Supervised Word Sense Disambiguation on Taiwan Hakka Polysemy with Neural Network Models: A Case Study of *BUN*, *TUNG* and *LAU*

Huei-Ling Lai
National Chengchi University
hllai@nccu.edu.tw

Hsiao-Ling Hsu
National Chengchi University
heidimavishsu@gmail.com

Jyi-Shane Liu
National Chengchi University
jsliu@cs.nccu.edu.tw

Chia-Hung Lin
National Chengchi University
linch0520@gmail.com

Yanhong Chen
National Chengchi University.
andy83918@gmail.com

Abstract

This research aims to explore an optimal model for automatic word sense disambiguation for highly polysemous markers *BUN*, *TUNG* and *LAU* in Taiwan Hakka, a low-resource language. The performance of word sense disambiguation tasks is carried out by examining DNN, BiLSTM and CNN models under different window spans. The results show that the CNN model can achieve the best performance with a multiple sliding window of L2R2+ L3R3 and L5R5.

Keywords

polysemy, word sense disambiguation, Taiwan Hakka, neural network models, CNN

1 Introduction

Polysemous phenomenon leading to ambiguity is one of the crucial problems that need to be resolved for natural language processing. Extensive studies on word sense disambiguation (WSD) that engage in solving polysemous problems have provided valuable findings (Iacobacci et al., 2016; Kågebäck & Salomonsson, 2016; Raganato et al., 2017; Liu & Nguyen, 2018; Li et al., 2019). While most put emphasis on a few dominant languages like English and Chinese, low-resource languages like Taiwan Hakka still gain relatively scanty attention because of the unavailability of data, neither raw nor labeled.

Lai et al. (2020), employing the DNN and the BiLSTM models, is an endeavor that investigates what information is needed to achieve the best performance of automatic polysemous word sense disambiguation in Taiwan Hakka. As a follow-up study, this research, incorporating the DNN, the BiLSTM, and the CNN in the experiment, aims to further explore what model can achieve the best performance for three semantically and syntactically intertwining polysemous markers.

Since Taiwan Hakka is a low-resource language, the characteristics of raw data employed for training and testing these models is one of the core challenges. The quantity and the quality of the raw data play a central role in the performance of the experiments featured on deep learning. In this research, a workable coding framework is schematized encompassing the following procedure: integrating and modifying the findings of previous studies on Taiwan Hakka polysemous phenomena, manually annotating the data to ensure the reliability of the labeled data, and then applying the three models to the massive corpus data.

2 Related Work

2.1 Polysemous *BUN* 分, *TUNG* 同 and *LAU* 摎

The focal point of this study is to differentiate the polysemous *BUN*, *TUNG* and *LAU* from different syntactic structures they occur and the various syntactic elements they are surrounded by. Drawing from findings of extant literature (Lai, 2001, 2003a, 2003b, 2004, 2015; Chiang, 2006; Huang, 2012, 2014, 2015), the coding schemes of *BUN*, *TUNG*, and *LAU* for human annotators are illustrated in Table 1, Table 2 and Table 3, respectively. The human annotators are well trained in linguistics and all the annotations are double checked to reach final agreement. Four labels are applied to the usages of *BUN*: VD for verb of giving in ditransitive constructions; VC for causative verb in causative constructions or purposive/pivotal constructions; P_dative for preposition in dative constructions; and P_passive for preposition in passive constructions. Six labels are applied to the usages of *TUNG*: VS for state verb; C for conjunction functioning as a comitative marker; P_goal for preposition functioning as a goal marker; P_source for preposition functioning as a source marker; P_patient for preposition functioning as a patient marker; and P_benefactive for preposition functioning as a benefactive marker. Six labels applied to the usages of *LAU*: VA for action verb; C for conjunction functioning as a comitative marker; P_goal for preposition functioning as a goal marker; P_source for preposition functioning as a source marker; P_patient for preposition functioning as a patient marker; and P_benefactive for preposition functioning as a benefactive marker.

Table 1. The coding scheme of *BUN*.

