Semantic and Expressive Variation in Image Captions Across Languages

Supplementary Material

6. Related work

Existing work in **multilingual multimodal modeling** investigates how vision-language models can perform better across a variety of languages. Many previous works have proposed methods to build non-English and multilingual models for specific vision tasks such as captioning, question answering, and retrieval [21, 32, 40, 45, 97]. To benchmark and build more multilingual models, many multilingual vision datasets have been introduced [6, 41, 64, 72, 78]. Many more recent large vision models are trained to be multilingual [19, 42, 51]. These models have been probed for biases across language and associated cultures [3, 34, 44]. To better measure and counteract these biases, vision datasets have been built which include images captured from diverse geographical regions around the world [7, 76, 77, 104], and which create diverse visual knowledge by annotating images with culture- and region-specific information, such as identifying regional dishes, dresses, and ideas [39, 74, 83, 127]. We build upon this rich lineage of multilingual vision work: rather than seeking to *propose new multilingual datasets* (which may offer new concepts) or *expand vision models' capabilities on non-English languages*, we seek to demonstrate that *multilingual vision datasets and models may already exhibit meaningful information differences across languages*.

Our inquiry is inspired by **research in cross-cultural Psychology**. Psychologists, anthropologists, and philosophers provide strong evidence that salient visual features differ systematically across cultures and languages, with broadly ranging studies including cross-cultural psychology [62, 85], usage-based linguistics [69, 117], organizational principles of perception and cognition [61, 123], and the cognitive realities of one's perceptual experience [43, 49, 110, 124]. Cognitive linguistics posits direct ties between what we say to the way we think and perceive the world [10, 66, 67]. They further suggest that expressed meaning depends on not only *what* is said — the *semantic content* of what we say, but also equally on *how* we say it — the very *manner of expression* we choose to say it (e.g., specificity of word choice, tone and mood of expression) [11, 69, 120]. That is, a speaker's conceptualizations has direct influence over what linguistic features or words they reach for and how put them together when they formulate our thoughts into words [28, 68, 109]. This leads us to study both *semantic* and *expressive* variation of captions across languages, e.g. in §2.2 and §2.3.

In general, **human-centric approaches** to computer vision center around considerations of human abilities and limitations in the development of models and applications. For example, methods highlighting saliency attempt to identify which image regions and features people find most important [8, 31, 107, 113, 128]. User-centric vision modeling adapts the models to user-specific preferences and knowledge [24, 98, 108, 118]. Similarly, our work looks closely at the differences between populations of humans "behind" multilingual vision datasets (and downstream multilingual vision models).

7. Limitations

Our chosen 7 languages. Our selection of languages is diverse but not representative of global linguistic diversity.
Mid-sized scale. Our experiments operate at a mid-sized scale (thousands of images), emphasizing breadth in languages over depth in images. Future studies may forego such a wide exploration to investigate more specific phenomena at a larger image scale, such as if models differ in their image understanding when trained on captions from different languages. Previous works have shown promise in this direction by showing how better-quality, denser, and more diverse captions can help with better image understanding [65, 84].
Risk of linguistic essentialism. Categorizing differences solely with languages may pose a risk of essentializing or stereotyping them, suggesting that all members that speak a language describe the world similarly. We emphasize that we do not make *categorical* but rather *distributional* claims, aiming to show general differences across a large set of samples.

8. Experimental Details

8.1. Translation into English

We prompted GPT-4 [86] to translate text with: "Return the translation (and only the translation) of the following text from `[SRC_LANG]` into `[TGT_LANG]` exactly with all details: `[TEXT]`". We find that this prompt produces translations which especially preserve the conceptual details of the original text.

Although some language-specific meanings will inevitably be lost in any translation between languages, we ensure that our English translations are as faithful as possible to the concepts expressed in the original language by conducting a human evaluation. We recruit 2-3 speakers for each of the six non-English languages (French, German, Russian, Chinese, Japanese, Korean), fluent in both the original image and English. Each subject evaluates 30 pairs of original and translated text. Of these 30 pairs, 10 are Vertex captions on Crossmodal images, 10 are LLaVA captions on Crossmodal images, and 10 are Vertex captions on Visual Genome images. This composition ensures wide coverage across image domains and caption format. Each translation evaluation has two parts. Firstly, subjects annotate the overall translation quality on a 1 to 5 scale, in which 1 is "entirely inaccurate", 2 is "some of the information is preserved", 3 is "only the most important information is preserved", 4 is "most of the information is preserved (the translation is adequate but not perfect)", and 5 is "entirely accurate". Secondly, subjects examine 11 general categories of concepts in natural visual scenes, provided by TIFA [48]: objects, animals/humans, attributes, activities, spatial relations, counting, food, materials, shapes, locations and colors. Subjects mark each category either as "Good" (the concept was present in the original text and faithfully represented in the translation), "Missed" (the concept was present in the original text but absent or not faithfully represented in the translation), or "N/A" (the concept was not present in the original text). Table 6 demonstrates that the translations are nearly entirely accurate, especially for European languages, and preserve nearly all of the salient content categories for understanding visual scenes.

Annotators were allowed to provide free-text explanations for areas in which the translation was inadequate. We provide a random sampling of comments to provide a holistic idea of the translation weaknesses. Overall, the changes to the translations indicated in the comments do not change the content or expression of the text in a substantive way.

One possible confounder in results like §4 is that language-specific syntactic artifacts introduced during translation. For instance, text translated from German into English might have a unique syntactic structure which distinguishes it from text originally written in English. If this is the case, then it should be possible to identify translated text from one language versus another. To test this limitation, we embed all translated captions using a BERT-based model [100]. We fit a logistic regression model to predict a sample's original language from these features, and find near-random chance performance at 16.43% (random chance is $1/7 \approx 14.29\%$). This suggests that the translation artifact confounder does not explain the observed results.

