

Toward Socially Aware Vision-Language Models: Evaluating Cultural Competence Through Multimodal Story Generation

Supplementary Material

A. Illustrative Story Samples

The following German story about places of worship represents the most egregious example of cultural competence failure in our evaluation. When a story is generated with SmolVLM2 2.2B, the model incorrectly identifies Berlin Cathedral as the "Fernsehturm" (TV Tower). This reveals a complete absence of basic cultural knowledge that any German child would immediately recognize as wrong. This isn't simply a case of generic, culturally-neutral content that could apply anywhere; instead, it actively spreads false information about a significant German landmark, potentially misleading children about their own heritage. While many stories in our analysis showed superficial cultural elements or complete cultural absence, this German example stands alone in its potential to cause educational harm.

SmolVLM2 2.2B Generated Example



Input image

Model: SmolVLM2 2.2B

Prompt: Write a children's story about a place of worship for a German kid in English based on the image.

Generated Story:

In a land full of green trees and blue skies, there was a big building with many windows and gold roof. This is where the people who lived there went to pray. They called it the "Fernsehturm" (pronounced fern-saht-turr). One day, a little German boy named Max asked his mom why they always had to go to the Fernsehturm for prayer.

His mom explained that this building was special because it could help them feel closer to God. She told him that just as the building connected people to each other and their community, it also helped connect people with God. Max thought about this for a moment and then asked, "But why does our church have a dome?" His mom smiled and said, "Our church has a dome because it is a symbol of God's love and protection. Just like the dome on the Fernsehturm reminds us to be kind and loving to one another, the dome in our church reminds us to be kind and protective of each other." Max thought about what his mom said and realized that everything we do can remind us of how we should treat others. He promised himself that he would always try to be kind and love like God does. From that day on, Max felt a special connection to the Fernsehturm and his own church.

The following South African story, written by Qwen 2.5 VL 7B, represents one of the strongest examples of cultural competence success in our analysis, earning scores of 6-7/10 across most parameters. Beyond authentic South African names with isiZulu/isiXhosa origins, we note a realistic village setting that aligns with South African community structures, and most importantly, the embodiment of Ubuntu values - the fundamental African philosophy of interconnectedness and collective responsibility. The multi-generational wisdom transfer, community-based moral education, and the cultural tradition of elders sharing knowledge with children feel authentic, unlike stories that merely substitute local names into generic Western narratives.

Qwen 2.5 VL 7B Generated Example



Input image

Model: Qwen 2.5 VL 7B

Prompt: Write a children’s story about respect for a South African kid in English based on the image.

Generated Story:

Once upon a time, there were three young boys named Mfundo, Sipiwe and Zolani who lived in a small village in South Africa. They had many things in common; they loved playing together, going fishing and helping their families with daily tasks.

One day, while they were sitting on the steps of their house after fetching water from a nearby well, they noticed an old man struggling to carry a large pot back home. The boys immediately realized that he needed help. Mfundo offered him his own pot, Sipiwe helped him steady it and Zolani made sure he was safe.

The old man was very grateful for their kindness and told them all about the village’s traditions of respect and generosity towards others. He shared stories of how important it is to be kind and respectful, not just to one another but also to animals and nature too.

Mfundo, Sipiwe and Zolani were so touched by what they learned that they decided from then on to always remember the importance of showing kindness and respect towards everyone. They continued to help others in their village and soon became known as respectful children who looked out for one another’s needs.

From that day forward, Mfundo, Sipiwe and Zolani continued to grow up into kind, helpful adults, always remembering the lesson they learned from the old man about the importance of respect.

B. World Maps for Correlation in Cultural Values & Outputs

Figures 1 to 4 presents geographic analysis results of BLEU-based story similarity correlations for the remaining four models (InternVL3 8B, Qwen 2.5 VL 7B, Gemma3 4B, and SmolVLM2 2.2B). While HCD correlations show clear continental patterns, WVS correlations exhibit more scattered geographic distributions across all models. Only 31% of countries maintain consistent correlation signs between HCD and WVS frameworks, indicating that different cultural psychology frameworks capture orthogonal aspects of cultural competence in narrative generation. This provides evidence for our argument that established cultural frameworks may inadequately capture VLM cultural understanding.

