

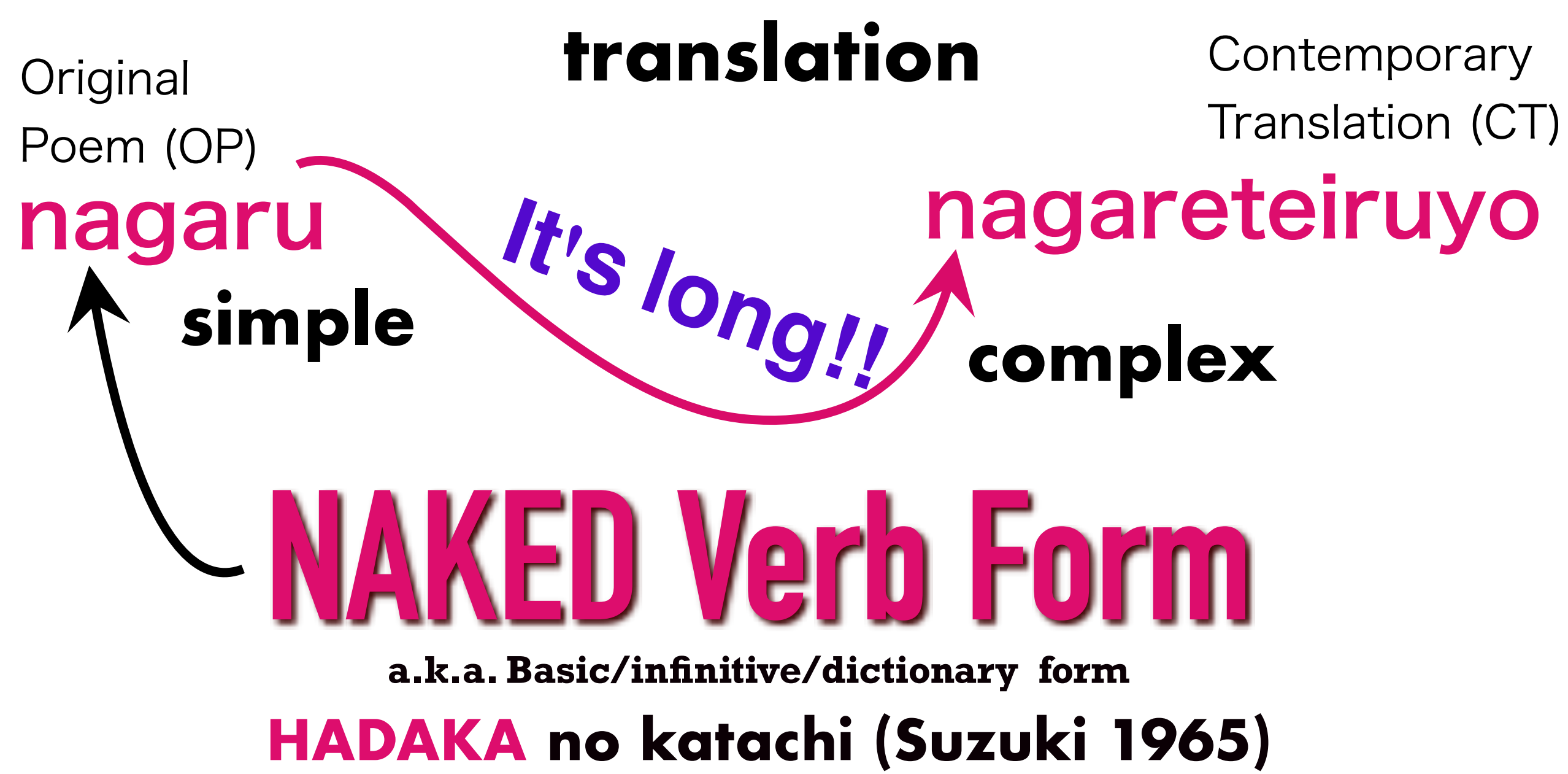
# Development of a dataset for comparison between predicate verb phrases in the Kokinshū and their contemporary translations

Hilofumi Yamamoto† Bor Hodošček‡ Xudong Chen†

†Tokyo Institute of Technology and ‡Osaka University

Development of a dataset for comparison between verb phrases in the Kokinshū and their contemporary translations