8.2. Probing Multilingual Capabilities in LLaVA

Models like LLaVA which are trained/fine-tuned with English data but which include multilingual LLM components can retain some of these multilingual capabilities. In order to request LLaVA generate captions in a target language, we change the prompt at all levels to correspond to that language language, displayed in Table 8. This works successfully across each of the non-English languages considered in this work, except for Korean and Japanese, which exhibit significantly worse quality.

8.3. Image Captioning User Study

We recruited 10 English speakers from the US and 10 Japanese speakers from Japan. The instructions given to them are presented in Figures 3a and 3b. A sample of the produced captions is given in Table 9.

Because large-scale image-text datasets do not conduct much annotator information, it is difficult to make detailed and strong inferences about the psychological causes of the observed results, so more work is needed in this direction. However, as a start, we recruited 10 English speakers from the U.S. and 10 Japanese speakers from Japan to caption 30 Visual Genome images and repeated the semantic content evaluation for human-produced captions. We find, in the same pattern as before with model captions, that unioning English scene graphs with Japanese scene graphs expands the size by 8.4% objects, 7.7% relations, and 6.5% attributes over unioning English scene graphs with other English scene graphs. Moreover, a manual inspection of the captions suggests that the captions roughly echo the predictions from cross-cultural perceptual psychology – Japanese captions tend to mention background objects and information more than English ones (see Figures 1 and 9).

9. Supplementary Data and Figures

Our results across all evaluations are displayed in Table 10.

9.1. Semantics Evaluations

Figure 4 shows that despite an expected diminishing-returns trajectory, continuously unioning even a well-developed existing scene graph with a new language's scene graph expands it. This suggests that different languages continue to have new information to add to the existing scene graph of visual knowledge. Table 11 displays the sizes of intersections between monolingual scene graphs as measured by the number of objects and relations, using the formula $M(A) + M(B) - M(A \cup B) = M(A \cap B)$. It is an alternative way to understand the conceptual overlap of different languages. Table 12 shows that scene graphs constructed from captions *from the same model but different languages* are only slightly smaller than those constructed from captions *from the same language but different models*. Table 13 shows the intersection sizes between monolingual and multi-model scene graphs. Table 16 shows some samples in which multilingual scene graphs identify objects in the image which are not mentioned in the Visual Genome annotations. Figure 5 shows several examples of scene graphs generated in different languages for different samples.

9.2. Multilingual Embedding Space Coverage

Recall from §2.1 that many of the tools we use to measure semantics and expressions are not available in different languages (e.g., scene graph parsers, linguistic measures). However, in the case of embedding space coverage, we can use multilingual embeddings rather than monolingual (English) embeddings (with translation of all captions into English). We reproduce the expressive variation experiment described in §2.3 using multilingual embeddings *without translation*, and find that the same result holds as in the main paper using English embeddings with translation 5. This provides further empirical support that translation bias does not interfere with our results. However, note that multilingual embeddings have documented language biases [18, 87], which is why we prefer to use monolingual embeddings with translation for a fairer comparison.

	mono			multi		
	en	fr	avg	en,fr,de	en,ru,zh	avg
XM	.274	.279	.280	.327	.328	.340
LLaVA	.475	.507	.521	.704	.795	.753
Vertex	.340	.321	.321	.600	.647	.612

Table 5. Model representations experiment from the paper, repeated using multilingual Sentence-BERT without translation. 'avg' is the mean cosine distance across all monolingual and multilingual caption sets; the difference is significant ($p < 0.001$).

9.3. Model outputs evaluations

Tables 14 and 15 repeat the same fine-tuning experiment as outlined in §4, but training on LLaVA and XM captions instead of Vertex captions.

Table 6. Human evaluations for translation quality using GPT-4 on multilingual captions. TIFA categories represent the mean proportion of non-N/A responses which are marked “Good” (as opposed to “Missed”).

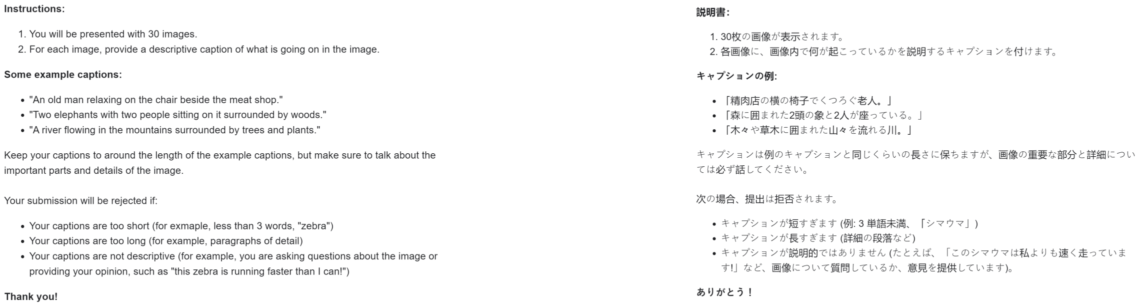
Metric		**de**	**fr**	**ru**	**zh**	**ja**	**ko**
Quality Ratings	Mean	4.95	4.76	4.82	4.63	4.48	4.48
	Median	5.00	5.00	5.00	5.00	5.00	5.00
	25th Percentile	5.00	5.00	5.00	4.00	4.00	4.00
TIFA Categories	Objects	1.00	0.99	1.0	0.97	0.98	0.90
	Animals/Humans	1.00	1.00	1.00	1.00	1.00	1.00
	Attributes	1.00	0.89	1.00	0.93	1.00	1.00
	Activities	1.00	1.00	1.00	0.98	0.91	0.96
	Spatial Relations	1.00	1.00	1.00	0.92	0.94	0.96
	Counting	1.00	1.00	1.00	0.99	1.00	0.90
	Food	1.00	1.00	1.00	1.00	1.00	1.00
	Material	1.00	1.00	1.00	1.00	1.00	1.00
	Shape	1.00	1.00	1.00	1.00	1.00	1.00
	Location	1.00	0.98	1.00	0.98	0.96	0.89
	Color	1.00	0.96	1.00	0.97	1.00	1.00

Table 7. Example annotator comments suggesting corrections to translations.