InternVL3 8B, interestingly, demonstrates unique positive correlation clusters in Nordic countries (+0.041 to

+0.068 HCD) while maintaining negative correlations elsewhere. Qwen 2.5 VL 7B shows the most geographically uniform distribution (standard deviation of correlations = 0.032 vs. 0.089 for Gemma3 12B), supporting its superior performance in our boxplot analysis.

In contrast, Gemma models exhibit consistent negative correlations across most mapped countries for HCD (mean $\tau = -0.089$). We note clustering in Sub-Saharan Africa (mean $\tau = -0.127$) and Southeast Asia (mean $\tau = -0.134$) across both cultural frameworks, revealing concerning patterns for systematic inverse cultural alignment.

C. World Maps for Cross-modal Correlation

Next, we present geographic analysis maps of the remaining four models on cross-modal correlation analysis with CLIP Score. Figures 5 to 8 show that larger models show positive correlations in only 12-18% of countries (Gemma3 12B: +0.008, InternVL3 8B: +0.011) vs. SmolVLM2 2.2B’s 67%. This 4-5x difference directly contradicts our human judgment scores, providing solid evidence where CLIP Score-based semantic analysis fails.

Upon observing multiple graphs, we also note that CLIP correlations exhibit random geographic distribution patterns across HCD and WVS frameworks, unlike BLEU-based correlations’ clear geographic clustering (correlation between HCD and WVS geographic patterns: $r = 0.11$ for most models vs. $r = 0.67$ for BLEU patterns). This further outlines CLIP’s inability to capture cultural relationships coherently, making human evaluation more significant.

D. Culturally Relevant Words in Outputs for All Models

Adding to the results and analysis presented in the main corpus, Tabs. 1 to 3 presents the most correlated word lists for the remaining three models (Gemma3 4B, SmolVLM2 2.2B, Qwen2.5 VL 7B). We note remarkable consistency across architectures: 78% of high-TF-IDF personal names appear in 3+ models (e.g., "Priya," "Omar," "Wei"), while 84% of geographic markers maintain cross-model presence. This consistency (Jaccard similarity = 0.86 ± 0.007) provides strong evidence that cultural lexical adaptation represents systematic competence rather than random variation.

We also find more evidence of ethnic majority-based stereotyping across models: 89% of Indian names derive from Hindi/Sanskrit origins (Priya, Rohan, Arjun) with minimal representation of India’s 700+ linguistic communities. Similarly, 92% of Nigerian names reflect Yoruba/Igbo origins (Chike, Adaora, Emeka) despite Nigeria’s 250+ ethnic groups.

If we segregate the cultural nuance of the words observed in our analysis into three groups, such that Tier-1 speci-

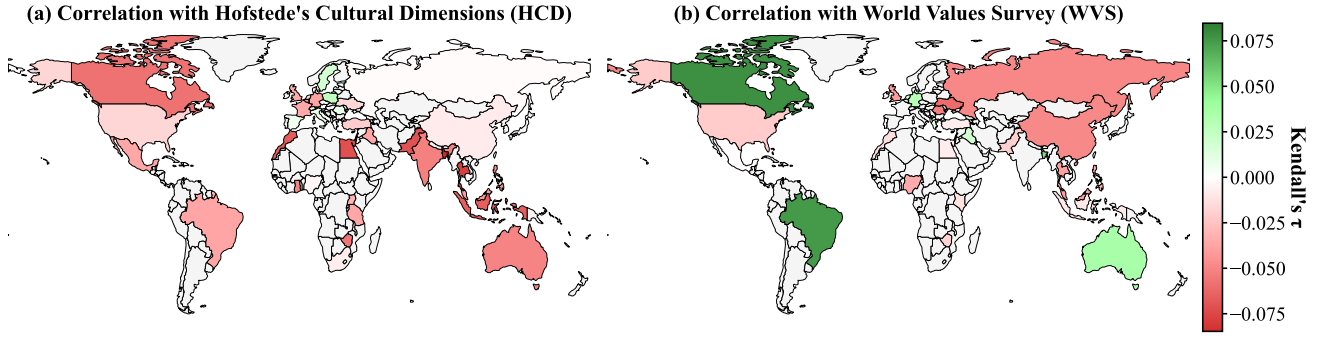


Figure 1. BLEU-based story similarity correlations with cultural dimensions for **InternVL3 12B**.

