CASS-LING's Linguistic Infrastructure: Resources, Platforms and Services
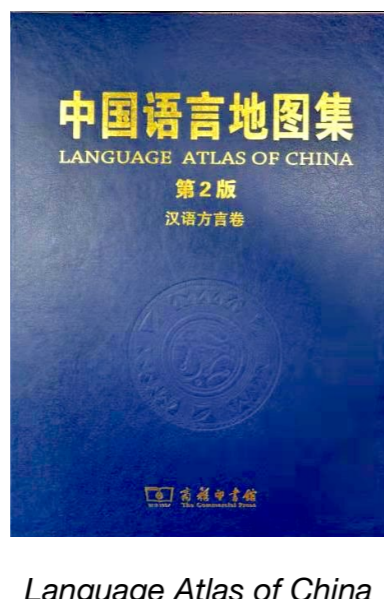
Wei WANG, Aijun LI and Danqing LIU
Institute of Linguistics, Chinese Academy of Social Sciences, China

Abstract

As the highest academic institution of linguistic research in the People's Republic of China, the Institute of Linguistics of Chinese Academy of Social Sciences (CASS-LING) has collected, compiled and created many language resources and established dedicated platforms to provide varied services, based on solid academic studies, to a society of high linguistic diversity. Such resources, platforms and services include: China's most authoritative *Contemporary Chinese Dictionary* and *Xinhua Dictionary*, world's most popular dictionary according to the Guinness World Records; the 41-volume *Dictionaries of Contemporary Chinese Dialects* and the 40-volume *Speech Archives of Contemporary Chinese Dialects*; a supportive online system for dictionary compilation; the Database of the Grammars of the Chinese Dialects with data of 10 dialects; a benchmark pronunciation model of 1.5-6 y native Putonghua speakers based on speech data collected from 4,000 plus children; a visual 3D pronunciation model for English phonetic learning based on 10,000 plus hours of speech recordings of English learners across major Chinese dialects as well as languages of non-Han ethnic groups; the state government's examination and standardization of the pronunciation of characters and words.

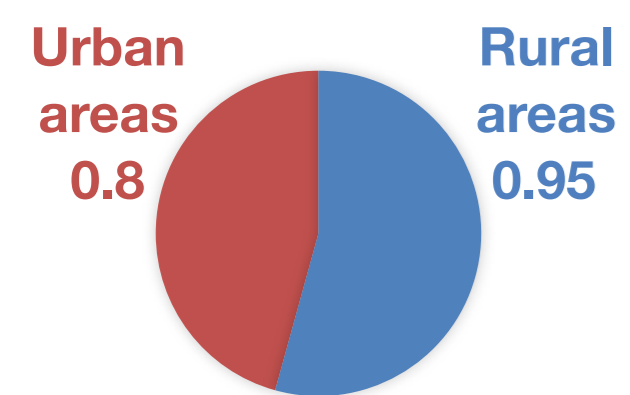## Population Distribution of Languages and Dialects in China

Language Atlas of China

Ninety-five percent of the population speaks Chinese dialects, which include ten major dialect groups: Mandarin, Jin, Wu, Hui, Xiang, Gan, Min, Cantonese (Yue), Pinghua and Hakka. The languages of non-Han ethnic groups are spoken by 5% of the population, with the largest number of speakers being the Zhuang and the smallest being the Hezhen.
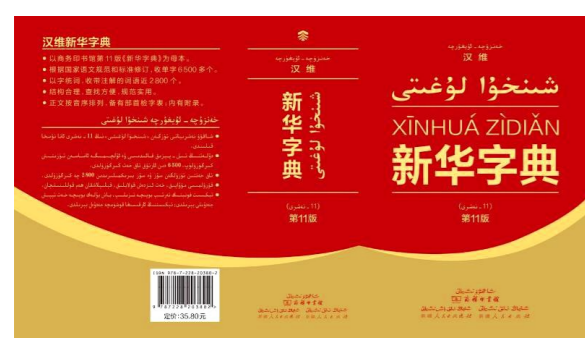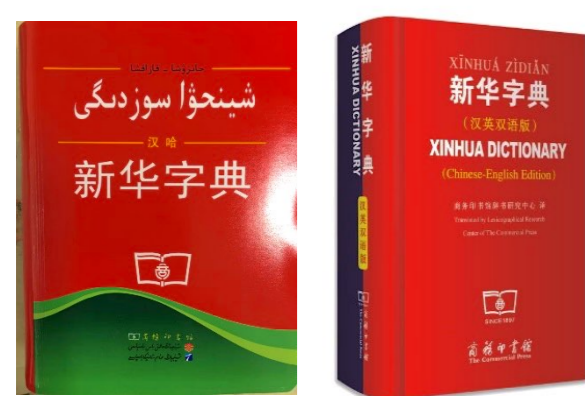
Xinhua Dictionary and *Contemporary Chinese Dictionary* in New China's Cultural Development

