nama biasa yang dilakukan telah menyumbang kepada peningkatan jumlah kekerapan *token* kata adjektif dalam KNBI bahasa Melayu sebanyak 855 *token*, iaitu lebih tiga kali ganda daripada jumlah kekerapan *token* kata adjektif dalam KNBI asal. Walau bagaimanapun, kadar peningkatan jumlah kekerapan *token* kata adjektif ini masih tidak dapat menandingi kadar peningkatan jumlah kekerapan *token* kata nama dalam KNBI bahasa Melayu.

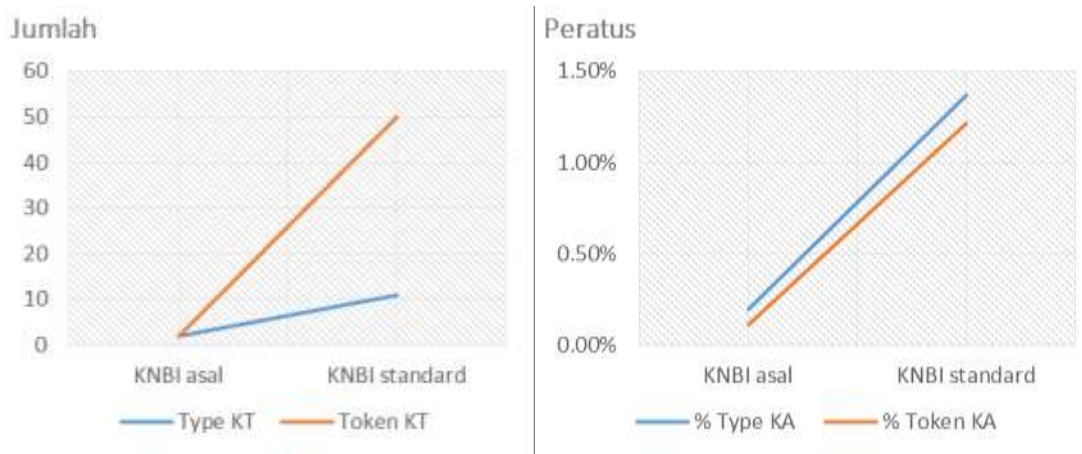
Sementara itu, hasil penelitian dalam kajian ini juga mendapati bahawa peratus kekerapan *token* kata adjektif dalam KNBI asal ialah 15.06%. Peratus kekerapan *token* kata adjektif dalam KNBI yang telah distandardkan pula ialah 27.41%. Perbandingan kedua-dua nilai peratus ini menunjukkan bahawa peratus kekerapan *token* kata adjektif dalam KNBI asal adalah lebih rendah berbanding peratus kekerapan *token* kata adjektif dalam KNBI yang telah distandardkan. Hal ini menunjukkan bahawa proses penyelarasan nama biasa yang dilakukan telah menyebabkan peratus kekerapan *token* kata adjektif dalam KNBI bahasa Melayu meningkat sebanyak 12.35%. Berdasarkan rajah 3, dapat dilihat bahawa peningkatan peratus kekerapan

token kata adjektif ini lebih drastik daripada peningkatan peratus kekerapan *type* kata adjektif dalam KNBI bahasa Melayu.

Sehubungan itu, kata adjektif ialah kata yang menerangkan sifat atau keadaan sesuatu benda (Baharudin dan Radzi, 2015). Oleh itu, peningkatan jumlah kekerapan *type*, peningkatan peratus kekerapan *type*, peningkatan jumlah kekerapan *token* dan peningkatan peratus kekerapan *token* yang terdiri daripada kata adjektif ini dapat menjelaskan bahawa proses penyeragaman yang dilakukan telah menyumbang kepada penambahan saiz variasi maklumat, penambahan imbalan variasi maklumat, penambahan saiz taburan maklumat dan penambahan imbalan taburan maklumat tentang sifat spesies ikan atau sifat bagi perkara berkaitan dengan spesies ikan yang dirujuk dalam KNBI bahasa Melayu.