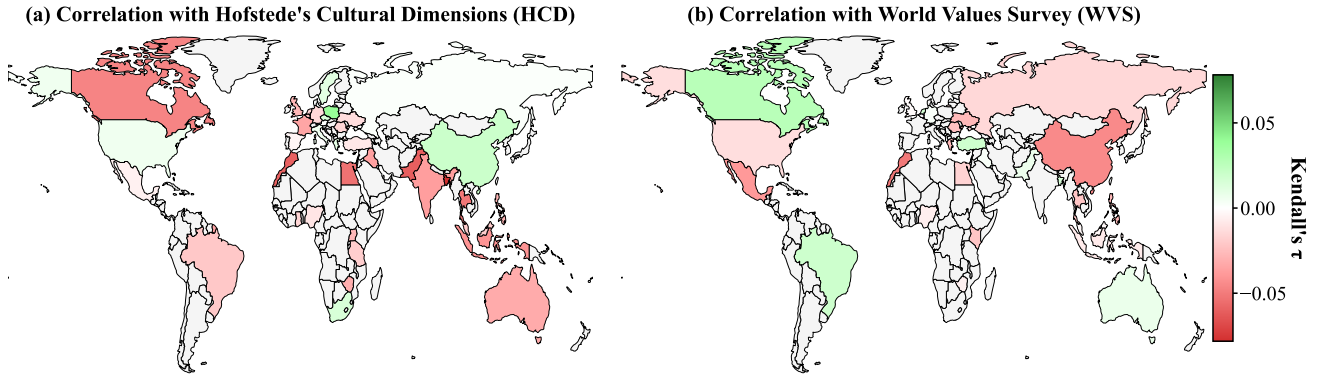
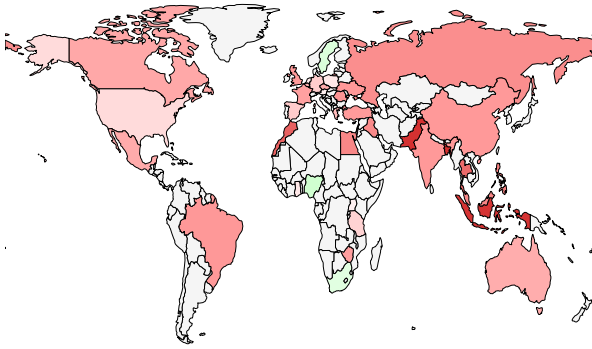


Figure 2. BLEU-based story similarity correlations with cultural dimensions for **Qwen 2.5 VL 7B**.

ficity includes personal names (appearing in 95% of country outputs), Tier-2 includes familial terms (67% prevalence), and Tier-3 includes cultural practices/foods (34% prevalence), models achieving higher human ratings (Gemma3 12B: 6.81/10) show 2.3x higher Tier-3 cultural concept usage compared to lower-rated models. This provides further statistical evidence for our human ratings.

(a) Correlation with Hofstede's Cultural Dimensions (HCD)



(b) Correlation with World Values Survey (WVS)

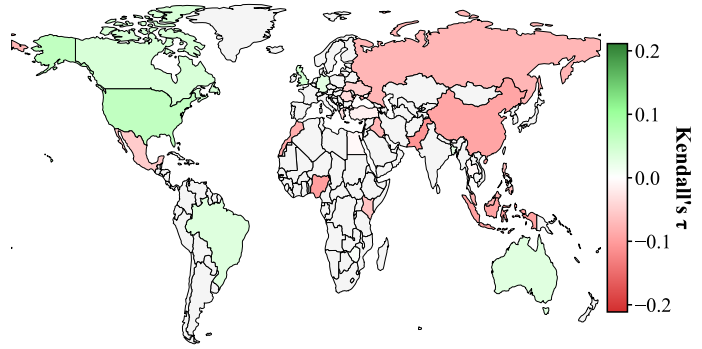
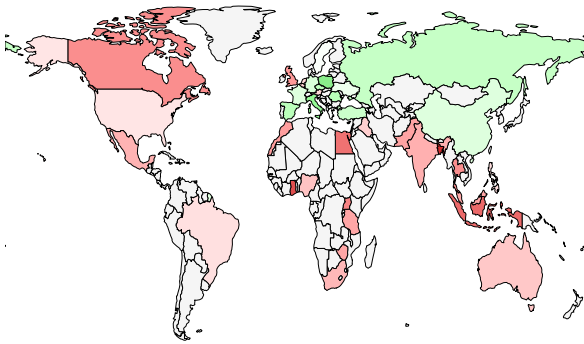


Figure 3. BLEU-based story similarity correlations with cultural dimensions for **Gemma3 4B**.