Instances		Construction	Grammatical function	Label
1a	佢**分**一枝筆偃 Gi *BUN* yi gi bid ngai. he *BUN* one CL pen me 'He gave a pen to me'	Ditransitive	verb of giving	VD
1b	佢**分**偃一枝筆 Gi *BUN* ngai yi gi bid. he *BUN* me one CL pen 'He gave me a pen'.	Ditransitive	verb of giving	
1c	佢送一枝筆**分**偃 Gi sung yi gi bid *BUN* ngai. he give one CL pen *BUN* me 'He gave a pen to me'	Dative	Preposition	P_dative
1d	佢帶東西**分**狗仔食 Gi dai dung-xi *BUN* geu-e sid. he bring thing *BUN* dog eat 'He brought food for the dog to eat'.	Pivotal (Chiang, 2006) Purposive (Huang, 2015) Causative (Lai, 2015)	Causative verb	VC
1e	佢會**分**偃去台北 Gi voi *BUN* ngai hi toibed. he would *BUN* me go Taipei 'He would let me go to Taipei'.			
1f	佢**分**偃打 Gi *BUN* ngai da. he *BUN* me beat 'He was beaten by me'.	Passive	Preposition	P_passive

Table 2. The coding scheme of *TUNG*.

	Instances	Grammatical function	Label
2a	佢兩儕**同**名**同**姓。 Gi liong sa *TUNG* miang *TUNG* xiang He two CL *TUNG* name *TUNG* suname ‘They two have the same first and last name.’	State verb	VS
2b	暗晡夜，𠊎愛**同**阿爸去食喜酒。 Ambuya, ngai oi *TUNG* aba hi siid hi.jiu Night, 1SG MOD *TUNG* father go eat wedding.feast ‘At night, I am going to attend the wedding feast with my father.’	Conjunction (comitative marker)	C
2c	先生希望大家有麼个問題全做得**同**佢講，毋好放在自家心肝肚。 Xin.sang hi.mong tai.ga iu ma.ge mun.ti qion zo.ded *TUNG* gi gung, m ho biong di qid.ga xim.gon du Teacher hope everyone have what question all can *TUNG* 3SG talk, NEG good put at self mind inside ‘The teacher asked everyone to tell him if they have any question; they should not hold it inside their own mind.’	Preposition (goal marker)	P_goal
2d	該師父就**同**佢咬一個手指包轉來。 Ge sii.fu qiu *TUNG* gi ngau id ge su.zii.bau zon loi DEM master thus *TUNG* 3SG bite one CL knuckle turn come ‘That master thus bit off a knuckle from him.’	Preposition (source marker)	P_source
2e	你**同**厥花盎仔打爛忒，就愛賠錢分人。 Ngi *TUNG* gia fa.ang.e da lam ted, qiu oi poi qien *BUN* ngin 2SG *TUNG* 3SG.poss vase hit shattered PRT, thus MOD compensate money *BUN* human ‘You broke his vase, and you should compensate him with money.’	Preposition (patient marker)	P_patient
2f	太白星君就**同**佢賜兩支。 Tai.pag.sen.giun qiu *TUNG* gi su liong gi Tai-pag-sen-giun thus *TUNG* 3SG grant two CL. ‘Tai-pag-sen-giun thus granted him two (things).’	Preposition (benefactive marker)	P_benefactive

Table 3. The coding scheme of *LAU*.