4.4 Kekerapan Kata Tugas

Pelbagai *type* dan *token* kata tugas wujud dalam KNBI asal dan KNBI standard yang dibandingkan dalam kajian ini. Antara *type* kata tugas ini ialah *type* ‘hujung’ yang terdapat pada nama biasa ikan ‘yu hujung sirip hitam’, *type* ‘betul’ pada nama biasa ikan ‘tamban betul’ dan *type* ‘dua’ pada nama biasa ikan ‘debam dua jalur’. Contoh *token* kata tugas pula ialah seperti *token* ‘dua’ pada nama biasa ikan ‘debam dua jalur’, ‘nuri dua tanda’ dan ‘pasir-pasir dua garis’.



Rajah 2.4. Perbandingan kekerapan kata tugas antara KNBI asal dengan KNBI standard.

Kajian ini mendapati bahawa jumlah kekerapan *type* kata tugas dalam KNBI asal ialah 2 *type*. *Type* ini terdiri daripada kata tugas ‘betul’ dan ‘hujung’. Jumlah kekerapan *type* kata tugas dalam KNBI yang telah distandardkan pula ialah 11 *type*. Perbandingan di antara jumlah-jumlah ini menunjukkan bahawa jumlah kekerapan *type* kata tugas dalam KNBI asal adalah lebih rendah berbanding jumlah kekerapan *type* kata tugas dalam KNBI yang telah distandardkan. Perbandingan di antara kedua-dua jumlah kekerapan *type* ini juga menunjukkan bahawa proses penyelarasan nama biasa yang dilakukan bukan hanya menyumbang dalam peningkatan jumlah kekerapan *type* kata adjektif dalam KNBI bahasa Melayu, malah proses ini juga turut menyumbang kepada peningkatan jumlah kekerapan *type* kata tugas dalam KNBI bahasa Melayu. Walau bagaimanapun, kadar peningkatan jumlah kekerapan *type* kata tugas dalam KNBI bahasa Melayu ini lebih rendah, iaitu sebanyak 9 *type* atau kurang satu per tiga berbanding kadar peningkatan jumlah kekerapan *type* kata adjektif dalam KNBI bahasa Melayu.

Pada masa yang sama, analisis dalam kajian ini mendapati bahawa peratus kekerapan *type* kata tugas dalam KNBI asal ialah 0.20 %, sementara peratus kekerapan *type* kata tugas dalam KNBI yang telah distandardkan pula ialah 1.37%. Selisih di antara kedua-dua nilai peratus ini menunjukkan bahawa peratus

kekerapan *type* kata tugas dalam KNBI asal adalah lebih rendah berbanding peratus kekerapan *type* kata tugas dalam KNBI yang telah distandardkan. Selisih di antara kedua-dua nilai peratus ini juga menunjukkan bahawa meskipun proses penyeragaman nama biasa ikan yang telah dilakukan menyebabkan penurunan peratus kekerapan *type* kata nama dan kata kerja dalam KNBI bahasa Melayu, namun proses ini telah menyumbang kepada peningkatan sebanyak 1.17% kekerapan *type* kata tugas dalam KNBI bahasa Melayu.

Selain pengelasan *type*, pengelasan *token* dalam kajian ini mendapati bahawa jumlah kekerapan *token* kata tugas dalam KNBI asal ialah 2 *token*. Jumlah *token* ini adalah sama dengan jumlah *type* kata tugas dalam KNBI asal kerana bilangan *type* kata tugas ‘betul’ dan ‘hujung’ masing-masing terdiri daripada 1 *token*. Sementara itu, jumlah kekerapan *token* kata tugas dalam KNBI yang telah distandardkan ialah 50 *token*. Perbandingan di antara kedua-dua jumlah ini menunjukkan bahawa jumlah kekerapan *token* kata tugas dalam KNBI asal adalah jauh lebih rendah berbanding jumlah kekerapan *token* kata tugas dalam KNBI yang telah distandardkan. Beza di antara kedua-dua jumlah kekerapan *token* ini juga menunjukkan bahawa proses penyeragaman nama biasa yang dilakukan telah menyumbang kepada peningkatan jumlah kekerapan *token* kata tugas dalam KNBI bahasa Melayu sebanyak 48 *token*, iaitu 24 kali ganda daripada jumlah kekerapan *token* kata tugas dalam KNBI asal. Walau bagaimanapun, kadar peningkatan jumlah kekerapan *token* kata tugas ini adalah lebih kecil berbanding kadar peningkatan jumlah kekerapan *token* kata nama dan kata adjektif dalam KNBI bahasa Melayu.

