

# HUANG Po-Hsuan

No.71, Sanduo St. • Yuanlin City • Changhua County • Taiwan  
+886 935 726 090 | [benson32169@gmail.com](mailto:benson32169@gmail.com) | <https://peh-suan.github.io>

## ACADEMICS

---

2020–2023	<b>National Taiwan University</b> <b>MA in Linguistics</b>	GPA 4.19/4.3
-----------	---	--------------

Relevant courses: Introduction to Syntax; Introduction to Phonology; Linguistic Fieldwork; Advanced Phonetics; Seminar on Cognitive Linguistics and Psychology; Special Topics on Text Mining and Natural Language Processing; Introduction to Psycholinguistics; Introduction to French Linguistics; Second Language Acquisition; Chinese Phonology; The Use of Ultrasound for Linguistic Research; Language and Society; Physiological Basis and Functions in Speech Production; Research Methodology; Data Analysis and Machine Learning with Python

---

2015–2020	<b>National Chengchi University</b> <b>BAs in English</b>	GPA 4.00/4.3
-----------	--	--------------

**European Languages and Cultures**  
**(Division of French Language and Culture)**

Relevant courses: Introduction to English Linguistics; Syntax; Introduction to Phonology; Language and Cognition; Development of the English Language; Conference Presentation and Journal Publication

## RESEARCH EXPERIENCE

---

<b>Natural Language Processing and Sentiment Analysis Lab</b> <b>(PI: Dr. Lun-Wei Ku)</b> Full-time Research Assistant <u>Projects (on-going)</u> <b>Enhancing abstract reasoning for LLMs and VLMs</b> <ul style="list-style-type: none"><li>Built a multi-hop instance-association dataset aimed for enhancing language models' abstract reasoning ability.</li></ul> <b>VLM-based ultrasonic-optical speech production learning tool</b> <ul style="list-style-type: none"><li>Using a large open-source lingual ultrasound corpus to train VLMs as speech production assistants for L2 acquisition/speech therapy.</li></ul> <b>Linguistic relativity in LLMs</b> <ul style="list-style-type: none"><li>Investigating linguistic influence on multilingual LLMs' performances.</li></ul> <b>Speech Behavior and Science Lab</b> <b>(PI: Dr. Chenhao Chiu)</b> Lab Member & Research Assistant <u>Notable projects</u> <b>High/low vowel differences in laryngeal height and EGG amplitude for different tones</b> <ul style="list-style-type: none"><li>Supervised lab interns on experiment design, experiment conduction, and analyses.</li><li>Developed pipelines for data analysis and visualization for multi-channel data.</li></ul> <b>Lip postures of Taiwan Mandarin High vowels</b> <ul style="list-style-type: none"><li>Developed automated face recognition for linguistic purposes with MediaPipe.</li><li>Designed automated lip landmark tracing for side faces with cluster analysis.</li></ul> <b>Squliq Atayal corpus construction</b> <ul style="list-style-type: none"><li>Labeled 20-hour-long recordings of Squliq Atayal produced by one female informant.</li><li>Cataloged the sound inventory of Squliq Atayal.</li><li>Assisted in Squliq Atayal speech auto-segmentation.</li></ul> <b>Prof. Hsuan-Lei Shao</b>	<b>Institute of Information Science</b> <b>Academia Sinica</b> <i>February, 2024–</i>
---	---

**Graduate Institute of Linguistics**  
**National Taiwan University**  
*July, 2021–*

**Research Center for Digital Law**  
**National Taiwan University**  
*August, 2023–December, 2023*

Research Assistant

Projects

**Legal judgment similarity computation**

- Devised a novel similarity computation based on graph analysis of textual data.
- Overcame the lack of prior-case citations in the continental law system.

**Association-rule-based domain-specific collocation extraction**

- Extracted domain-specific collocations in two different corpora (legal judgments and political talks) through iterative word re-segmentation based on (average) MI (mutual information).
- Devised an algorithm for (average) MI threshold determination based on mean sentential word ratios and the elbow method.