	Instances	Grammatical function	Label
3a	食米篩目**摎**糖水，已合嘴。 Siid mi.qi.mug *LAU* tong-sui, i hab zoi Eat rice.noodle mix sugar-water, already match mouth ‘Eating rice noodle with sugar water is a good match to mouth.’	Action verb	VA
3b	佢**摎**吾爸從細共下到大，故所𠊎喊佢阿姑。 Gi *LAU* nga ba cong se kiong.ha do tai, gu.so ngai hen gi a.gu 3SG with 1SG.poss father from small together to big, therefore 1SG call 3SG aunt ‘She grew up with my father, and therefore I call her aunt.’	Conjunction (comitative marker)	C
3c	𠊎愛大聲**摎**別人講客話。 Ngai oi tai sang *LAU* ped ngin gong hag.fa 1SG MOD big sound *LAU* other human speak Hakka ‘I would speak Hakka with other people loudly.’	Preposition (goal marker)	P_goal
3d	該儕人屋下無錢又尋無頭路，將就去**摎**人分飯食。 Ge sa ngin vug.ka mo qien iu mo qin teu.lu, jiong qui hi *LAU* ngin *BUN* fan siid DEM CL human home NEG money and find NEG job, altogether go *LAU* human share rice eat ‘That person had no money in his home and couldn’t find a job, so he altogether shared food from other people.’	Preposition (source marker)	P_source
3e	砰一聲，斯**摎**門關起來。 Bang id sang, sii *LAU* mun gon hi loi Bang one sound, then *LAU* door close up come ‘With a “bang” sound, (somebody) then shut the door.’	Preposition (patient marker)	P_patient
3f	阿英盡會**摎**人作媒人。 a.in qin voi *LAU* ngin zo moi.ngin A-in very be.capable *LAU* human do matchmaking ‘A-in is very capable of matching couples.’	Preposition (benefactive marker)	P_benefactive

2.2 Neural Network Models for WSD tasks: DNN, BiLSTM, and CNN

The two models, a feed-forward DNN with 10 hidden layers and a Bi-LSTM (Graves and Schmidhuber, 2005; Graves, Mohamed, and Hinton, 2013) are compared and contrasted in Lai et al. (2020). In this study, CNN sentence classification is additionally adopted for further comparison. Originally designed to cope with computer vision problems (Lecun et al.,1998), Convolutional Neural Networks (CNN) have also been shown to be capable of dealing with natural language processing tasks (Collobert et al., 2011; Kalchbrenner et al., 2014; Shen et al., 2014; Yih et al, 2014). Among many NLP task applications, sentence classification by Kim (2014) is one of the applications that fit our research the most. CNN sentence classification concatenates pretrained word2vec vectors as a sentence matrix, from which the model can extract multiple types of features with filters of different size from the textual vector space, as other CNN models extract that from the image vector space.

3 Methods

3.1 Overall architecture

Three kinds of information in the context contingent to the target are extracted: the neighboring POS, word, and character. In Lai et al. (2020), four types of input are fed into the model to investigate which type of input can achieve the best performance for classifying *BUN*, and it is reported that the type that includes all the features achieves the best performance under a window span of L3R3 with DNN and BiLSTM: POS representation + word embeddings + character-based embeddings. Thus, in this study, we employ this type of input as an initial attempt to explore which model can achieve the best performance among DNN, BiLSTM, and CNN.