Sementara itu, hasil penelitian dalam kajian ini juga mendapati bahawa peratus kekerapan *token* kata tugas dalam KNBI asal ialah 0.11%, manakala kekerapan *token* kata tugas dalam KNBI yang telah distandardkan pula ialah 1.22%. Perbandingan kedua-dua nilai peratus ini menunjukkan bahawa peratus kekerapan *token* kata tugas dalam KNBI asal adalah lebih rendah berbanding peratus kekerapan *token* kata tugas dalam KNBI yang telah distandardkan. Hal ini menunjukkan bahawa proses penyeragaman nama biasa yang dilakukan telah menyebabkan peratus kekerapan *token* kata tugas dalam KNBI bahasa Melayu meningkat sebanyak 1.11%. Peningkatan peratus kekerapan *token* kata tugas dalam KNBI bahasa Melayu ini hampir selari dengan peningkatan peratus kekerapan *type* kata tugas dalam KNBI bahasa Melayu. Beza antara kedua-dua kadar peningkatan peratus ini sangat kecil, iaitu sebanyak 0.06%.

Menurut Kirkham (1857), kata tugas ialah kata yang berperanan dalam menspesifikasi makna yang terkandung dalam kata lain. Oleh itu, hubung kait antara pernyataan ini dengan peningkatan jumlah kekerapan *type*, peningkatan peratus kekerapan *type*, peningkatan jumlah kekerapan *token* dan peningkatan peratus kekerapan *token* yang terdiri daripada kata tugas ini dapat menjelaskan bahawa proses penyeragaman yang dilakukan telah menyumbang kepada penambahan saiz variasi maklumat, imbalan variasi maklumat, saiz taburan maklumat dan imbalan taburan maklumat yang lebih spesifik tentang spesies ikan atau perkara berkaitan dengan spesies ikan yang dirujuk dalam KNBI bahasa Melayu.

5. Kesimpulan

Secara keseluruhannya, perbandingan kekerapan kata dalam kajian ini menunjukkan bahawa proses penyeragaman nama biasa ikan yang dilakukan telah menyumbang kepada peningkatan jumlah kekerapan *type*, jumlah kekerapan *token*, peratus kekerapan *type* dan peratus kekerapan *token* yang terdiri daripada golongan kata tertentu dalam KNBI bahasa Melayu. Semua peningkatan yang berlaku dapat menjelaskan bahawa proses penyeragaman nama biasa ikan ini telah menyumbang kepada penambahan saiz taburan dan imbalan maklumat tertentu dalam koleksi atau korpus nama biasa ikan bahasa Melayu. Peningkatan dan penambahan ini boleh dilihat sebagai sebahagian daripada perkembangan dalam bahasa Melayu daripada hasil inovasi atau penyeragaman yang telah dilakukan. Perkembangan bahasa daripada hasil inovasi bahasa ini perlu dikawal dan ditapis dengan teliti (Nor Hashimah, 2015), supaya perkembangan ini tidak menyebabkan bahasa Melayu semakin jauh daripada identitinya yang tersendiri.

Perbandingan kekerapan kata yang dilakukan dalam kajian ini juga menunjukkan bahawa proses penyeragaman nama biasa ikan yang dilakukan juga telah menyumbang kepada penurunan jumlah kekerapan *type*, jumlah kekerapan *token*, peratus kekerapan *type* dan peratus kekerapan *token* yang terdiri daripada golongan kata tertentu yang lain dalam KNBI bahasa Melayu. Semua penurunan yang berlaku dapat

menjelaskan bahawa proses penyelarasan nama biasa ikan ini juga telah memberikan kesan kepada pengurangan saiz taburan dan imbangan maklumat tertentu dalam koleksi atau korpus nama biasa ikan bahasa Melayu. Berdasarkan saranan Kikusawa (2012), kesan penurunan dan pengurangan nilai dalam aspek-aspek bahasa yang disebabkan oleh proses penyelarasan dan penstandardan ini tidak boleh dipandang ringan. Oleh itu, proses penyelarasan dan penstandardan nama biasa ikan dalam bahasa Melayu ini perlu dilakukan dengan lebih berhati-hati.