Comment
↪ Should use “above” instead of “on”
↪ More appropriate to use ”memories” instead of ”impressions”
↪ should be ‘small’ balls (remove ‘round’, add ‘small’)
↪ Particle suggests that numbers are written “using” sheet of paper, not “on” it.
↪ “On the side” is translated as “next to”.
↪ “toile d’ araignée” can be directly translated to “cobweb”

Table 8. Prompt information for probing multilingual behavior in LLaVA.

Prompt Type	**Language**	**Prompt**
Roles	English	(user, assistant)
	French	(utilisateur, assistant)
	German	(Benutzer, Assistent)
	...	...
System	English	A conversation between a user and an LLM-based AI assistant. The assistant gives helpful and honest answers.
	French	Une conversation entre un utilisateur et un assistant IA basé sur LLM. L’assistant donne des réponses utiles et honnêtes.
	German	Ein Gespräch zwischen einem Benutzer und einem auf LLM basierenden KI-Assistenten. Der Assistent gibt hilfreiche und ehrliche Antworten.
	...	...
User Prompt	English	What is in this image? Answer in English.
	French	Qu’est-ce qu’il y a dans cette image? Répondez en français.
	German	Was ist auf diesem Bild? Antwort auf Deutsch.
	...	...



Instructions:

1. You will be presented with 30 images.
2. For each image, provide a descriptive caption of what is going on in the image.

Some example captions:

- "An old man relaxing on the chair beside the meat shop."
- "Two elephants with two people sitting on it surrounded by woods."
- "A river flowing in the mountains surrounded by trees and plants."

Keep your captions to around the length of the example captions, but make sure to talk about the important parts and details of the image.

Your submission will be rejected if:

- Your captions are too short (for exmaple, less than 3 words, "zebra")
- Your captions are too long (for example, paragraphs of detail)
- Your captions are not descriptive (for example, you are asking questions about the image or providing your opinion, such as "this zebra is running faster than I can!")

Thank you!

(a) English instructions.

説明書:

1. 30枚の画像が表示されます。
2. 各画像に、画像内で何が起こっているかを説明するキャプションを付けます。

キャプションの例:

- 「精肉店の横の椅子でくつろぐ老人。」
- 「森に囲まれた2頭の象と2人が座っている。」
- 「木々や草木に囲まれた山々を流れる川。」

キャプションは例のキャプションと同じくらいの長さに保ちますが、画像の重要な部分と詳細については必ず話してください。

次の場合、提出は拒否されます。

- キャプションが短すぎます (例: 3 単語未満、「シマウマ」)
- キャプションが長すぎます (詳細の段落など)
- キャプションが説明的ではありません (たとえば、「このシマウマは私よりも速く走っています!」など、画像について質問しているか、意見を提供しています)。

ありがとう！

(b) Japanese instructions.

Figure 3. Instructions and examples presented to human evaluation participants for image captioning.

Table 9. A few examples of captions collected from the human study across English and Japanese speakers show differences in the observed content for each image. Japanese captions tend to include more context (e.g., background objects, added details). Samples are selected but representative of broader trends.

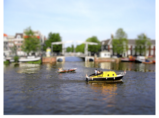
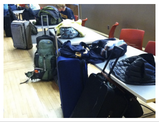
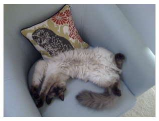
	English	I.	Two very small boats on a river
		II.	Toy boats in the water
		III.	A yellow boat and a red boat that appear to be models.
	Japanese	I.	Two boats on the water and a building in the back
		II.	close-up of a model of a boat and people on the waterfront
		III.	Two boats floating on the river and a model of the town in the distance
	English	I.	Luggage left unattended at a table.
		II.	Luggage lined up next to tables with jackets resting on the tables.
		III.	luggage sitting next to tables
	Japanese	I.	A man sitting in a lobby with lots of suitcases and bags
		II.	A man is sitting in a room, and there are several tables filled with luggage nearby.
		III.	Man waiting with a lot of luggage
	English	I.	Cat laying down in an arm chair.
		II.	A Siamese cat laying on its back on a couch next to a pillow.
		III.	A cat stretched out and upside down on a chair
	Japanese	I.	A cat stretches out on a blue chair and a pillow with an embroidered owl next to it.
		II.	A cat is relaxing next to a cushion with a picture of an owl on it.
		III.	Cat sitting on his back in an armchair with an owl-patterned cushion

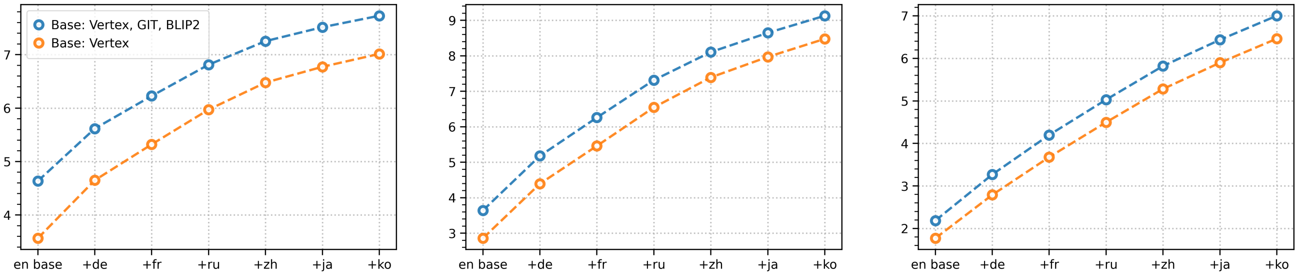
Table 10. Primary results across each of four evaluations. Shaded comparisons emphasize comparison between monolingual English caption sets and multilingual caption sets (including English and other languages), which have higher coverage/diversity across their respect evaluations.