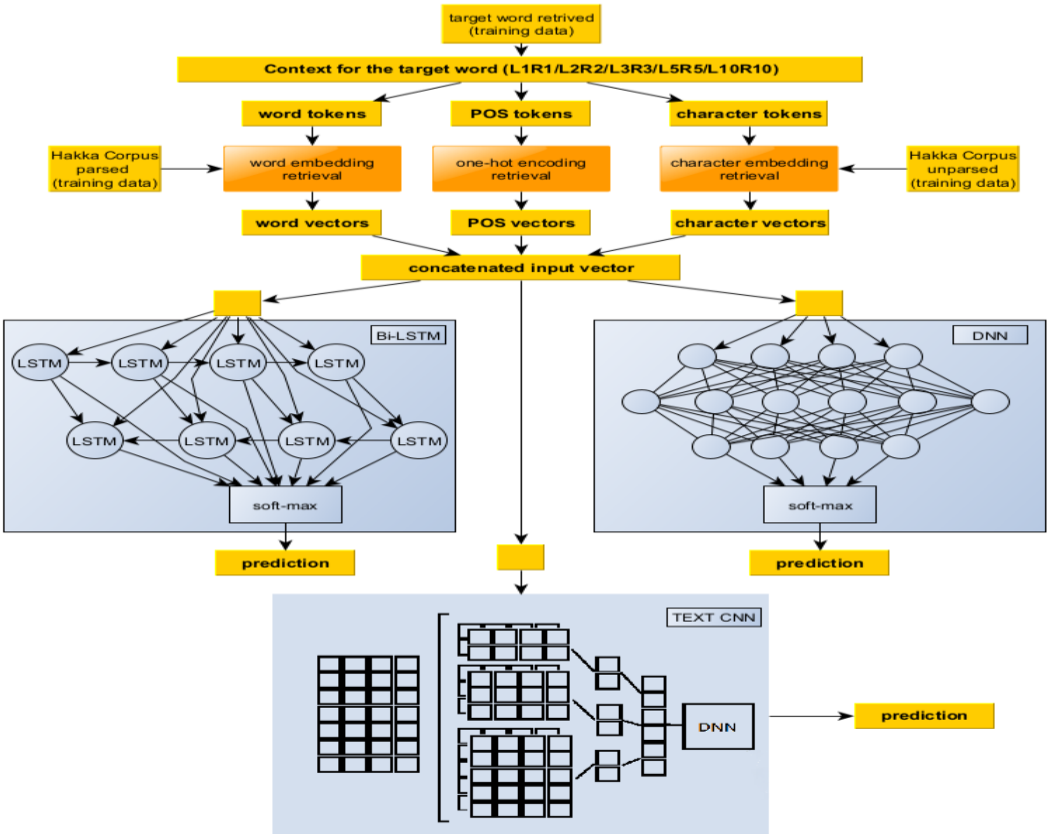
As demonstrated in Figure 1, the model we used to classify the polysemous words in Taiwan Hakka mainly consists of two parts: embeddings and classification. The embeddings part vectorizes and concatenates the features obtained from the data. The feature of POS is represented with one-hot encoding, while word and character-based embeddings are generalized by using word2vec algorithm. Then, the vectors are concatenated. As for the classification part, three neural networks are employed: DNN, Bi-LSTM and CNN.

3.2 Input layers and output layers

The inputs are n-dimensional real-valued vectors. A range of window span is selected to investigate how much contextual information is needed to achieve the best performance in the task of classification: L1R1; L2R2; L3R3; L5R5 and L10R10. For instance, in an instance containing either *BUN*, *LAU* or *TUNG* with a window span mentioned, every POS/word/character in that span is converted into an n-dimensional real-valued vector. POS features are represented with one-hot encoding (24-dimensional embeddings). Word and character-based embeddings (128-dimensional embeddings) are trained in the dataset 4a, 4b and 4c, as shown in Table 5. As for the vector concatenation, the input vector for each word is concatenated in a consecutive order: POS, word-embedding, and character-based. To ensure the same input shape, the character-based embedding vectors are restricted in three characters.

The output layers present the results of classification, which report the probabilities of the POS classes in *BUN*, *LAU* or *TUNG*.

Figure 1. The disambiguation model structure.



4 Experiments

4.1 Dataset

The main challenge for this research is the characteristics of raw data as its quantity and quality can crucially influence WSD tasks featured on deep learning. To obtain a substantial amount of raw data, we have continuously retrieved raw data from the Taiwan Hakka Corpus along with its development with a total of two million characters. To ensure the quality of the target data, we have removed cases which are wrongly segmented and have conducted several laborious examinations to ensure the reliability of hand-labeled POS annotations on the three target polysemous words. In addition, after a careful examination of the data, we have found that a considerable number of sentences that express the same ideas differ only in dialectal variations but can be quite similar in their syntactic structures and vocabulary. These examples with such subtle differences may result in overfitting problems for the experiments. Hence, we then have decided to run the experiments by dialects.

In total, twelve datasets are used in the experiments: manually annotated instances containing *BUN* in three dialects (Dataset 1a, 1b, and 1c); manually annotated instances containing *TUNG* in three dialects (Dataset 2a, 2b, and 2c);

manually annotated instances containing *LAU* in three dialects (Dataset 3a, 3b, and 3c); raw data retrieved from Taiwan Hakka Corpus in three dialects (Dataset 4a, 4b, and 4c). The manually annotated instances containing *BUN*, *TUNG*, or *LAU* are used as training sets and test sets. The raw data retrieved from Taiwan Hakka Corpus are used to train the word and character embeddings. The detailed statistical descriptions are presented in Table 4 and Table 5.