Illiteracy rate when New China was founded in 1949

Urban areas 0.8
Rural areas 0.95

Xinhua Dictionary

First released in December 1953, *Xinhua Dictionary* has been named after the Chinese words 新华 meaning 'a new China'. Shouldering the task of setting rules for the Contemporary Chinese and eliminating illiteracy in an old rural country then, it has been compared with a jack that has lifted the whole country's new cause of culture and education. As of July 28, 2015, its global circulation has reached 567 million copies, which scored two Guinness World Records as the "most popular dictionary" and the "best-selling book (regularly updated)."

Following the Chinese-English edition published in 2013, versions in languages of China's non-Han ethnic groups of the dictionary will also be available soon, with the Uygur and Kazakh ones being the first two.

Contemporary Chinese Dictionary

Till now a total of 70 million copies of the dictionary have been sold. While it has fulfilled its task of setting up lexical rules and enhancing the education of the nation's common language, it has also played a key role in helping overseas Chinese people to learn and use the language and to identify with their motherland.
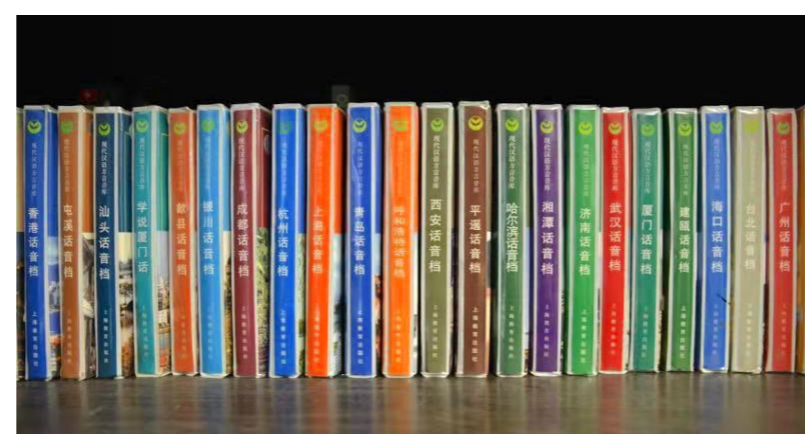

Translation of the Chinese-English bilingual version of *Contemporary Chinese Dictionary* has almost been finished and the translation of versions in Georgian, Arabic, Russian, Spanish and Persian will start soon.

The current global trend of digitization is leading our society into the era of Internet+ and artificial intelligence. The computer and smart-phone applications of *Xinhua Dictionary* and *Contemporary Chinese Dictionary* have been released respectively in 2017 and 2019 to better meet our readers' requirements of online language use.

Language Diversity: The Flourishing of Linguistic Typology Studies

A Handbook for Grammatical Investigation and Research

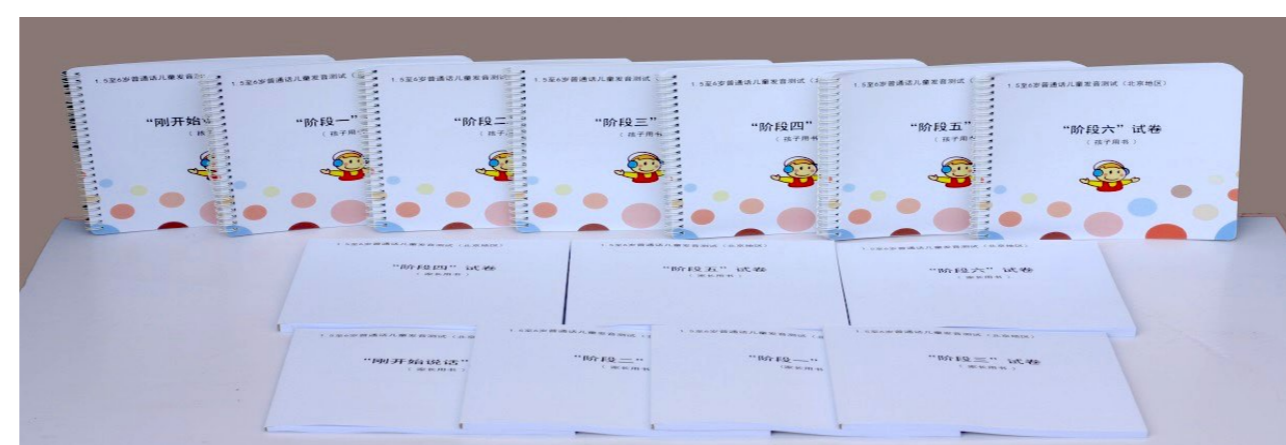The 40-volume *Speech Archives of Contemporary Chinese Dialects*