Rujukan

- Baharudin, M., & Radzi, R. M. (2015). Pengajaran Bahasa Melayu sebagai Bahasa Asing kepada Pelajar dari Negara China: Perkongsian Pengalaman di Universiti Sains Malaysia. *Dinamika Ilmu: Jurnal Pendidikan*, 15(1), 85-98.
- Claborn, D. M. (2010). The Biology and Control of Leishmaniasis Vectors. *Journal of Global Infectious Diseases*, 2(2), 127-134. <http://doi.org/10.4103/0974-777X.62866>
- Collin, P. M. & Hakala, M. (2011, June). *Standardizing the Corpus of Early English Correspondence (CEEC)*. Poster dalam seminar ICAME ke-32. (pp. 1-5).
- Freire, K. M. F., & Filho, A. C. (2009). Richness of common names of Brazilian reef fishes. *Pan-American Journal of Aquatic Sciences*, 4(2), 96-145.
- Gibb, T. J. (2014). *Contemporary Insect Diagnostics*. USA: Elsevier Science.
- Hishamudin, I., & Norsimah, M. A., (2012). Nilai setia dari perspektif prosodi semantik: Analisis berbantu data korpus. *GEMA: Online Journal of Language Studies*, 12(2 (special edition)), 359-374.
- Imran, H. A. (2011). Produktiviti Imbuan Mem-, Memper- Dan Memper-Kan: Pendekatan Linguistik Korpus. *Jurnal Linguistik*. 12(1), 41-62
- Johnson, S., & Ensslin, A. (2006). Language in the news: Some reflections on keyword analysis using WordSmith Tools and the BNC. *Leeds Working Papers in Linguistics and Phonetics*, 11, 96-109.
- Junaini K., Harshita, A. H., Nor Suhaila. C. P., & Zuhairah, I. (2017). Gandaan Separa dalam Terminologi Bahasa Melayu: Analisis Sosioterminologi (Partial Reduplication in Malay Terminology: A Socio-terminological Analysis). *GEMA Online® Journal of Language Studies*, 17(1).
- Kikusawa, R. (2012). Standardization as language loss: Potentially endangered Malagasy languages and their linguistic features. *People and culture in Oceania*, 28, 23-44.
- Kirkham, S. (1857) *English Grammar, in Familiar Lectures*. New York: Collins & Brother. [Web.] Retrieved from the Library of Congress, <https://lccn.loc.gov/10027832>.
- Kos, M. P. (2011). Local Bird Names and Their Standard Name Equivalents. (*Tesis Doktor Falsafah*). Charles University, United Kingdom.
- Lailatul, F. (2014). Morphological pattern of adjective words used in descriptive essay written by third semester students of English departement of IAIN Tulungagung. *Tesisi Master*. Tulungagung: State Islamic Institute (IAIN)
- Nik Safiah, K. et al. (2010). *Tatabahasa Dewan Edisi Ketiga*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Nor Hashimah, J. (2015). Pelestarian dan Perekayasaan Bahasa Melayu. *Jurnal Linguistik*. 19(1), 1-13
- Nugraha, F. D. (2015). Sufiks infleksi dalam artikel bisnis pada surat kabar The Jakarta Post: Kajian morfologis. *Tesis doktor falsafah*. Bandung: Universitas Widyatama
- Rozaimah, R. (2015). Metafora Konsepsi MARAH dalam Data Korpus Teks Tradisional Melayu. *Jurnal Linguistik*. 19(1), 29-47
- Saily, T., Nurmi, A., & Sairio A. (2013). *Linguistic Change In Its Social Contexts In Eighteenth-Century English: New Methods For Historical Sociolinguistics*. Poster dalam seminar ICAME ke-34, Santiago de Compostela.
- Singnoi, U. (2011). A reflection of Thai culture in Thai plant names. *MANUSYA: Journal of Humanities*, 14(1), 79.
- Stevenson, M., & Gaizauskas, R. (2000, April). Using corpus-derived name lists for named entity recognition. Dalam *Proceedings of the sixth conference on Applied natural language processing* (pp. 290-295). Association for Computational Linguistics.
- Yusri A. et. al. (2010). *Ikan Laut Malaysia: Glosari Nama Sahih Spesies Ikan*. Kuala Lumpur: Dawama Sdn. Bhd.