

Successful Linguistic Validation of Clinical Outcomes Assessments: Key Considerations for Chinese Language Variants



BY TIM POESEL, PHD; BARBARA BRANDT, MA; ELIZABETH YOHE MOORE, MPH; ELIZABETH MCCULLOUGH, MA; SHAWN MCKOWN, MA; HAYLEY SIMPSON-FINCH, BA, OF RWS LIFE SCIENCES

Introduction

As the scale and frequency of clinical trials conducted in the Asia-Pacific region has increased, pharmaceutical and life sciences companies have encountered challenges in selecting the appropriate written variants of Chinese in a variety of countries. Selecting the appropriate variant is particularly important when translating and linguistically validating Clinical Outcomes Assessments (COAs) and Patient Reported Outcomes (PROs), which are used to directly collect patient data. To help provide clarity, the RWS Life Sciences Linguistic Validation team analyzed the written and spoken variants of Chinese and their applicability across countries and regions. By providing a detailed discussion of both the history and current usage of each variant, we can recommend best practices by country.

Objective

There are several predominant variants of both written and spoken Chinese. The dominant written variants are Traditional and Simplified Chinese, and well-known spoken variants include Mandarin, Cantonese, and Taiwanese Hokkien. Although

these variants are related, they do not have perfect mutual intelligibility, and their use is strongly influenced by historical, political, and geographical factors. As a result, understanding the relationship between these written and spoken variants, and the appropriate circumstances for selecting one variant over another, is vital to successful translation and linguistic validation of COAs targeted for use in populations that speak and write modern Chinese.

Background

Modern spoken Chinese is divided into at least seven major dialectal groups containing over 200 dialects. Although many of these spoken dialects are not mutually intelligible, historically, there has been a common written system underlying the spoken variants. Written Chinese is composed of logograms, or characters consisting of many individual strokes. The majority of these characters consist of two distinct parts: a semantic component (the “radical”) conveying general meaning; and a phonetic component suggesting a pronunciation, although written Chinese is not an alphabetic script ^[1].

TABLE 1: COMPARISON OF TRADITIONAL AND SIMPLIFIED CHINESE CHARACTERS

TRADITIONAL CHARACTER	SIMPLIFIED CHARACTER	TRANSLATION	SIMPLIFIED BY:
麼	么	“What”	removing the radical
個	个	“One”	retaining only radical
樂	乐	“Fun”	using a printed vs cursive form
幹	干	“Dry”	replacing a character with one that sounds similar
塵	尘	“Dust”	adopting historical variants

This common written system officially diverged in the 1950s into two variants, Traditional and Simplified Chinese. In mainland China, development and use of Simplified Chinese was promoted by the People’s Republic of China, one goal being to increase literacy by greatly simplifying the forms of the most common Chinese characters so they would be easier to write and learn. Simplified Chinese is used and taught across the majority of mainland China, and is also the dominant variant in Singapore and Malaysia, with the worldwide number of users currently estimated at more than 1.3 billion. Traditional Chinese, a retronym referring to the system of writing that existed prior to the establishment of Simplified Chinese, was retained as the dominant written form in Taiwan, Hong Kong, and Macau, and is also commonly found in overseas Chinese communities, with a worldwide number of users estimated at 50 million.

Traditional Chinese characters are more complex than their Simplified counterparts, on average having 30% more strokes among common characters, and nearly twice as many strokes (100% more) for less common characters (see Table 1 for a comparison of words written with each variant)^[2]. Although Traditional Chinese is considered more difficult to learn, users contend that it better preserves the rich history and traditions of the Chinese written language, with this perceived loss being one among many tensions that exist between users of the two variants^[3, 4].

While the total number of modern Chinese characters appearing in comprehensive dictionaries is numbered at or above 80,000, Traditional Chinese frequently uses approximately 4,800 characters, with another 6,300 less common characters, whereas Simplified Chinese is composed of approximately 2,500 frequently used characters, with another 1,000 that are less common. Traditional and Simplified Chinese share a subset of these characters, which many can recognize and interpret. However, formal education is typically given in only one of the written variants, which have additionally experienced 60 years of separation as a result of political tensions and geography. Thus, although the variants are related and many speakers encounter both in daily life, each has undergone natural changes in usage and vocabulary that increase the difficulty of interpretation across the forms^[5, 6].