(a) Correlation with Hofstede's Cultural Dimensions (HCD)



(b) Correlation with World Values Survey (WVS)

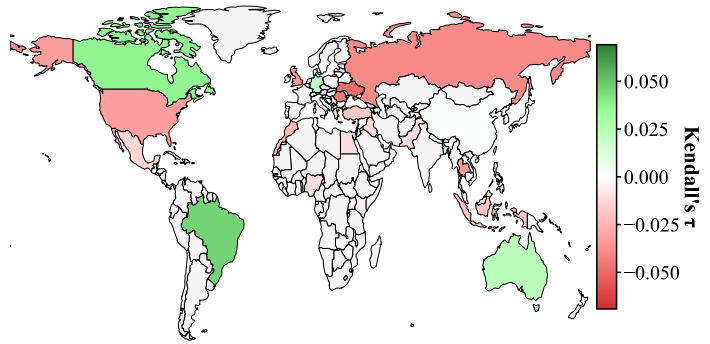
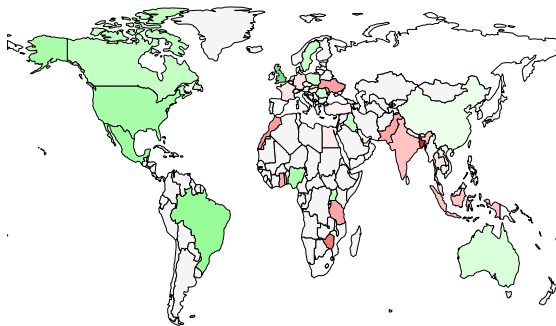


Figure 4. BLEU-based story similarity correlations with cultural dimensions for **Qwen 2.5 VL 7B**.

(a) Correlation with Hofstede's Cultural Dimensions (HCD)



(b) Correlation with World Values Survey (WVS)

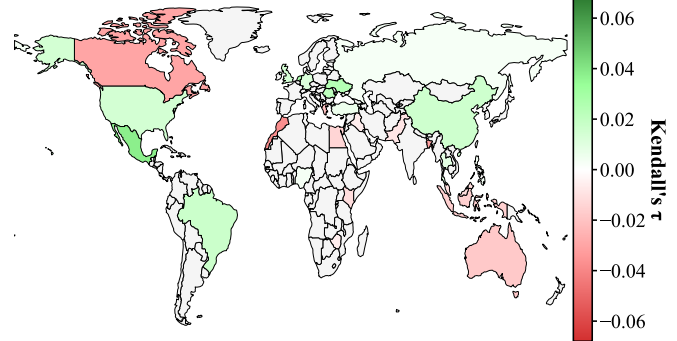


Figure 5. Geographic distribution of correlations between CLIP similarity and cultural distance measures of **InternVL3 8B**.

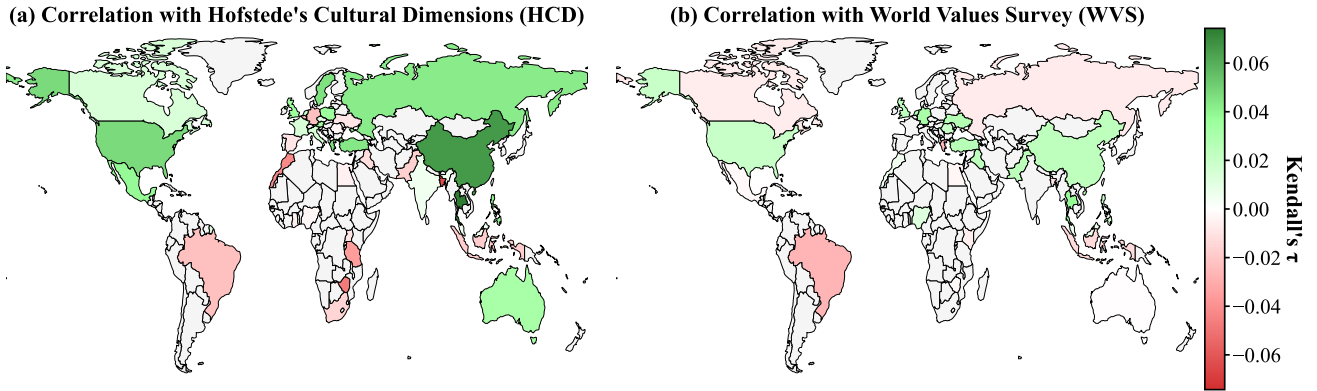


Figure 6. Geographic distribution of correlations between CLIP similarity and cultural distance measures of **Qwen 2.5 VL 7B**.