Hilo Yamamoto  
Bor Hodošček  
Xudong Cheng



## Comparative study using OP/CT parallel text dataset

### Methods

Examine the following points:

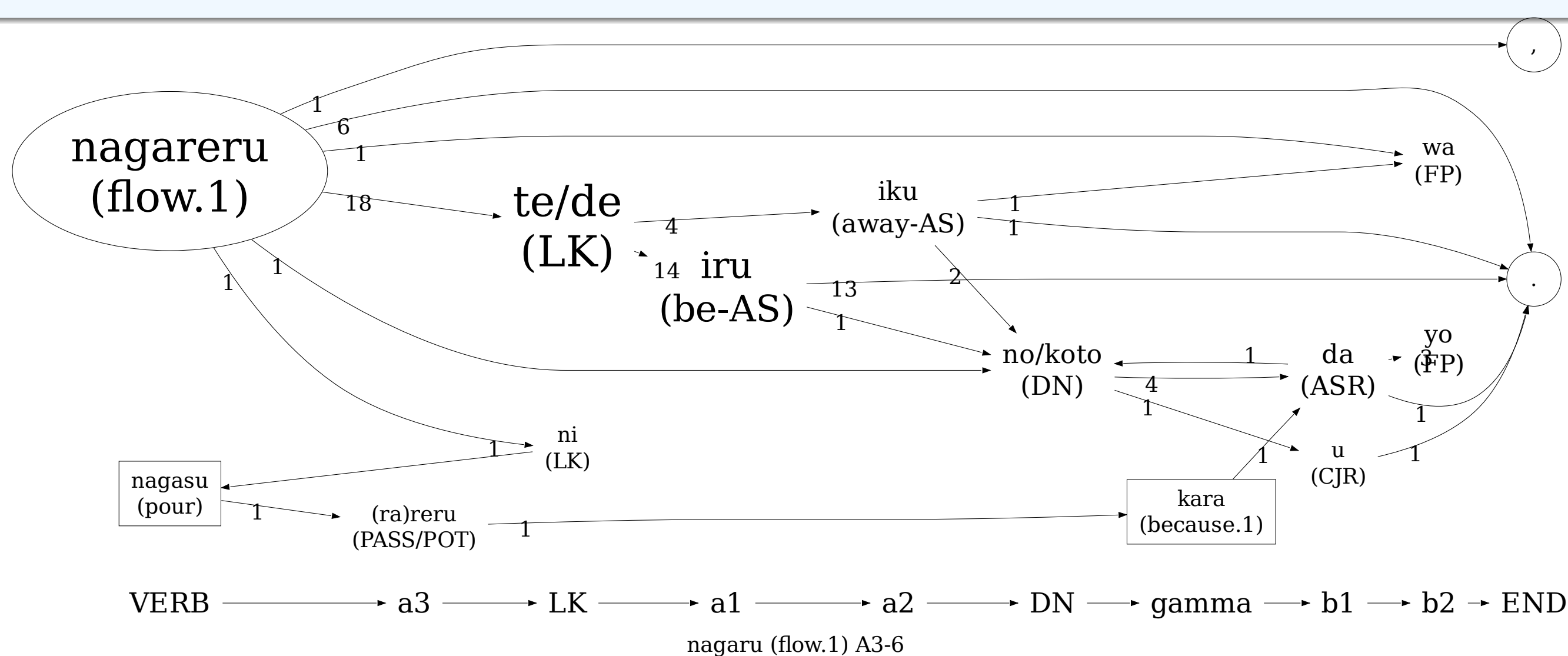
- Whether the same verb is used in OP and CT
- Whether syntactical elements are added to CT to maintain senses of OP
- Whether there are any other elements added to CT based on the context

### Problem of NAKED VERB: Infinitive form of a verb

**Table 1:** Four conversion patterns from Classical Japanese verbs (CJ) to Modern Japanese verbs (MJ) based on the description in Yamamoto (2005);  $V_{cj}$  and  $V_{mj}$  indicates a verb of CJ and the corresponding verb with MJ respectively; *nagaru* (flow) of CJ is assumed to be as the same as *nagareru* (flow) of MJ.