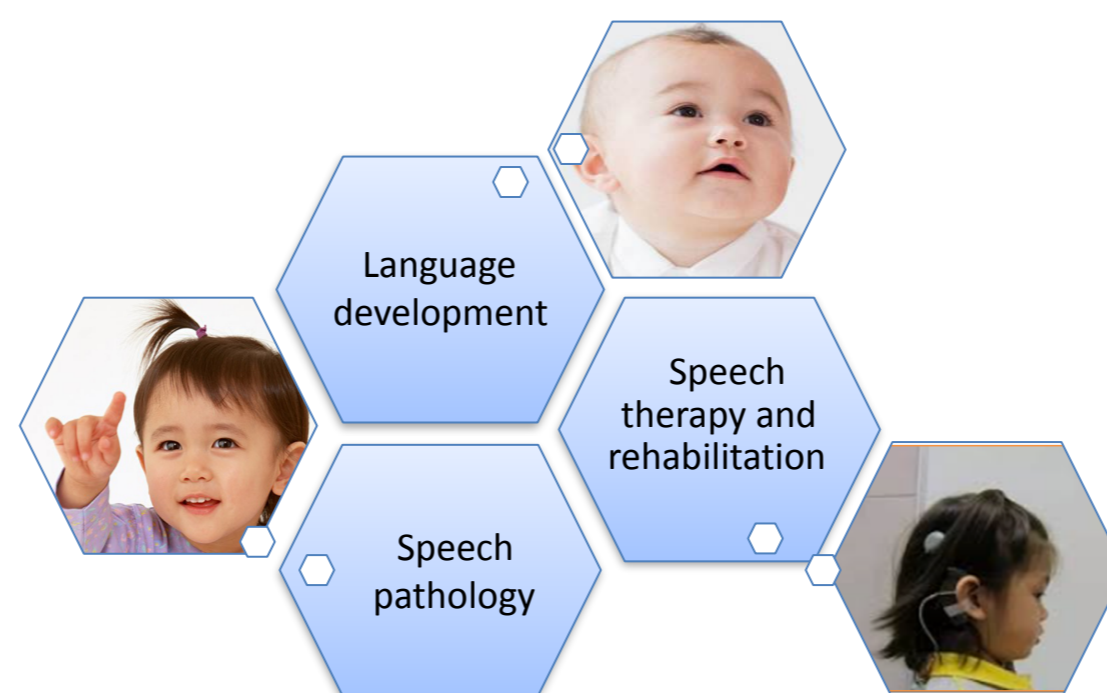The 41-volume *Dictionaries of Contemporary Chinese Dialects*