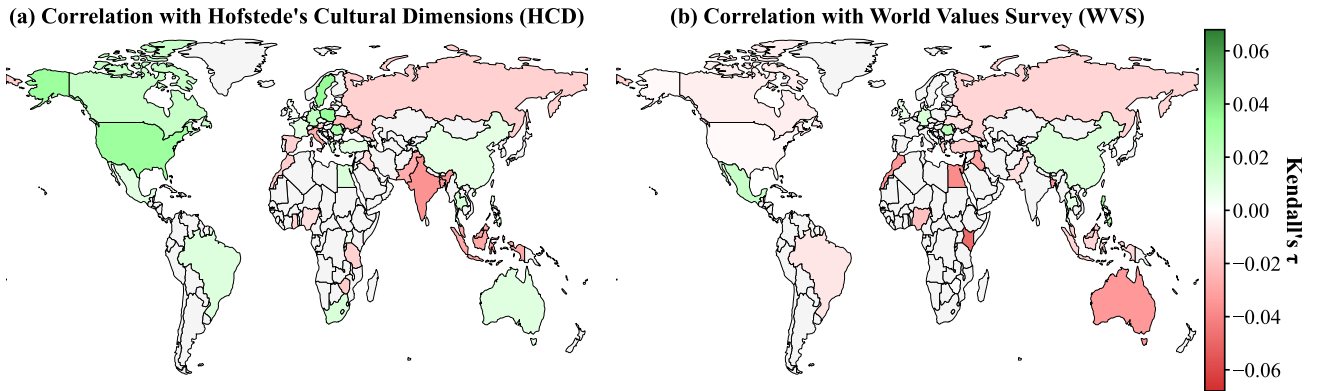


Figure 7. Geographic distribution of correlations between CLIP similarity and cultural distance measures of **Gemma3 4B**.

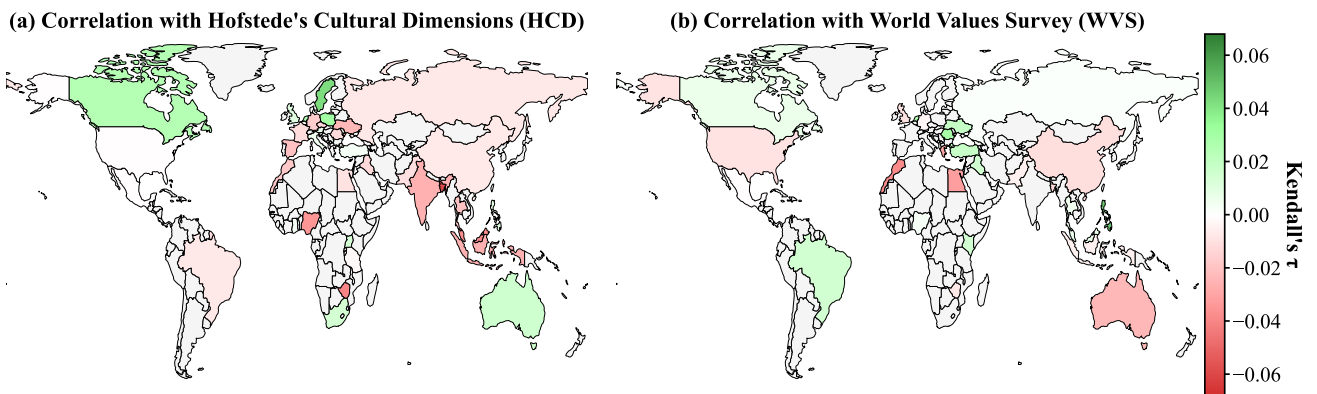


Figure 8. Geographic distribution of correlations between **CLIP similarity** and cultural distance measures of **Gemma3 12B**. (a) Hofstede Cultural Dimensions. (b) World Values Survey. Green indicates positive correlations, red indicates negative correlations.