rules		examples	
Classical J.	→ Modern J.	Classical J.	→ Modern J.
1 $V_{cj}$	$V_{mj}$	<i>nagaru</i>	<i>nagareru</i>
2 $V_{cj}$	$V_{mj}$ -teiru	<i>nagaru</i>	<i>nagareteiru</i>
3 $V_{cj}$	$V_{mj}$ -ta	<i>nagaru</i>	<i>nagareta</i>
4 $V_{cj}$	$V_{mj}$ (-u)	<i>nagaru</i>	<i>nagareru(darō/dearō)</i>

- MJ verbs can also express both past and non-past. (Takahashi, 1983)
- The Naked CJ verbs CANNOT be simply replaced with the Naked MJ verbs (Kato, 1986, 62).



**Figure 1:** Construction of the predicate of *nagaru* (flow)

### Conclusion

- We found that most elements in CT predicates are based on the literal elements present in the OP.
- We found that most CT predicates are expressed with greater length than OP predicates, and the ways in which they are expressed vary.
- The concise form of the CJ naked verb is not inherently bound to the modern Japanese construction of “to be ...ing”; rather, its conciseness allows for a range of interpretations beyond this specific usage.
- Predicate verb phrases in classical Japanese poetry exhibit various patterns beyond the plain form, necessitating the development of methods to address other complex forms within the current dataset.

### Introduction

- Waka poems are limited to 31 characters
- Each word should be expressed in the most concise manner possible.
- Developed a parallel dataset comprising ten modern translations.
- To compare the Kokinshū (ca. 905) verbs with contemporary verbs Kokinshū includes 1,111 poems; compare each with ten translation sentences.
- Verbs of Original Poems (OP) / Verbs of Contemporary Translations (CT)
- Classified all elements in CT into syntactical, contextual, and other elements.  
 $\alpha_1 = \{ (ru), ta, iru \}; \alpha_2 = \{ aru, iku, kuru, shimau, .. \};$   
 $\alpha_3 = \{ (ra)reru, nai, .. \};$   
 $\beta_1 = \{ (u), darō, dearō, deshō, .. \}; \beta_2 = \{ ka, nā, yo, zo, sa, .. \};$   
 $LK = \{ te, de, ni, .. \} DN = \{ no, koto \}.$
- Focus on the analyses on infinitive (plain) form. (Suzuki, 1965)