Level	Evaluation	Data Source	Metric	Monolingual							Multilingual			
				en	**fr**	**de**	**ru**	**zh**	**ja**	**ko**	**en,fr,zh**	**fr,zh,ru**	**de,fr,ru**	**avg**[a]
Language	Logical meaning	Vertex	Objects	3.65	3.51	3.60	3.86	3.46	3.13	3.18	4.13	4.24	4.13	4.17
			Relations	2.96	2.83	2.89	3.20	2.68	2.37	2.47	3.45	3.48	3.38	3.40
			Attributes	1.67	1.67	1.79	1.86	1.66	1.59	1.62	2.29	2.40	2.33	2.33
		LLaVA	Objects	4.54	5.05	5.26	4.52	4.54	-	[b] -	6.14	6.15	6.25	5.93
			Relations	3.79	4.21	4.42	3.67	3.66	-	-	4.85	4.76	4.97	4.54
			Attributes	2.75	3.47	3.50	2.76	3.25	-	-	4.00	3.99	4.07	3.86
		XM	Objects	2.59	2.92	3.16	3.03	2.99	3.41	2.71	3.71	3.92	3.93	4.35
			Relations	1.54	1.76	1.94	1.88	1.71	1.99	1.59	2.41	2.57	2.57	2.94
			Attributes	1.27	1.66	1.97	1.74	2.01	2.47	1.46	2.36	2.59	2.55	2.97
	Expression	Vertex	Concreteness	1.64	1.64	1.67	1.66	1.51	1.50	1.56	1.75	1.74	1.73	1.81
			Analytic	0.85	0.43	0.62	1.08	2.3	2.6	2.17	2.02	2.05	1.0	2.3
			Clout	4.94	5.05	5.40	7.14	6.92	6.49	5.56	10.59	11.29	9.8	11.01
			Authentic	23.21	22.8	21.68	23.07	23.51	25.01	21.67	40.16	36.94	31.85	38.06
			Tone	1.85	1.84	2.03	2.63	2.05	1.96	2.05	3.98	4.10	4.19	3.92
		LLaVA	Concreteness	2.17	2.44	2.53	2.27	2.33	-	-	2.54	2.54	2.57	2.56
			Analytic	2.85	4.64	14.72	7.53	23.05	-	-	22.03	22.29	23.76	19.61
			Clout	14.19	19.96	15.81	23.15	21.3	-	-	23.06	26.78	26.84	28.72
			Authentic	35.97	38.01	34.33	37.20	44.63	-	-	54.94	54.64	54.82	53.15
			Tone	6.79	9.53	16.64	16.48	11.56	-	-	16.74	20.79	19.80	16.56
		XM	Concreteness	1.80	2.24	2.11	2.03	2.26	2.25	2.14	2.08	2.13	2.10	2.17
			Analytic	1.88	2.31	1.62	0.76	5.02	5.28	5.13	4.12	4.37	4.47	4.49
			Clout	9.66	14.77	15.00	14.14	13.32	13.10	12.24	18.8	20.51	19.54	19.39
			Authentic	33.21	40.09	42.22	32.35	34.02	33.83	35.28	53.12	54.41	53.35	53.21
			Tone	8.62	12.16	9.74	10.90	10.59	9.18	9.07	13.78	14.69	15.42	15.40
Model	Embeddings	Vertex	Coverage	0.19	0.18	0.19	0.22	0.20	0.19	0.37	0.37	0.33	0.38	0.49
		LLaVA		0.22	0.29	0.28	0.29	0.36	-	-	0.45	0.47	0.43	0.47
		XM		0.38	0.0.42	0.43	0.40	0.46	0.42	0.43	0.54	0.54	0.49	0.52
	Fine-tuning	Vertex	Best Model[c]	en	de	fr	ru	zh	ja	ko	-	-	-	multi
		LLaVA		en	de	fr	ru	zh	-	-	-	-	-	multi
		XM		en	de	fr	ru	zh	ja	ko	-	-	-	multi

[a] The 'avg' column is computed by randomly sampling across all languages.

[b] Japanese and Korean are excluded from LLaVA results due to extremely poor caption generation in the languages.

[c] Each cell value is the original language of the caption set used to train the model which performed best as evaluated by the caption set originally from the language specified in the column.



(a) # objects in scene graph

(b) # relations in scene graph

(c) # attributes in scene graph

Figure 4. Scene graphs of captions unioned cumulatively from different languages lead to more coverage in objects, relations, and attributes.

Table 11. Sizes of intersections between monolingual unioned scene graphs, listed in the form "number of objects / number of relations". Sizes listed along the diagonal arc of monolingual graphs and can be used for reference.

	en	de	fr	ru	zh	ja	ko
en	3.65 / 2.96	2.34 / 1.20	2.41 / 1.26	2.39 / 1.24	2.15 / 0.97	2.04 / 0.91	2.02 / 0.89
de		3.51 / 2.83	2.37 / 1.24	2.47 / 1.32	2.14 / 0.98	2.04 / 0.91	2.06 / 0.94
fr			3.60 / 2.89	2.44 / 1.27	2.16 / 0.97	2.07 / 0.91	2.09 / 0.95
ru				3.86 / 3.2	2.25 / 1.06	2.13 / 0.98	2.12 / 0.96
zh					3.46 / 2.68	2.08 / 0.95	2.04 / 0.92
ja						3.13 / 2.37	2.10 / 1.02
ko							3.18 / 2.47