10 dialects of Mandarin, Xiang, Gan, Hakka, Wu, Yue groups, with 10 others coming.
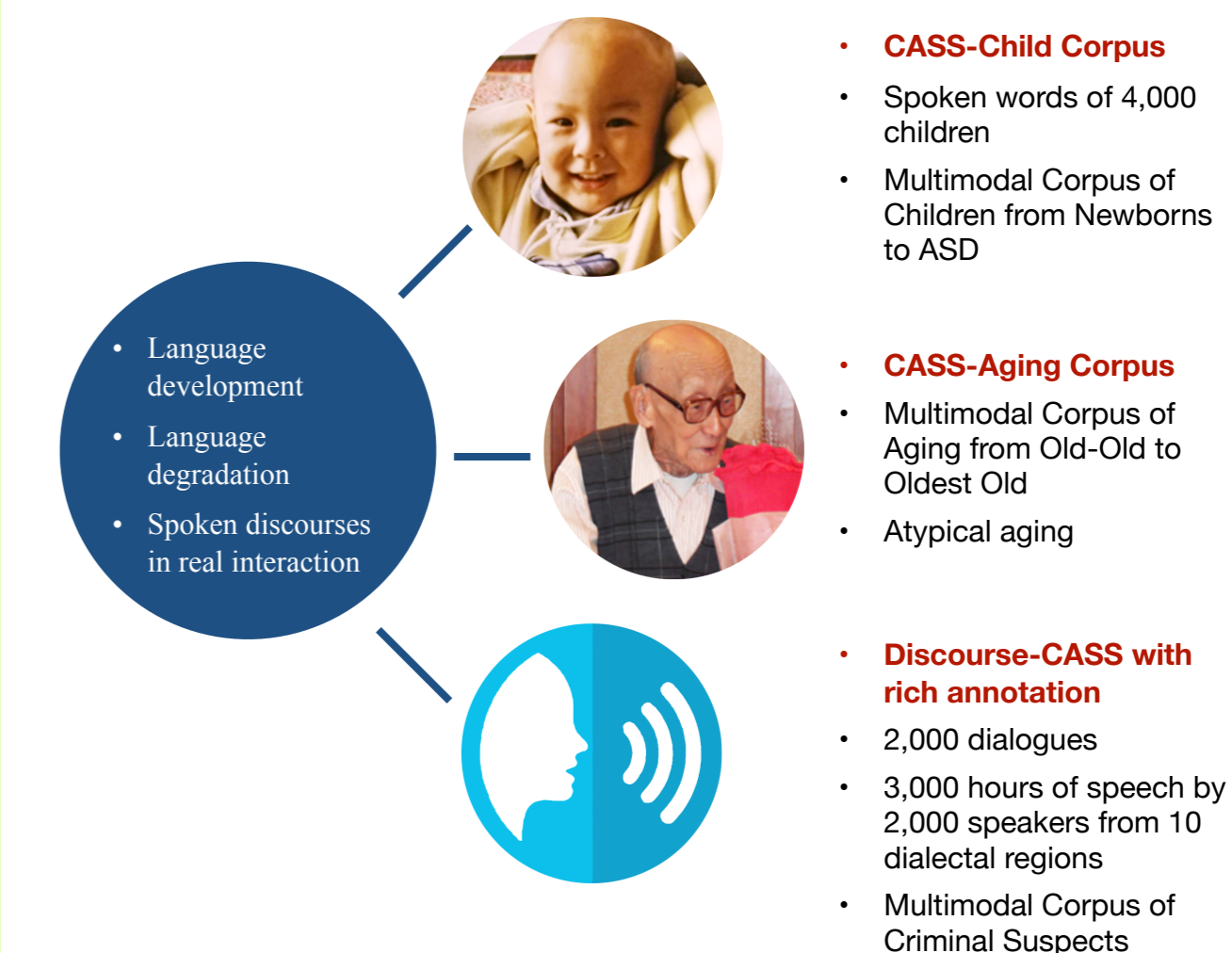
The Database of the Grammars of the Chinese Dialects

Linguistic and Cognitive Studies on Children and Aged People

Phonetic and Phonologic Development Tests (picture books for 1.5-6-year-old Chinese)

Language development
Speech therapy and rehabilitation
Speech pathology

Multimodal Corpus of Life Course from Womb to Tomb

- Language development
- Language degradation
- Spoken discourses in real interaction

CASS-Child Corpus
- Spoken words of 4,000 children
- Multimodal Corpus of Children from Newborns to ASD

CASS-Aging Corpus
- Multimodal Corpus of Aging from Old-Old to Oldest Old
- Atypical aging

Discourse-CASS with rich annotation
- 2,000 dialogues
- 3,000 hours of speech by 2,000 speakers from 10 dialectal regions
- Multimodal Corpus of Criminal Suspects

Corpus for L2 learners: AESOP-CASS, 10,000 hours English speech

Platforms and APPs

Contemporary Chinese Dictionary
现代汉语词典

Xinhua Dictionary
新华字典

Wechat Applet: **JIUZHOU YINJI**
九州音集
A network platform for collecting and demonstrating the speech sounds around the world
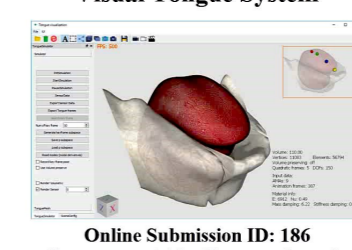
Online and offline tests for phonetic and phonological development based on Pronunciation Normal

线下发音测试软件

3D Pronunciation Training Platform

Acoustic VR in the Mouth: A Real-time Speech-driven Visual Tongue System

Online Submission ID: 186
Some parts of this video contain audio

国际音标IPA可视化发音训练
Consonants