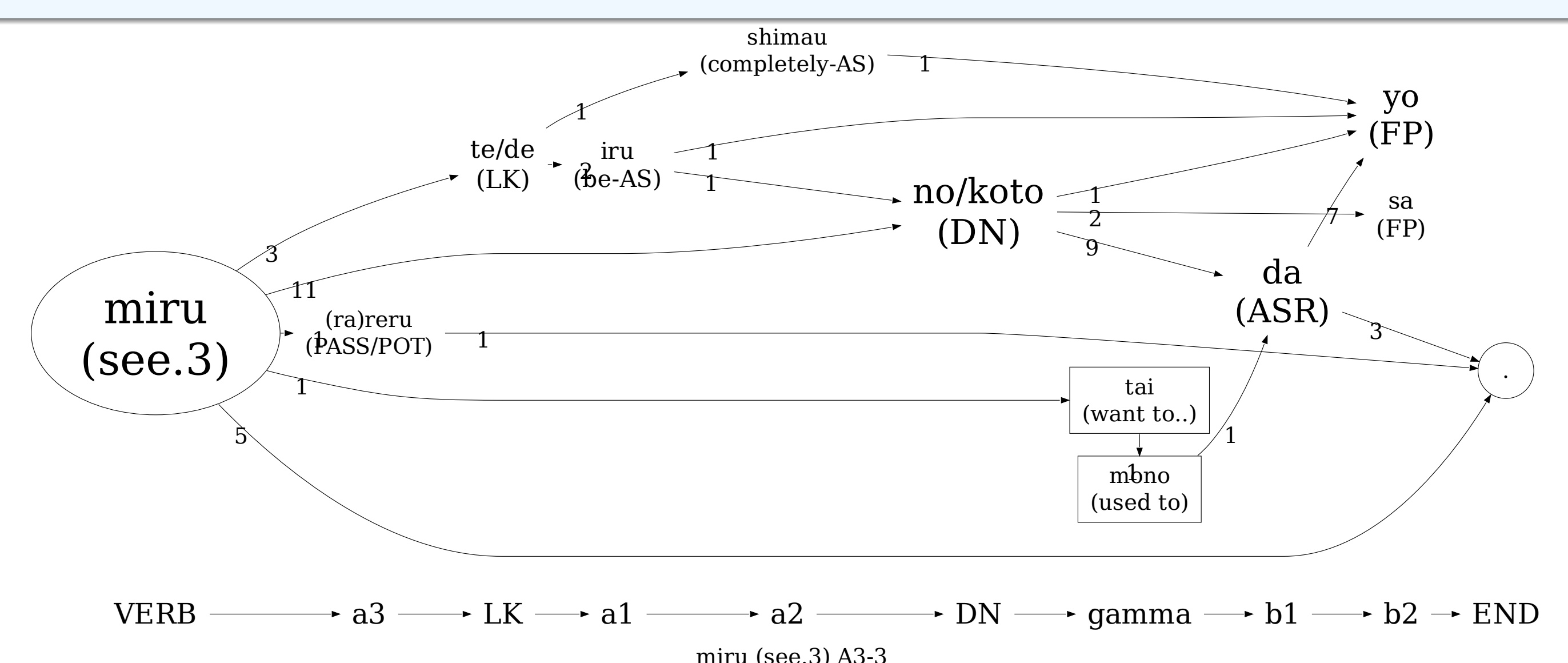
### Materials

- Use of the Kokinshū: 1,000 poems excluding those not in the 5/7/5/7/7 form
- Zenodo Hachidaishū vocabulary dataset (Yamamoto and Hodošček, 2021)
- Annotations were taken from sources: Shin-Nihon Koten Bungaku Taikei Bon Nijūichidaishū (New Japanese Classical Literature Compendium), etc.

### Results

**Table 2:** Predicates of CT corresponding to the predicate of *nagaru* in OP.

translator	predicate
0 Kokinshū	<i>nagaru</i> (to flow; infinitive)
1 Kaneko	<i>nagare-te-iku-wa-</i> .
2 Katagiri	<i>nagare-te-iku-koto-de-aru-yo-</i> .
3 Komachiya	<i>nagare-te-iku-</i> .
4 Kubota	<i>nagare-te-iru-koto-de-aru-yo-</i> .
5 Kyusojin	<i>nagareru-koto-de-aru-yo-</i> .
6 Kojima-Arai	<i>nagare-te-iku-koto-da-</i> .
7 Ozawa	<i>nagare-ni-nagasa-reru-kara-na-no-daro-u-</i> .
8 Takeoka	<i>nagareru-</i> .
9 Okumura	<i>nagasa-re-te-tachimachi-bosshi-te-shimau-</i> .
10 Matsuda	<i>utsut-te-iku-yo-</i> .



**Figure 2:** Construction of the predicate of *miru* (see)

### Reference

- Kato, Yasuhide (1986) “Bunmatsu ni shiyō sareru dōshi no imi yōhō no shidō (Teaching methods of meanings and uses of verbs used as a predicate)”, *Nihongogaku*, Vol. 5, No. 4, pp. 55–63.
- Suzuki, Shigeyuki (1965) “Gendai nihongo no dōshi no tensu —iikiri no jutsugo ni tsukawareta baai— (Tense of verbs of Modern Japanese: the case of verbs at the end of the predicate.)”, in *Kotoba no kenkyū 2 (Study of language)*, Vol. 2 of Report of the National Language Research Institute, Tokyo: Shuei shuppan, pp. 1–38.
- Takahashi, Taro (1983) “Suru tomo shita tomo ieru toki (The cases that can be expressed by both ‘suru’ and ‘shita’), in *Kindaichi Haruhiko Hakushi koki kinen ronbunshū*, Vol. 2, Tokyo: Sanseido, pp. 405–34.
- Yamamoto, Hilofumi and Bor Hodošček (2021) “Hachidaishū vocabulary dataset”, <https://doi.org/10.5281/zenodo.4744170>.
- Yamamoto, Hirofumi Hilo (2005) “A Mathematical Analysis of the Connotations of Classical Japanese Poetic Vocabulary”, Ph.D. dissertation, Australian National University.