Table 12. Scene graph metrics across Vertex and LLaVA captions in different languages show that multilingual scene graph unions are richer than monolingual ones. Increases are relative to the English average.

		en,fr,zh	**fr,de,ru**	**multi-model**
Vertex	**Objects**	4.31	4.25	4.63
	Relations	3.60	3.56	3.64
	Attributes	2.13	2.15	2.19
LLaVA	**Objects**	5.87	6.02	6.65
	Relations	4.84	4.97	5.42
	Attributes	4.10	4.07	2.88

Table 13. Intersection sizes between 3 unioned monolingual Vertex captions and an English multimodel baseline (a unioned BLIP2 ∪ GIT scene graph, held constant across all languages) are both relatively **small** and **smaller for Asian than European languages**. All relationships between European languages and Asian languages are statistically significant with Bonferroni correction. The 'mm' column includes the size of the unioned GIT and BLIP model scene graph for reference.

		Language							
		en	**de**	**fr**	**ru**	**zh**	**ja**	**ko**	**mm**
Metric	**Objects**	1.96	1.92	1.93	1.97	1.85	1.73	1.76	3.59
	Relations	0.79	0.76	0.74	0.78	0.70	0.62	0.64	2.51
	Attributes	0.44	0.36	0.37	0.42	0.37	0.37	0.33	1.45

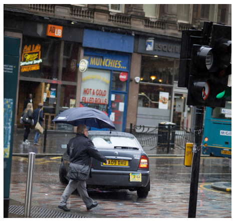
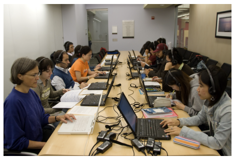
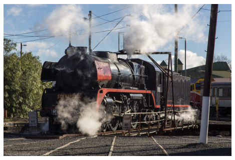
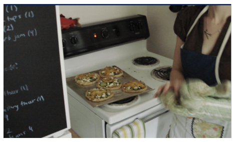
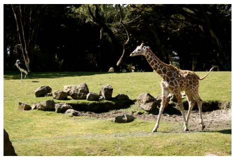
Table 14. Evaluations for models fine-tuned on LLaVA captions. Generally speaking, a model fine-tuning on a particular language performs best on that language.

		Evaluated on					
		en	**de**	**fr**	**ru**	**zh**	**multi**
Fine-tuned on	**en**	0.271	0.225	0.229	0.219	0.218	0.230
	de	0.213	0.245	0.219	0.217	0.215	0.219
	fr	0.248	0.240	0.259	0.234	0.236	0.246
	ru	0.226	0.234	0.228	0.254	0.231	0.239
	zh	0.199	0.202	0.199	0.207	0.247	0.216
	multi	0.239	0.233	0.234	0.233	0.235	0.244

Table 15. Evaluations for models fine-tuned on XM captions. Generally speaking, a model fine-tuning on a particular language performs best on that language.

		Evaluated on							
		en	**de**	**fr**	**ru**	**zh**	**ja**	**ko**	**multi**
Fine-tuned on	**en**	0.254	0.124	0.1421	0.120	0.114	0.129	0.130	0.148
	de	0.158	0.153	0.152	0.143	0.124	0.140	0.146	0.149
	fr	0.182	0.142	0.181	0.143	0.130	0.146	0.150	0.154
	ru	0.172	0.136	0.152	0.159	0.125	0.137	0.142	0.148
	zh	0.144	0.116	0.129	0.120	0.124	0.130	0.142	0.130
	ja	0.144	0.128	0.137	0.125	0.124	0.154	0.144	0.135
	ko	0.151	0.116	0.131	0.116	0.115	0.134	0.159	0.134
	multi	0.179	0.140	0.153	0.145	0.131	0.149	0.151	0.151

Table 16. Examples in which multilingual distributions identify visual features which are not documented in the Visual Genome dataset. Rightmost column indicates objects mentioned in multilingual scene graphs but which are not covered in the Visual Genome object list, shown in the left column.

Image	**VG Objects**	**Scene Graph Objects**
	woman, sign, man, bag, license plate, car, person, leg, satchel	umbrella, sandwich restaurant, street, rain
	woman, key, notes, page, keyboard, pencil case, laptop, student	table
	leaves, sign, sky, cloud, trees, roof, train, steam cloud, ground, lamp, green leaves, cables, pole, tracks, locomotive, train car, tree, steeples, gravel, steam, bush, door, wheel	number, logo, inscription
	tray, writing, cloth, stove door, light, oven back, bird necklace, mitt, shirt, apron, stove, burner, strings, aprontop, towel, board, pizza, shortsleeveshirt, menu, woman, necklace, pizzas, pan, oven, sheet	chalkboard
	giraffe tail, spot, rock, giraffe, rocks, grass	bird (left of image)

