Don't Rule Out Monolingual Speakers:
A Method For Crowdsourcing Machine Translation Data

Rajat Bhatnagar, Ananya Ganesh and **Katharina Kann**
University of Colorado Boulder
{rajat.bhatnagar, ananya.ganesh, katharina.kann}@colorado.edu

Abstract

High-performing machine translation (MT) systems can help overcome language barriers while making it possible for everyone to communicate and use language technologies in the language of their choice. However, such systems require large amounts of parallel sentences for training, and translators can be difficult to find and expensive. Here, we present a data collection strategy for MT which, in contrast, is cheap and simple, as it does not require bilingual speakers. Based on the insight that humans pay specific attention to movements, we use graphics interchange formats (GIFs) as a pivot to collect parallel sentences from monolingual annotators. We use our strategy to collect data in Hindi, Tamil and English. As a baseline, we also collect data using images as a pivot. We perform an intrinsic evaluation by manually evaluating a subset of the sentence pairs and an extrinsic evaluation by finetuning mBART (Liu et al., 2020) on the collected data. We find that sentences collected via GIFs are indeed of higher quality.

1 Introduction

Machine translation (MT) – automatic translation of text from one natural language into another – provides access to information written in foreign languages and enables communication between speakers of different languages. However, developing high performing MT systems requires large amounts of training data in the form of parallel sentences – a resource which is often difficult and expensive to obtain, especially for languages less frequently studied in natural language processing (NLP), endangered languages, or dialects.

For some languages, it is possible to scrape data from the web (Resnik and Smith, 2003), or to leverage existing translations, e.g., of movie subtitles (Zhang et al., 2014) or religious texts (Resnik et al., 1999). However, such sources of data are only available for a limited number of languages,
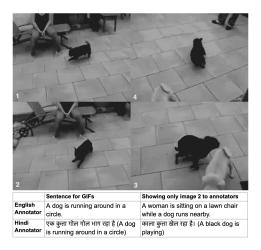

Figure 1: Sentences written by English and Hindi annotators using GIFs or images as a pivot.

and it is impossible to collect large MT corpora for a diverse set of languages using these methods. Professional translators, which are a straightforward alternative, are often rare or expensive.

In this paper, we propose a new data collection strategy which is cheap, simple, effective and, importantly, does not require professional translators or even bilingual speakers. It is based on two assumptions: (1) **non-textual modalities can serve as a pivot for the annotation process** (Madaan et al., 2020); and (2) **annotators subconsciously pay increased attention to moving objects**, since humans are extremely good at detecting motion, a crucial skill for survival (Albright and Stoner, 1995). Thus, we propose to leverage graphics interchange formats (GIFs) as a pivot to collect parallel data in two or more languages.

We prefer GIFs over videos as they are short in duration, do not require audio for understanding and describe a comprehensive story visually. Furthermore, we hypothesize that GIFs are better pivots than images – which are suggested by Madaan et al. (2020) for MT data collection – based on our second assumption. We expect that people who

are looking at the same GIF tend to focus on the main action and characters within the GIF and, thus, tend to write more similar sentences. This is in contrast to using images as a pivot, where people are more likely to focus on different parts of the image and, hence, to write different sentences, cf. Figure 1.

We experiment with collecting Hindi, Tamil and English sentences via Amazon Mechanical Turk (MTurk), using both GIFs and images as pivots. As an additional baseline, we compare to data collected in previous work (Madaan et al., 2020). We perform both intrinsic and extrinsic evaluations – by manually evaluating the collected sentences and by training MT systems on the collected data, respectively – and find that leveraging GIFs indeed results in parallel sentences of higher quality as compared to our baselines.[1]

2 Related Work

In recent years, especially with the success of transfer learning (Wang et al., 2018) and pretraining in NLP (Devlin et al., 2019), several techniques for improving neural MT for low-resource languages have been proposed (Sennrich et al., 2016; Fadaee et al., 2017; Xia et al., 2019; Lample et al., 2017; Lewis et al., 2019; Liu et al., 2020).

However, supervised methods still outperform their unsupervised and semi-supervised counterparts, which makes collecting training data for MT important. Prior work scrapes data from the web (Lai et al., 2020; Resnik and Smith, 2003), or uses movie subtitles (Zhang et al., 2014), religious texts (Resnik et al., 1999), or multilingual parliament proceedings (Koehn, 2005). However, those and similar resources are only available for a limited set of languages. A large amount of data for a diverse set of low-resource languages cannot be collected using these methods.

For low-resource languages, Hasan et al. (2020) propose a method to convert noisy parallel documents into parallel sentences. Zhang et al. (2020) filter noisy sentence pairs from MT training data.

The closest work to ours is Madaan et al. (2020). The authors collect (pseudo-)parallel sentences with images from the Flickr8k dataset (Hodosh et al., 2013) as a pivot, filtering to obtain images which are simplistic and do not contain culture-specific references. Since Flickr8k already contains 5 English captions per image, they select images whose captions are short and of high similarity to each other. Culture-specific images are manually discarded. We compare to the data from Madaan et al. (2020) in Section 4, denoting it as M20.

3 Experiments

3.1 Pivot Selection

We propose to use GIFs as a pivot to collect parallel sentences in two or more languages. As a baseline, we further collect parallel data via images as similar to our GIFs as possible. In this subsection, we describe our selection of both mediums.