As reported in Table 4, the number of manually annotated instances containing *BUN* are 3,586 in *xiyen*, 2,676 in *namxiyen*, and 3,400 in *hoilu*. The distribution of the usages of *BUN* in three dialects all reflect the following pattern: VC occurs the most frequently, followed by P_passive, P_dative, and VD. As for *TUNG*, the number of manually annotated instances are 2,305 in *xiyen*, 2,492 in *namxiyen*, and 2,652 in *hoilu*. For *LAU*, the number of manually annotated instances are 2,807 in *xiyen*, 3,453 in *namxiyen*, and 2,454 in *hoilu*. It is interesting that the distribution of the usages of *TUNG* and *LAU* demonstrate a similar pattern in three dialects: C occurs the most frequently, followed by P_patient, P_goal, P_benefactive, P_source, and VS (for *TUNG*) or VA (for *LAU*).

The number of tokens in raw data are reported in Table 5. For *xiyen*, 635,610 characters (453,451 words) are retrieved. For *namxiyen*, 186,197 characters (133,625 words) are obtained. For *hoilu*, 590,544 characters (421,539 words) are extracted.

Table 4. The distribution of *BUN*, *TUNG* and *LAU* in manually annotated data

Dataset	Uses (POS types)	Token	Percentage
Dataset 1a (manually annotated instances containing *BUN xiyen*)	VD (ditransitive verb)	137	3.82%
	VC (causative verb)	1,541	42.97%
	P_ dative	514	14.33%
	P_ passive	1,394	38.87%
	Subtotal	3,586	100%
Dataset 1b (manually annotated instances containing *BUN namxiyen*)	VD (ditransitive verb)	89	3.33%
	VC (causative verb)	1,225	45.78%
	P_ dative	386	14.42%
	P_ passive	976	36.47%
	Subtotal	2,676	100%
Dataset 1c (manually annotated instances containing *BUN hoilu*)	VD (ditransitive verb)	115	3.38%
	VC (causative verb)	1,396	41.06%
	P_ dative	511	15.03%
	P_ passive	1,378	40.53%
	Subtotal	3,400	100%
Dataset 2a (manually annotated instances containing *TUNG xiyen*)	VS (state verb)	20	0.87%
	C (conjunction)	1,135	49.24%
	P_ goal	291	12.62%
	P_ source	86	3.73%
	P_ patient	571	24.77%
	P_ benefactive	202	8.76%
	Subtotal	2,305	100%
Dataset 2b (manually annotated instances containing *TUNG namxiyen*)	VS (state verb)	24	0.96%
	C (conjunction)	1,195	47.95%
	P_ goal	347	13.92%
	P_ source	102	4.09%
	P_ patient	592	23.76%
	P_ benefactive	232	9.31%

	Subtotal	2,492	100%
Dataset 2c (manually annotated instances containing *TUNG hoilu*)	VS (state verb)	18	0.68%
	C (conjunction)	1,270	47.89%
	P_ goal	373	14.06 %
	P_ source	113	4.26%
	P_ patient	609	22.96%
	P_ benefactive	269	10.14%
	Subtotal	2,652	100%
Dataset 3a (manually annotated instances containing *LAU xiyen*)	VA (action verb)	23	0.82%
	C (conjunction)	1,197	42.64%
	P_ goal	325	11.58%
	P_ source	118	4.20%
	P_ patient	870	30.99%
	P_ benefactive	274	9.76%
	Subtotal	2,807	100%
Dataset 3b (manually annotated instances containing *LAU namxiyen*)	VA (action verb)	27	0.78%
	C (conjunction)	1,478	42.80%
	P_ goal	391	11.32%
	P_ source	150	4.34%
	P_ patient	1,063	30.78%
	P_ benefactive	344	9.96%
	Subtotal	3,453	100%
Dataset 3c (manually annotated instances containing *LAU hoilu*)	VA (action verb)	26	1.06%
	C (conjunction)	1,093	44.54%
	P_ goal	245	9.98%
	P_ source	83	3.38%
	P_ patient	773	31.50%
	P_ benefactive	234	9.54%
	Subtotal	2,454	100%