- Built a package that performs calculation of (average) MI for adjacent bigrams, re-segmentation of the original sentences, and visualization of the sentences as binary trees.

**Prof. Te-Hsin Liu**

**Graduate Program of Teaching Chinese as a Second Language  
National Taiwan University**

Research Assistant

*July, 2021–October, 2022*

Project

**Interaction between L1 transfer and universal constraints for French-speaking Mandarin L2 learners**

- Conducted thorough literature reviews for weekly seminars.
- Devised and conducted suitable statistical analyses for the corpus data.

**Prof. One-Soon Her**

**Graduate Institute of Linguistics  
National Chengchi University**

Research student

*June, 2018–July, 2020*

Project

**Classification of Paiwan questions**

- Conducted a thorough literature review.
- Conducted thorough fieldwork.
- Visualized the prosodic patterns of the collected materials.
- Authored a journal paper.

Thesis

**Production and perception of tonal coarticulation in Taiwan Mandarin and Taiwan Southern Min**

- Conducted a thorough literature review and found research gaps.
- Formulated hypotheses pertaining to the research gaps and designed experiments to validate these hypotheses.
- Independently completed three large-scale experiments.
- Devised and conducted suitable statistical analyses for the experiment data.
- Provided explanation for the experiment results.
- Authored a journal paper (under review).

Course/Personal Projects

**Ultrasonic-and-optical-imaging-assisted automated speech correctness judgment model**

**Data Analysis and Machine Learning with Python**

Lecturer: Prof. Tommy C.Y. Ho

*Fall semester, 2023*

- Conducted a thorough literature review on unfamiliar areas and found research gaps.
- Applied newly learned knowledge and skills to existing expertise; integrated distinct fields with linguistics.
- Designed a sophisticated neural network structure that combines two different forms of data, with a sound identification accuracy of 83%.

**Word networks for ill-formed versus well-formed documents in Taiwan Mandarin**

**Special Topics on Text Mining and Natural Language Processing**

Lecturer: Prof. Hsuan-Lei Shao

*Fall semester, 2023*

- Devised a pipeline to collect documents to form corpora with a static web crawler.
- Self-taught word vectorization.
- Built a word2vec model based on 1 million documents.
- Conducted a thorough literature review on unfamiliar areas and found research gaps.
- Applied word networks to graph analyses.
- Developed a novel kind of word network based on word similarity and compared its properties with conventional word co-occurrence networks.

**Simulated production and perception of tonal coarticulation through communication neural agents**

- Developed speaker/listener neural agents in tone communication.
- Self-taught related deep learning knowledge, including multilayer perceptrons and communication networks.
- Successfully induced human-like behaviors, including perceptual normalization and avoidance of coarticulation based on different tone distributions.

## **TEACHING EXPERIENCE**

**Technology of Brain Science**

Lecturers: Prof. Joshua G.O. Soo & Prof. Chia-Lin Lee

Teaching Assistant

- Supervised students' term project progress.
- Provided students with coding assistance.
- Evaluated and graded students.

**Graduate Institute of Brain and Mind Sciences**

**National Taiwan University**

*Spring semester, 2023*

## Uncovering Languages

Lecturer: Prof. Chenhao Chiu

Teaching Assistant

- Supervised students' term project progress.
- Provided students with coding assistance.
- Designed pop-up quizzes and mid-term exams.
- Supervised and led student's discussions in weekly tutorials.
- Evaluated and graded students.
- Prepared class materials.

## Freshman English

Lecturer: Prof. Wen-Hsien Hsu

Teaching Assistant

- Gave oral presentation tutorials and example presentations.
- Supervised student's mid-term and final presentations.
- Evaluated and graded students.
- Prepared class materials.