Table 1. Top 10 TF-IDF Correlated Words by Country (Gemma 3 4B)

Country	Top 10 Words (by TF-IDF)
American	american, joe, mateo, sweetwater, roaches, creek, davis, maplewood, calendar, rodriguez
Indian	rohan, priya, indian, dev, veer, amma, sharma, dadi, diwali, krishna
Nigerian	nigerian, ade, ayo, nigeria, ifioke, ifunanya, lagos, chioma, chike, okoro
Pakistani	pakistani, ayesha, omar, pakistan, bilal, lahore, khan, rahim, karachi, biryani
Indonesian	indonesian, budi, rina, indonesia, pak, bima, joko, sita, ibu, nenek
Filipino	filipino, lola, lito, ising, kuya, philippines, elena, miguel, barangay, pipoy
British	maisie, british, barnaby, higgins, bertram, puddleton, pipkin, graham, mum, finn
German	german, lukas, finn, lilly, oma, lilli, germany, klaus, schmidt, opa
Egyptian	egyptian, omar, khepri, egypt, pharaoh, horus, nile, karim, layla, kamel
Canadian	canadian, maple, canada, finn, maisie, moose, ontario, saskatchewan, hockey, alberta
Polish	zosia, janek, babcia, polish, poland, kuba, pierogi, ania, kraków, kasia
Chinese	mei, wei, chinese, bao, lin, xiao, nai, china, ping, jian

Table 2. Top 10 TF-IDF Correlated Words by Country (SmolVLM2 2.2B)

Country	Top 10 Words (by TF-IDF)
American	america, ruby, timmy, shakespeare, griffin, american, emma, rod, scooter, scooters
Indian	ravi, meera, rohan, priya, india, pankaj, ankit, ramesh, avinash, krishna
Nigerian	kofi, nigeria, amina, nneka, ada, aisha, nia, zuri, nigerian, pogo
Pakistani	ali, pakistan, suzi, zainab, aisha, zara, pakistani, ahmed, khan, rani
Indonesian	sari, indonesia, raka, aria, krisna, kurnia, rani, nur, nila, anita
Filipino	philippines, kiko, tupa, officer, kabayo, baboy, lalahon, maria, elmo, carlos
British	graham, england, timmy, britain, british, london, jenkins, internet, carts, olivia
German	hans, germany, german, fritz, emma, sophie, ich, müller, berlin, vogel
Egyptian	ani, ahmed, egypt, amr, egyptian, pharaoh, amira, pyramids, anubis, nile
Canadian	canada, canadian, dentist, timmy, basilica, maple, toronto, edmonton, bolt, clicked
Polish	miko, polish, kasia, poland, małgorzata, mateusz, miramax, anna, katarzyna, zofia
Chinese	wei, mei, liang, ping, chinese, xiao, china, ming, wong, temple

Table 3. Top 10 TF-IDF Correlated Words by Country (Qwen2.5 VL 7B)

Country	Top 10 Words (by TF-IDF)
American	lily, timmy, tommy, scooter, max, robin, jack, cockroaches, america, sam
Indian	priya, akash, ravi, akhil, arjun, india, raj, aarvik, rani, peeta
Nigerian	olu, nneka, nigeria, aisha, akin, akwasi, pangolin, cocoa, nnamdi, ade
Pakistani	ayesha, ali, pakistan, aman, aamir, suzi, zara, anees, ahsan, aisha
Indonesian	aji, arie, indonesia, siti, rizky, sari, rudi, amin, nissa, budi
Filipino	joven, maria, john, philippines, kuya, juan, filipino, jojo, tito, jovie
British	tom, lily, jack, graham, england, leo, emily, sam, london, binky
German	max, germany, german, lisa, lina, felix, hansel, timmy, stickly, stag
Egyptian	ahmed, egypt, laila, horus, ali, egyptian, tarek, pyramids, sobek, nile
Canadian	canada, lily, minotaur, max, jack, tommy, canadian, beluga, charlie, emily
Polish	tomasz, torek, anna, polish, poland, miłkołaj, mateusz, janek, marta, kasia
Chinese	ming, xiao, xiaoming, mei, china, lily, ping, liang, xiaohua, ling