Table 5. The number of tokens and types in each dataset

Dataset	**Segmentation**	**POS tagging**	**Training**	**Token**	**Type**
Dataset 4a *xiyen* (raw data retrieved from Taiwan Hakka Corpus in January, 2021)			Character embedding	635,610	4,093
	Yes	Yes	Word embedding	453,451	22,058
Dataset 4b *namxiyen* (raw data retrieved from Taiwan Hakka Corpus in January, 2021)			Character embedding	186,197	3,093
	Yes	Yes	Word embedding	133,625	11,479
Dataset 4c *hoilu* (raw data retrieved from Taiwan Hakka Corpus in January, 2021)			Character embedding	590,544	4,056
	Yes	Yes	Word embedding	421,539	21,583

4.2 Procedure

The experiments are designed to explore which model can achieve the best performance on WSD tasks for *BUN*, *LAU*, and *TUNG*. The four input features are all employed in the experiments: POS only, POS + word embedding, POS + character embedding, and POS + word embedding + character embedding. In addition, we also explore in which window span can the model get the best performance when the all input features are employed. The detailed procedures of each experiment are shown as follows.

Experiments
a. The input features with POS + word embedding + character embedding demonstrated in Table 4 and Table 5 are vectorized.

b.	The input features with five different window spans (L1R1, L2R2, L3R3, L5R5, and L10R10) are employed to train and test BiLSTM.
c.	The input features with five different window spans (L1R1, L2R2, L3R3, L5R5, and L10R10) are employed to train and test DNN (10 hidden layers).
d.	The input features with five different window spans (L1R1, L2R2, L3R3, L5R5, and L10R10) are employed to train and test CNN (single sliding window).
e.	The input features with five different window spans (L1R1, L2R2, L3R3, L5R5, and L10R10) are employed to train and test CNN (multiple sliding window).

4.3 Results and analysis

The results show that, by employing the CNN model, the highest accuracy rate of all the three polysemous words *BUN*, *LAU* and *TUNG* in different dialects can reach up to 80%. The accuracy rates of each model are presented in Table 6.

For *BUN*, CNN gains the highest accuracy rate among the three dialects. In *BUN xiyen* and *hoilu*, the highest accuracy rate (88.03% for *xiyen*; 89.5% for *hoilu*) is achieved when the CNN multiple sliding window with a window span of L5R5 is employed and in *BUN namxiyen*, the highest accuracy rate (91.25%) is gained when the CNN single sliding window with a window span of L5R5 is used. For *TUNG*, the CNN multiple sliding window achieves the highest accuracy rate among the three dialects (85.96% for *xiyen*; 83.84% for *namxiyen*; 84.51% for *hoilu*). For *LAU*, the CNN single sliding window and multiple sliding window gain the highest accuracy rate among the three dialects. In *LAU xiyen and LAU namxiyen*, the highest accuracy rate is achieved when the CNN multiple sliding window with a window span of L5R5 is employed (88.62% for *xiyen*; 82.79% for *namxiyen*). In *LAU hoilu*, the highest accuracy is reached when the CNN single sliding window with a window span of L3R3 is used (83.55% for *hoilu*).

Table 6. The accuracy rates of BiLSTM, DNN and CNN with the most input features under five window spans

Types of Model	**Window Span**				
	L1R1	**L2R2**	**L3R3**	**L5R5**	**L10R10**
BUN* in *xiyen					
BiLSTM	77.4	**82.2**	80.36	77.81	72.39
DNN (10 hidden layers)	75.86	82.61	68.5	**84.35**	76.27
CNN (single)	79.24	87.73	87.32	**87.83**	84.35
CNN (multiple 2+3+5)				**88.03**	85.27
BUN* in *namxiyen					
BiLSTM	**82.65**	81.28	79.09	73.49	71.85
DNN (10 hidden layers)	75.0	**85.24**	70.62	58.19	46.44
CNN (single)	82.78	88.38	90.57	**91.25**	90.02
CNN (multiple 2+3+5)				**90.84**	90.43
BUN* in *hoilu					
BiLSTM	77.16	**83.87**	78.57	82.9	75.32
DNN (10 hidden layers)	64.61	55.73	**87.66**	71.10	54.32
CNN (single)	79.00	87.98	87.44	**89.17**	87.66
CNN (multiple 2+3+5)				**89.50**	86.79