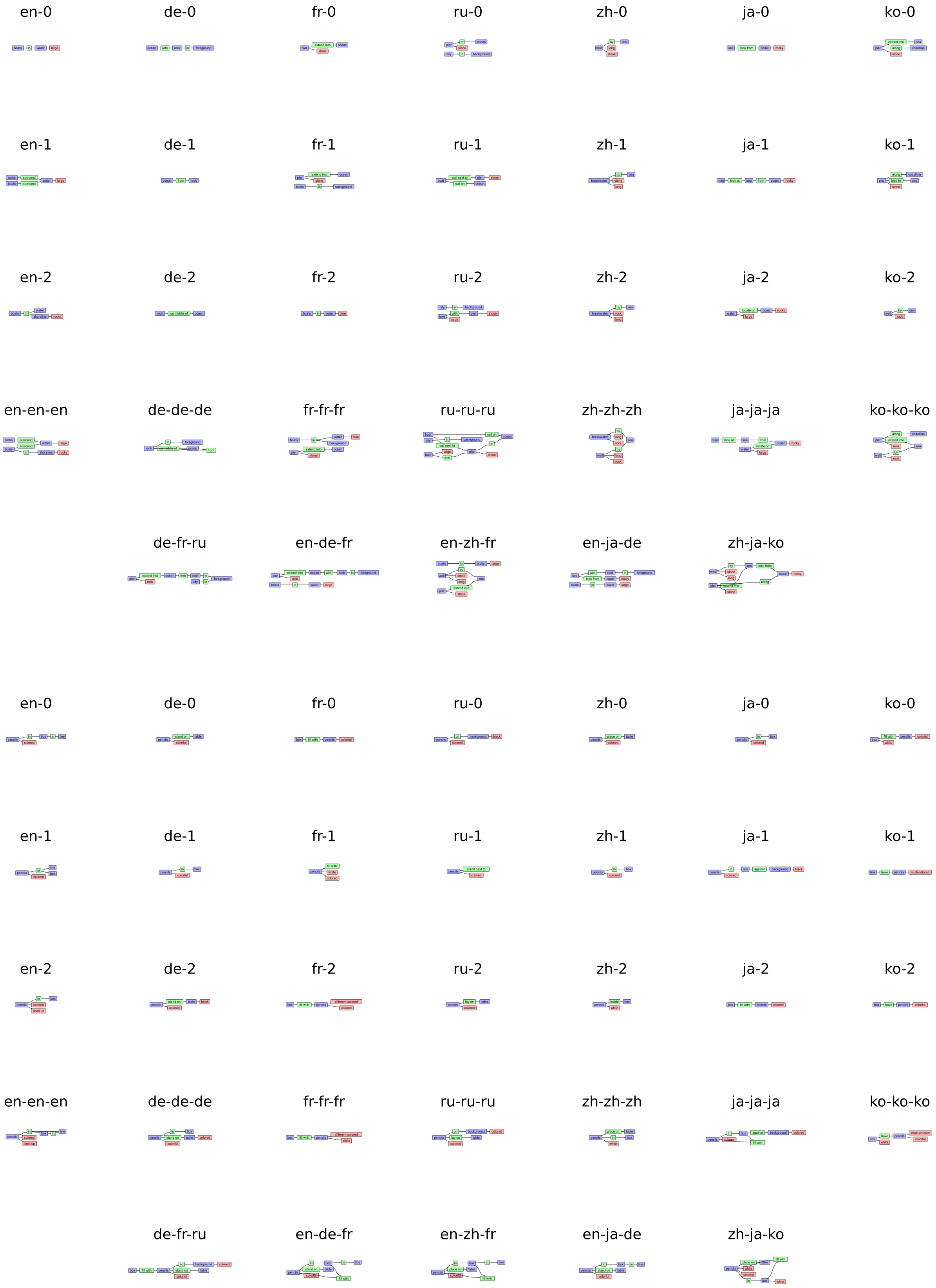
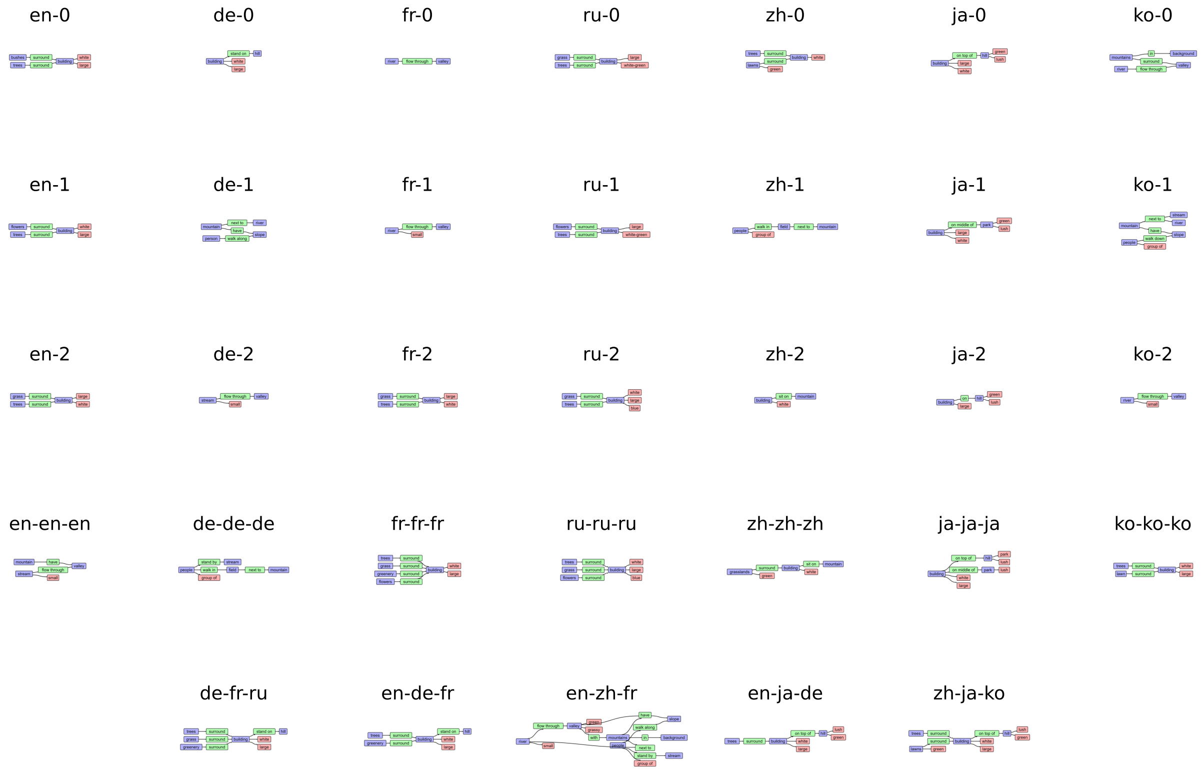