Graduate Institute of Linguistics

National Taiwan University

Spring semester, 2022/2023

Department of Foreign Languages and Literatures


National Taiwan University

Fall semester, 2020/Spring semester, 2021


## PUBLICATIONS


### Peer-reviewed Publications


#### Journal articles

**Huang, Po-Hsuan** and Her, One-Soon. 2024. On a dichotomy of question types: The case of Paiwan. *Concentric: Studies in Linguistics*, 50(1), 20–56. doi: [10.1075/cons.00033.her](https://doi.org/10.1075/cons.00033.her) 


#### Proceedings


**Huang, Po-Hsuan** and Shao, Hsuan-Lei. 2024. Applying mutual information to extract legal domain-specific collocation nouns in Mandarin. *Proceedings of the Thirtieth Annual Meeting of the Association for Natural Language Processing*, 1329–1333. Sanbi Printing. 


**Huang, Po-Hsuan** and Chiu, Chenhao. 2023. Perception of coarticulated tones in Taiwan Mandarin and Taiwan Southern Min. In: Radek Skarnitzl & Jan Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*, 446–450. Guarant International. 


Chiu, Chenhao and **Huang, Po-Hsuan**. 2023. Lip postures of high vowels in Taiwan Mandarin. In: Radek Skarnitzl & Jan Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences*, 1052–1056. Guarant International. 


#### Presentations


Chiu, Chenhao, **Huang, Po-Hsuan** and Huang, Jian-Zhi. 2024, June 24–25. Coordination of tongue and lips in Taiwan Mandarin high rounded vowels. Ultrafest XI: Conference on Ultrasound Imaging for Speech and Language Research. Aizu, Japan. 


**Huang, Po-Hsuan**. 2024, May 13–17. Production and perception of tonal coarticulation through computational communication simulation. The 13<sup>th</sup> International Seminar on Speech Production (ISSP 2024). Autrans, France. [Poster] 


Chiu, Chenhao, Chang, Cheng-Hsiang, Huang, Jian-Zhi and **Huang, Po-Hsuan**. 2024, May 13–17. Enhancing lip contrasts between /u/ and /y/ in Taiwan Mandarin. The 13<sup>th</sup> International Seminar on Speech Production (ISSP 2024). Autrans, France. [Poster] 


Chiu, Chenhao, Huang, Jian-Zhi and **Huang, Po-Hsuan**. 2023, December 4–8. Perceptual identification of high vowels in Taiwan Mandarin. Acoustics 2023 Conference. Sydney, Australia. 




**Huang, Po-Hsuan** and Liu, Yi-Chen. 2023, October 21. Ultrasonic-and-optical-imaging-assisted automated speech correctness judgment model. Tunghai International Conference on Second Language Teaching and Research (SLTR.THU). Taichung, Taiwan. 

**Huang, Po-Hsuan** and Chiu, Chenhao. 2023, May 26–27. Tonal coarticulation in Taiwan Mandarin and Taiwan Southern Min. Hanyang International Symposium on Phonetics and Cognitive Sciences of Language (HisPhonCog). Seoul, Korea. [Poster] 


Wang, Chi-Wei, Chen, Bo-Wei, **Huang, Po-Hsuan**, Lai, Ching-Hung, and Chiu, Chenhao. 2023, May 26–27. Evaluating forced alignment for under-resourced languages: A test on Squliq Atayal. Hanyang International Symposium on Phonetics and Cognitive Sciences of Language (HisPhonCog). Seoul, Korea. [Poster] 

**Huang, Po-Hsuan** and Chiu, Chenhao. 2022, November 3–4. Articulation and syllabic affiliation of prenuclear glide in Taiwan Mandarin. Ultrafest X: Conference on Ultrasound Imaging for Speech and Language Research. Manchester, United Kingdom. 

**Huang, Po-Hsuan**, Lai, Ching-Hung and Liu, Te-Hsin. 2022, September 23–25. On the interaction between L1 Transfer and universal constraints – Evidence from the acquisition of Mandarin tones by French speakers. North American Conference on Chinese Linguistics (NACCL). Bloomington, IN, United States. 