TUNG* in *xiyen					
BiLSTM	67.94	76.39	**79.1**	74.8	67.14
DNN (10 hidden layers)	54.86	**70.65**	66.02	68.58	63.47
CNN (single)	72.24	82.77	85.16	**85.8**	83.89
CNN (multiple 2+3+5)				**85.96**	84.84
TUNG* in *namxiyen					
BiLSTM	64.31	72.98	**73.56**	67.69	54.33
DNN (10 hidden layers)	55.5	65.34	62.4	**68.57**	67.98
CNN (single)	67.69	81.64	**83.25**	82.08	79.0
CNN (multiple 2+3+5)				**83.84**	79.58
TUNG* in *hoilu					
BiLSTM	67.78	78.1	**78.38**	75.17	72.94
DNN (10 hidden layers)	64.01	**72.52**	67.78	67.92	56.62
CNN (single)	67.92	82.98	**84.37**	83.12	81.86
CNN (multiple 2+3+5)				**84.51**	82.14
LAU in xiyen					
BiLSTM	63.39	**76.33**	75.94	73.33	61.69
DNN (10 hidden layers)	62.09	66.01	65.35	**66.66**	64.05
CNN (single)	66.92	83.79	86.53	**87.71**	86.53
CNN (multiple 2+3+5)				**88.62**	88.23
LAU in namxiyen					
BiLSTM	59.18	**71.47**	67.84	58.76	53.84
DNN (10 hidden layers)	56.3	**73.61**	56.62	68.48	63.99
CNN (single)	62.92	78.2	80.44	**82.69**	79.48
CNN (multiple 2+3+5)				**82.79**	81.3
LAU in hoilu					
BiLSTM	64.57	72.94	**73.84**	71.15	65.17
DNN (10 hidden layers)	60.53	64.57	**74.14**	59.49	70.25
CNN (single)	65.76	79.52	**83.55**	81.91	79.07
CNN (multiple 2+3+5)				**82.51**	80.71

5 Discussion

To explore the optimal model for tasks of automatic word sense disambiguation of polysemous *BUN*, *TUNG* and *LAU* in Taiwan Hakka, we conduct experiments with most input features under different ranges of window spans.

Overall, in *BUN*, *TUNG*, and *LAU*, the performance carried out by the CNN model is better (82% to 91%) than the ones carried out by the DNN model (66% to 87%) and the BiLSTM model (71% to 83%). As illustrated in Table 6, the results of the nine test sets reveal a tendency: except for *BUN namxiyen* and *LAU hoilu*, the highest accuracy rates in all the other seven test sets are achieved with the CNN multiple sliding window under a window span of L2R2 + L3R3 + L5R5. And this tendency may indicate that the CNN multiple sliding window is the optimal model for the automatic WSD tasks in Taiwan Hakka under a window span of L2R2 + L3R3 + L5R5. As for *BUN namxiyen*, the highest accuracy rate is achieved with the CNN single sliding window under a window span of L5R5; as for *LAU hoilu*, the highest accuracy rate is achieved with the CNN single sliding window under a window span of L3R3. While this inconsistency is detected, their best accuracy rate is not that far from the accuracy rate gained with the CNN multiple sliding window under a window span of L2R2 + L3R3 + L5R5: 91.25% versus 90.84% in *BUN namxiyen*; 83.55% versus 82.51% in *LAU hoilu*. However, this inconsistency remains unexplained and should be further studied.

Several computational implications come to light from our empirical study. First, the high accuracy rates of CNN (ranging from 82% to 91%) suggest that this model may optimize the development of automatic WSD system in Taiwan Hakka. Second, the results revealing a consistent tendency that the CNN multiple sliding window is the optimal model for the automatic WSD tasks under a window span of L2R2 + L3R3 + L5R5. This may indicate that in the case of Taiwan Hakka, to perform a successful classification, the contextual information in multiple window spans should be taken into consideration simultaneously.