The widespread use of technology in education and daily



Figure 1 Countries and regions colored according to the dominant variant of written Chinese

life has further affected how speakers interact with and understand the written systems. Because 2500+ characters cannot be encoded on a traditional keyboard, speakers use an alphabetic system (i.e., consisting of letters that represent sounds) to encode the words they want to type (pinyin for Simplified Chinese and zhuyin for Traditional). Speakers are next presented with several characters that closely match the alphabetic input, and make the appropriate selection. As a result of reliance on these alphabetic systems, knowledge of how to write Traditional and Simplified Chinese is eroding in some populations, and may further complicate speakers’ ability to transfer their knowledge of one written variant to the other.^[7]

Further Considerations

There are at least two other Chinese written variants. One of these is specific to Hong Kong and Macau, and is a written form of Cantonese. As vocabulary differences between Cantonese and Standard Mandarin are estimated to be as high as 30%, written Cantonese diverges dramatically from both the Traditional and Simplified Chinese variants, and appropriate consideration of

TABLE 2: SPOKEN AND WRITTEN CHINESE VARIANTS BY COUNTRY

COUNTRY	MAJOR SPOKEN VARIANTS	DOMINANT WRITTEN VARIANT
China	Mandarin (70%); Min (6.5%); Wu (6.5%); Jin (5%); Cantonese (5%)	Simplified Chinese
Malaysia	Cantonese (27%); Mandarin (18%); Taiwanese Hokkien (5%)	Simplified Chinese*
Singapore	Mandarin (35%); Other Chinese dialects (12%)	Simplified Chinese*
Taiwan	Mandarin (100%) (lingua franca) Taiwanese Hokkien (70%); Hakka (2%)	Traditional Chinese*
Hong Kong	Cantonese (87.5%) Native Yue dialects (3.5%) Mandarin (1.5%)	Traditional Chinese**
Macau	Cantonese (86%); Other Chinese Dialects (7%) Mandarin (3%)	Traditional Chinese**

* Taiwanese Hokkien may also be appropriate (See Further Considerations)

** Written Cantonese may also be appropriate (See Further Considerations)

Table 2: The languages spoken in each country by percentage of population, along with the dominant written form^[10].

Simplified & Traditional Chinese: Users Worldwide

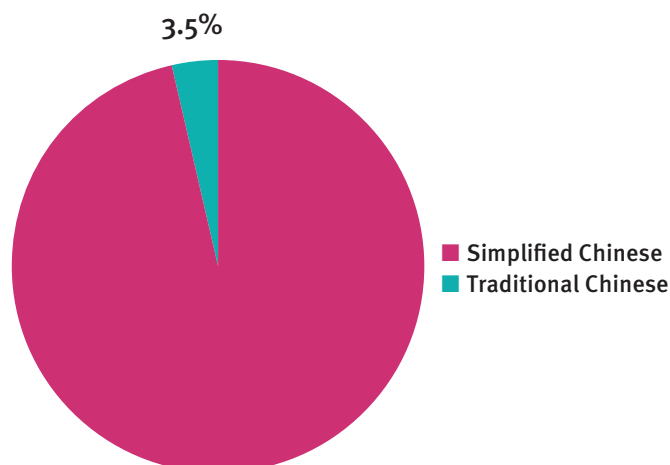


Figure 2: Simplified Chinese is the dominant variant of written Chinese by percentage of users. Traditional Chinese, though used by far fewer people, is nonetheless the dominant variant in Taiwan, Hong Kong, and Macau. Linguistic validation should consider both the location and context of a study in determining the proper language(s) to request.

both context and target population should be exercised when selecting this written variant for translation^[8]. An additional written form of Hokkien exists, and is used in Taiwan, the Fujian province of mainland China, and in areas of Malaysia and Singapore. Written Hokkien is still an unstandardized system,

with many variations across communities and places of use, and is typically used for informal and non-administrative purposes. All the same, written Hokkien is regularly encountered in novels, songs, and other media, and may be considered for use in the appropriate context and location^[9].

Conclusions

There are important differences between the Traditional and Simplified variants of written Chinese, as well as geographic and national variation in their use. Most formal education occurs in only one variant, and although related, their separation across time has allowed for a divergence in their evolution and a corresponding reduction in mutual intelligibility. There is little systematic relationship between spoken variants and their underlying written forms, and additional written variants specific to certain regions (e.g., written Cantonese and Hokkien) further complicate the process of selecting a language for translation. Together, these factors necessitate careful consideration of the written variant and the subject population requested for translation services and linguistic validation of COAs in locations where modern Chinese is used. Furthermore, these factors suggest that the development of “worldwide” translations, intended for speakers across multiple countries, may be difficult in the case of modern Chinese, and that a country or population-specific translation may provide the best results.