Figure 5. Sample scene graphs across six images. "lang-n" indicates the scene graph generated for the nth caption in lang. "lang1-lang2-lang3" indicates the scene graph unioned from three scene graphs originally from each of the three languages.



en-0

de-0

fr-0

ru-0

zh-0

ja-0

ko-0

en-1

de-1

fr-1

ru-1

zh-1

ja-1

ko-1

en-2

de-2

fr-2

ru-2

zh-2

ja-2

ko-2

en-en-en

de-de-de

fr-fr-fr

ru-ru-ru

zh-zh-zh

ja-ja-ja

ko-ko-ko

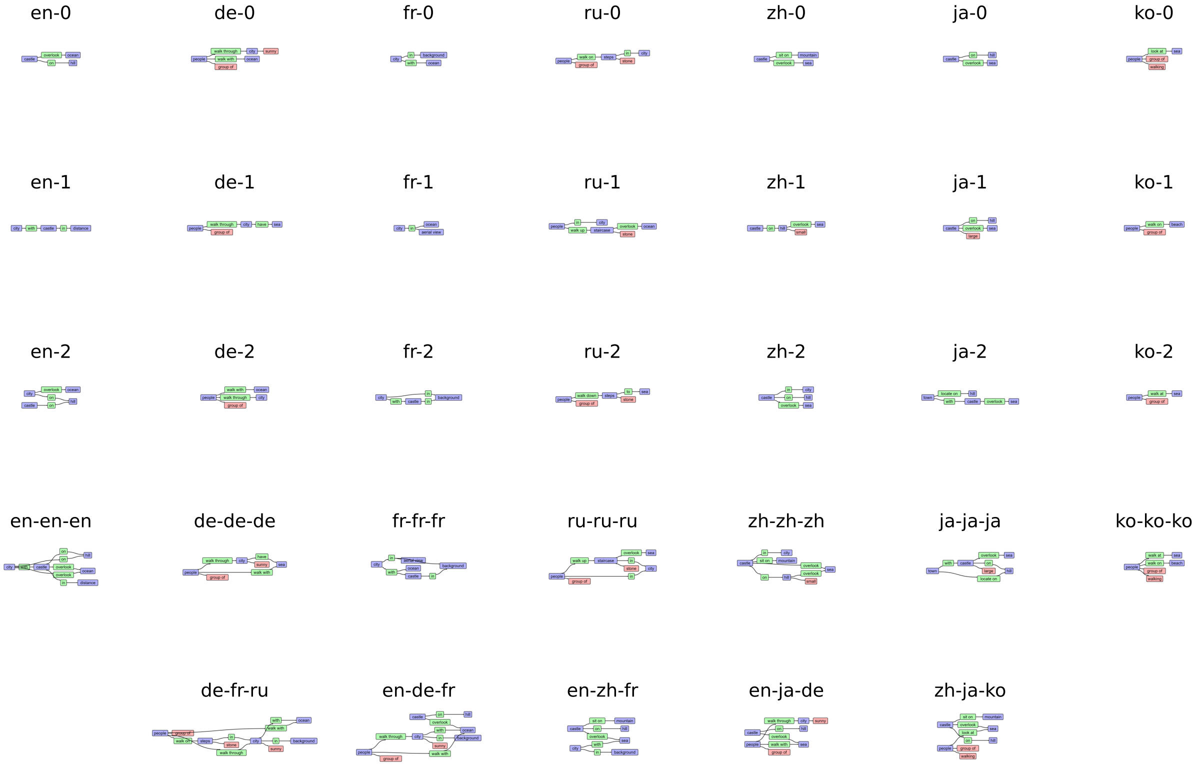
de-fr-ru

en-de-fr

en-zh-fr

en-ja-de

zh-ja-ko



en-0

de-0

fr-0

ru-0

zh-0

ja-0

ko-0

en-1

de-1

fr-1

ru-1

zh-1

ja-1

ko-1

en-2

de-2

fr-2

ru-2

zh-2