- Huang, Po-Hsuan** and Chiu, Chenhao. 2022, January 6–9. Vowel-glide distinction in high vocoid diphthong structures in Squliq Atayal: An ultrasound and acoustic study. Annual Meeting of the Linguistic Society of America (AML). Washington, DC, United States. [Poster] 
- Huang, Po-Hsuan**. 2021, October, 23. Vowel-glide distinction in high vocoid diphthong structures in Squliq Atayal: An ultrasound and acoustic study. National Conference on Linguistics (NCL). Taipei, Taiwan. 
- Huang, Po-Hsuan** and Her, One-Soon. 2021, October 15–16. On a two-way distinction of questions in Paiwan. Young Scholars' Symposium on Taiwan Languages (YSSTL). Online. 

### Works in progress

- Huang, Po-Hsuan and Shao, Hsuan-Lei. Comparison between the structures of word co-occurrence and word similarity networks for ill-formed and well-formed texts in Taiwan Mandarin. Manuscript. 
- Huang, Po-Hsuan. Improving text classification with construction embedding. In preparation.

### INVITED TALKS

---

“Electroglottography in linguistics research with Python” Institute of Linguistics, Academia Sinica, Taipei	<i>March, 2024</i>
“Face extraction with MediaPipe in linguistics research” Speech Behaviours and Speech Sciences In-class Talk, Taipei	<i>October, 2023</i>
“Face extraction with MediaPipe in linguistics research” MultiMoco NTU Workshop, Taipei	<i>November, 2022</i>
“Using PCA and LDA for ultrasound image classification” The 1 <sup>st</sup> NTU GIL Ultrasound Workshop, Taipei	<i>November, 2021</i>

### SERVICE & ACTIVITIES

---

Organizer, Education Camp for Rural Schools, NCCU	<i>February–August, 2017</i>
President of Activity, Changhua Student Association, NCCU	<i>July, 2017–June, 2018</i>
Vice organizer, Education Camp for Rural Schools, NCCU	<i>September, 2016–January, 2017</i>
Leading actor, Concours universitaire national de théâtre, Taiwan	<i>May, 2017</i>
Project manager, Education Camp for Rural Schools, NCCU	<i>August, 2016</i>
Voluntary after-school tutor, Wanhshing Elementary School, Taipei	<i>Spring semester, 2016</i>

### SKILLS & KNOWLEDGE

---

Languages: Mandarin (native); Southern Min (native); English (advanced); French (advanced); Spanish (basic)  
Programming languages/software: Python; R; Javascript; Praat Script; Psychopy; E-Prime; Matlab; SQL; bash; Swift  
Packages/modules:

- Data analyses: *numpy*; *scikit-learn*; *statsmodels*; *pandas*; *network*; *lme4*
- Visualization: *matplotlib*; *seaborn*; *Plotly*
- Image processing: *OpenCV*; *MediaPipe*; *FFmpeg*
- Acoustic analyses: *parselmouth*; *wave*
- Text mining: *jieba*; *ckip*; *gensim*
- Others: *tensorflow*; *PyTorch*; *tkinter*; *PyPI*; *pickle*; *cython*; *git*

Statistical analyses: fundamental analyses (t-test, ANOVA, etc.); dimensionality reduction (PCA, LDA, t-SNE, etc.); regression/modeling (linear/logistic regression, LMM, GLM, GAMM, etc.); cluster analysis (hierarchical; K-Means, etc.); correlation-related analyses (fundamental correlation analyses, CCA, Apriori Algorithm, mutual information, label ranking average precision scores, etc.); basic graph analyses


Machine learning: linear/logistic regression; SVM; decision tree; random forest; KNN; neural networks (CNN, spatial-transformer network, RNN, GAN, etc.); neural agent simulation, transformers and LLMs

Data structures & Algorithms: linked list, tree, array, A\* pathfinding, Levenshtein distance

Laboratory instrumentation: ultrasound; electromyography; electroglottography

### LICENSES & CERTIFICATIONS

---

Deep Learning Specialization  [\[View\]](#)

Issued by: *Coursera*