Linguistic Validation

Linguistic Validation is a process conducted to confirm that a Clinical Outcome Assessment (COA) questionnaire is acceptable for use in different languages and in different cultural contexts.

Both the harmonized translation and the English back translation are reviewed by a project manager and a survey research analyst, and adaptations to the translation are made as needed.

Without this careful development of a translation and subsequent cognitive debriefing, one cannot be reasonably certain that the adapted instrument is both conceptually equivalent to the original and clearly understood by the average patient.

The linguistic validation process begins with two translators independently translating the instrument into the target language. The translators then exchange drafts and work together to develop one reconciled or “harmonized” version. At that point, the harmonized translation is provided to a third translator who translates the text back into English without access to the original English. Both the harmonized translation and the English back translation are reviewed by a project manager and a survey research analyst, and adaptations to the translation are made as needed. Once the final translation has been approved, it is debriefed with a sample of in-country native speakers of the language, with varying demographic and educational backgrounds, to check for conceptual equivalence and clarity.

References

1. Kurpaska, M. (2010). Chinese language(s): A look through the prism of The Great dictionary of modern Chinese dialects (Vol. 215). Walter de Gruyter.
2. 郭曙纶. (2009). Dynamic Statistics and Comparison of the Number of Stroke and Traditional Chinese Characters. Journal of Beihua University (Social Science Edition), 2, 50-56.
3. Keller, A. (1997). Vernacular Literacy: A Re-Evaluation. Oxford University Press. ISBN 0-19-823635-2.
4. Wilkinson, E. (2012). Chinese History: A New Manual. Harvard-Yenching Institute Monograph Series 84. Cambridge, MA: Harvard-Yenching Institute; Harvard University Asia Center. ISBN 978-0-674-06715-8.
5. Norman, J. (1988). Chinese. Cambridge: Cambridge University Press. ISBN 978-0-521-29653-3.
6. Ramsey, S. (1987). The Languages of China. Princeton University Press. ISBN 978-0-691-01468-5.
7. Leung, C., & Ruan, J. (Eds.). (2012). Perspectives on Teaching and Learning Chinese Literacy in China (Vol. 2). Springer Science & Business Media.
8. Snow, D. Cantonese as Written Language: The Growth of a Written Chinese Vernacular. Hong Kong University Press, 2004.
9. Lu, J., Liu, C. “Simplified vs. Traditional Chinese: What Every Translation Agency Should Know.” The ATA Chronicle (2002): 31.
10. Australian Bureau of Statistics, Census of Population and Housing 2006 and 2011

NOTE: RWS Life Sciences presented this information in poster format at the International Society for Pharmacoeconomics and Outcomes Research International Meeting, November 2017, in Glasgow, Scotland.

About RWS Life Sciences



Under the umbrella of RWS, the world’s leading provider of global language solutions, RWS Life Sciences focuses exclusively on providing quality-driven translations for clients in the life sciences industry. We specialize in language support solutions for highly regulated, global markets in areas including clinical, regulatory, medical device, pharmacovigilance, health economics, outcomes research, and product labeling. Through our innovative technology platforms, we provide process automation, scalability, and business intelligence to serve our clients’ needs in fast-paced and demanding environments.

RWS Life Sciences is a leading and trusted authority on the linguistic validation of Clinical Outcomes Assessments (COA). Whether intended for a patient (PRO), clinician (ClinRO), or observer (ObsRO), our translations are accurate as well as culturally and conceptually equivalent to the source instrument. Our experience in translating COAs has expanded across a variety of therapeutic areas, including cardiovascular, allergy/respiratory, oncology, gastroenterology, inflammation, neurology, infectious diseases and vaccines.

Our state-of-the-art linguistic validation process and COA added-value services ensure both accuracy and timely completion of your documents. We model our linguistic validation process on the U.S. Food and Drug Administration PRO guidance document and the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Good Practice recommendations. We proudly deliver exceptional customer satisfaction, 99% on-time delivery and 98% first-pass yield.

Learn more about RWS Life Sciences at www.rws.com.