A careful observation of the erroneous predictions made by the CNN model reveals that the patterns of erroneous predictions correlate with possible ambiguous cases proposed in the extent literature to some extent. For instance, in the three dialects, the most frequent erroneous prediction for *BUN* is that P_passive (hand-labeled) is wrongly predicted to be VC (predicted); for *TUNG*, C (hand-labeled) is wrongly predicted to be C (predicted); for *LAU*, P_patient (hand-labeled) is wrongly predicted to be C (predicted). These outcomes may imply that the CNN model can learn most of the patterns of the various uses of *BUN*, *TUNG* and *LAU*, but further efforts need to be put for CNN to learn these highly ambiguous cases.

Acknowledgments

We would like to thank the PACLIC 2021 anonymous reviewers for the valuable comments on this paper and MOST grant (MOST-108-2410-H-004-050-MY3) for supporting the research discussed herein.

References

Chiang, M. H. (2006). Grammatical characteristics of *TUNG* and *BUN* in Dongshi Hakka and the relatedness of the two markers. *Languages and Linguistics*, 7(2), 339-364.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research*, 12, 2493-2537.

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM networks. In *Proceedings. 2005 IEEE International Joint Conference on Neural Networks*, 2005. (Vol. 4, pp. 2047-2052). IEEE.

Graves, A., Mohamed, A. R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645-6649). IEEE.

Huang, H. C. (2012). Dative constructions in Hakka: A constructional perspective. *Hakka Studies, 5,* 39-72

Huang, H. C. (2014). Semantic Extensions and the Convergence of the Beneficiary Role: A Case Study of *BUN* and *LAU* in Hakka. *Concentric: Studies in Linguistics*, *40*(1), 65-94.

Huang, H. C. (2015). Relating Causative and Passive

"*BUN*" Constructions in Hakka. *Tsing Hua Journal of Chinese Studies*, *45*(2), 167-200.

Iacobacci, I., Pilehvar, M. T., & Navigli, R. (2016). Embeddings for word sense disambiguation: An evaluation study. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (pp. 897-907).

Kågebäck, M., & Salomonsson, H. (2016). Word

Lai, H. L. (2001). On Hakka *BUN*: A case of polygrammaticalization. *Language and Linguistics*, *2*(2), 137-153.

Lai, H. L. (2003a). The Semantic Extension of Hakka *LAU*. Language and Linguistics, 4(3), 533-561.

Lai, H. L. (2003b). Hakka *LAU* Constructions: A Constructional Approach. *Language and Linguistics*, *4*(2), 353-378.

Lai, H. L. (2004). The Syntactic Grounding and Conceptualization of Hakka *BUN* and *LAU*. *Concentric: Studies in Linguistics*, *30*(1), 87-105.

Lai, H. L. (2015). Profiling Hakka *BUN* 1 Causative Constructions. Language and Linguistics, 16(3), 369-395.

Lai, H.L., Hsu, H.L., Liu, J.S., Lin, C.H., & Chen, Y.H. (2020). Supervised Word Sense Disambiguation on Polysemy with Neural Network Models: A Case Study of *BUN* in Taiwan Hakka. *International Journal of Asian Language Processing*, *30* (3), 1-17.

sense disambiguation using a bidirectional lstm. arXiv preprint arXiv:1606.03568.

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A Convolutional Neural Network for Modelling Sentences, Baltimore, Maryland.

Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324. doi:10.1109/5.726791

Li, Z. H. I., Yang, F. A. N., & Luo, Y. (2019). Context embedding based on bi-LSTM in semi-supervised biomedical word sense disambiguation. *IEEE Access*, 7, 72928-72935.

Raganato, A., Bovi, C. D., & Navigli, R. (2017). Neural sequence learning models for word sense disambiguation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 1156-1167).

Shen, Y., He, X., Gao, J., Deng, L., & Mesnil, G. (2014). Learning semantic representations using convolutional neural networks for web search. In *Proceedings of the 23rd international conference on world wide web (pp. 373-374).*

Yih, W. T., He, X., & Meek, C. (2014). Semantic parsing for single-relation question answering. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 643-648).