ja-2

ko-2

en-en-en

de-de-de

fr-fr-fr

ru-ru-ru

zh-zh-zh

ja-ja-ja

ko-ko-ko

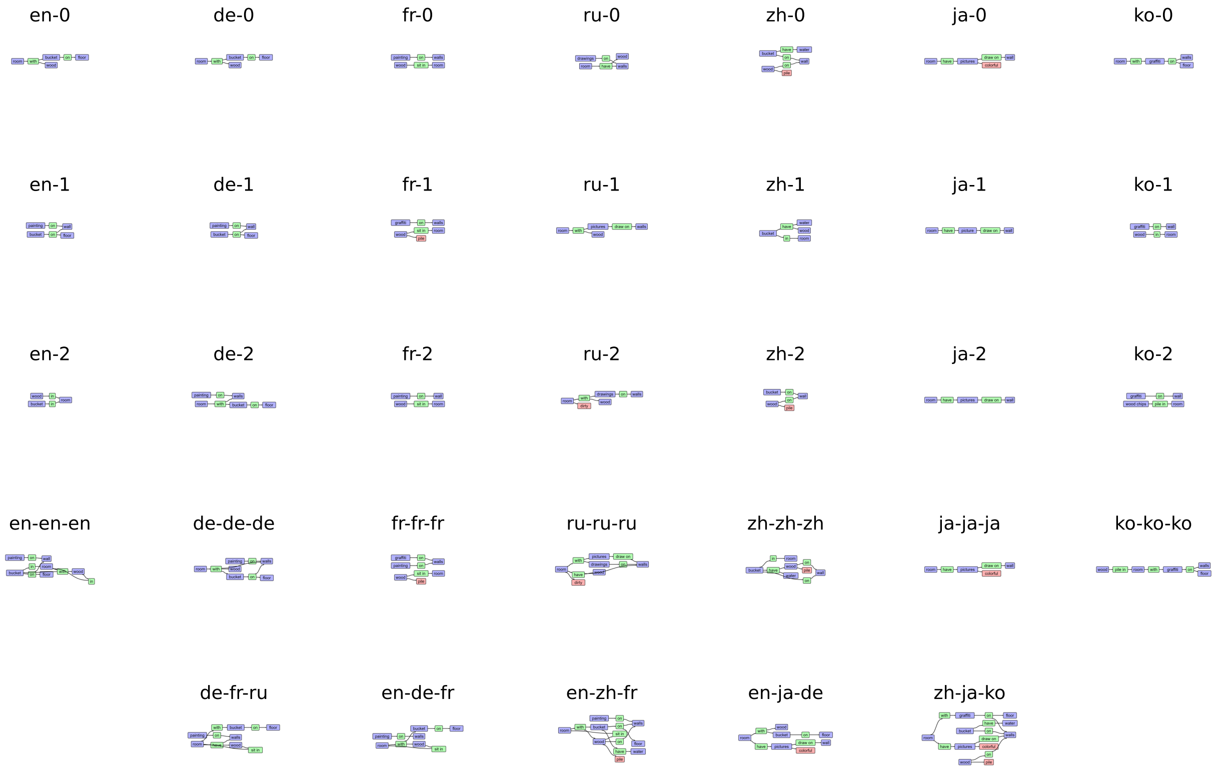
de-fr-ru

en-de-fr

en-zh-fr

en-ja-de

zh-ja-ko



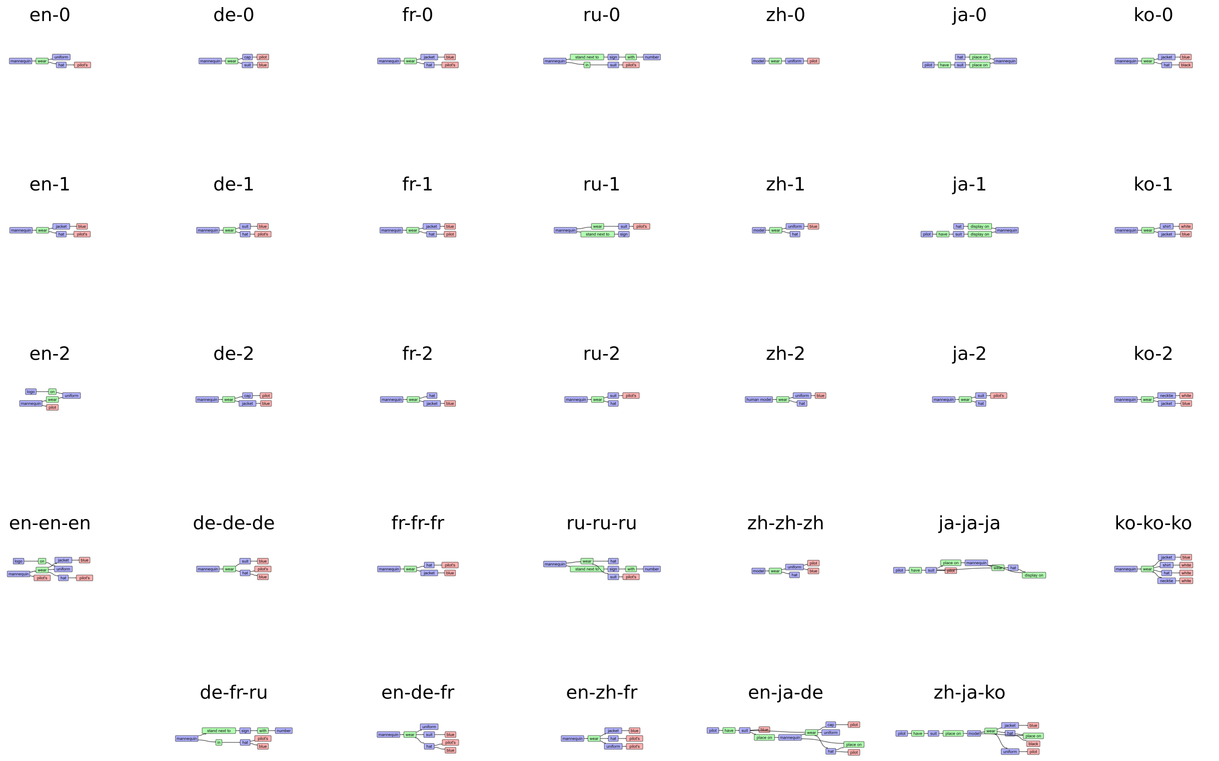
en-0 de-0 fr-0 ru-0 zh-0 ja-0 ko-0

en-1 de-1 fr-1 ru-1 zh-1 ja-1 ko-1

en-2 de-2 fr-2 ru-2 zh-2 ja-2 ko-2

en-en-en de-de-de fr-fr-fr ru-ru-ru zh-zh-zh ja-ja-ja ko-ko-ko

de-fr-ru en-de-fr en-zh-fr en-ja-de zh-ja-ko



en-0 de-0 fr-0 ru-0 zh-0 ja-0 ko-0

en-1 de-1 fr-1 ru-1 zh-1 ja-1 ko-1

en-2 de-2 fr-2 ru-2 zh-2 ja-2 ko-2

en-en-en de-de-de fr-fr-fr ru-ru-ru zh-zh-zh ja-ja-ja ko-ko-ko

de-fr-ru en-de-fr en-zh-fr en-ja-de zh-ja-ko