GIFs We take our GIFs from a dataset presented in Li et al. (2016), which consists of 100k GIFs with descriptions. Out of these, 10k GIFs have three English one-sentence descriptions each, which makes them a suitable starting point for our experiments. We compute the word overlap in F1 between each possible combination of the three sentences, take the average per GIF, and choose the highest scoring 2.5k GIFs for our experiments. This criterion filters for GIFs for which all annotators focus on the same main characters and story, and it eliminates GIFs which are overly complex. We thus expect speakers of non-English languages to focus on similar content.

Images Finding images which are comparable to our GIFs is non-trivial. While we could compare our GIFs' descriptions to image captions, we hypothesize that the similarity between the images obtained thereby and the GIFs would be too low for a clean comparison. Thus, we consider two alternatives: (1) using the *first* frame of all GIFs, and (2) using the *middle* frame of all GIFs.

In a preliminary study, we obtain two Hindi one-sentence descriptions from two different annotators for both the first and the middle frame for a subset of 100 GIFs. We then compare the BLEU (Papineni et al., 2002) scores of all sentence pairs. We find that, on average, sentences for the middle frame have a BLEU score of 7.66 as compared to 4.58 for the first frame. Since a higher BLEU score indicates higher similarity and, thus, higher potential suitability as MT training data, we use the middle frames for the image-as-pivot condition in our final experiments.

---

[1]All data collected for our experiments is available at https://nala-cub.github.io/resources.

Rating	Sentences from the GIF-as-Pivot Setting
1	A child flips on a trampoline. A girl enjoyed while playing.
3	A man in a hat is walking up the stairs holding a bottle of water. A man is walking with a plastic bottle.
5	A man is laughing while holding a gun. A man is laughing while holding a gun.
	Sentences from the Image-as-Pivot Setting
1	A woman makes a gesture in front of a group of other women. This woman is laughing.
3	An older woman with bright lip stick lights a cigarette in her mouth. This woman is lighting a cigarette.
5	A woman wearing leopard print dress and a white jacket is walking forward. A woman is walking with a leopard print dress and white coat.

Table 1: Sentences obtained in English and Hindi for each setting where both annotators agree on the rating. The first sentence is the sentence written in English and the second sentence is the corresponding English translation of the Hindi sentence, translated by the authors.

3.2 Data Collection

We use MTurk for all of our data collection. We collect the following datasets: (1) one single-sentence description in Hindi for each of our 2,500 GIFs; (2) one single-sentence description in Hindi for each of our 2,500 images, i.e., the GIFs' middle frames; (3) one single-sentence description in Tamil for each of the 2,500 GIFs; (4) one single-sentence description in Tamil for each of the 2,500 images; and (5) one single-sentence description in English for each of our 2,500 images. To build parallel data for the GIF-as-pivot condition, we randomly choose one of the available 3 English descriptions for each GIF.

For the collection of Hindi and Tamil sentences, we restrict the workers to be located in India and, for the English sentences, we restrict the workers to be located in the US. We use the instructions from Li et al. (2016) with minor changes for all settings, translating them for Indian workers.[2]

Each MTurk human intelligence task (HIT) consists of annotating five GIFs or images, and we expect each task to take a maximum of 6 minutes. We pay annotators in India $0.12 per HIT (or $1.2 per hour), which is above the minimum wage of $1 per hour in the capital Delhi.[3] Annotators in the US are paid $1.2 per HIT (or $12 per hour). We have obtained IRB approval for the experiments reported in this paper (protocol #: 20-0499).

	GIF-as-Pivot	Image-as-Pivot	M20
Hindi–English	**2.92**	2.20	2.63
Tamil–English	**3.03**	2.33	-

Table 2: Manual evaluation of a subset of our collected sentences; scores from 1 to 5; higher is better.

3.3 Test Set Collection

For the extrinsic evaluation of our data collection strategy we train and test an MT system. For this, we additionally collect in-domain development and test examples for both the GIF-as-pivot and the image-as-pivot setting.

Specifically, we first collect 250 English sentences for 250 images which are the middle frames of previously unused GIFs. We then combine them with the English descriptions of 250 additional unused GIFs from Li et al. (2016). For the resulting set of 500 sentences, we ask Indian MTurk workers to provide a translation into Hindi and Tamil. We manually verify the quality of a randomly chosen subset of these sentences. Workers are paid $1.2 per hour for this task. We use 100 sentence pairs from each setting as our development set and the remaining 300 for testing.

4 Evaluation

4.1 Intrinsic Evaluation

In order to compare the quality of the parallel sentences obtained under different experimental conditions, we first perform a manual evaluation of a subset of the collected data. For each lan-

---

[2]Our instructions can be found in the appendix.
[3]https://paycheck.in/salary/minimumwages/16749-delhi

	Rating	GIF-as-pivot	Image-as-pivot	M20
Hi-En	5	**13.08**	2.5	10.0
Ta-En		**6.0**	3.5	-
Hi-En	>= 4	**35.77**	15.5	26.43
Ta-En		**37.0**	14.0	-
Hi-En	>= 3	**61.15**	39.0	51.43
Ta-En		**67.5**	42.5	-
Hi-En	>= 2	**82.69**	63.0	75.0
Ta-En		**92.5**	72.5	-
Hi-En	>= 1	**100.0**	**100.0**	**100.0**
Ta-En		**100.0**	**100.0**	-

Table 3: Cumulative percentages with respect to each setting; GIF-as-pivot shows the best results;

guage pair, we select the same random 100 sentence pairs from the GIF-as-pivot and image-as-pivot settings. We further choose 100 random sentence pairs from M20. We randomly shuffle all sentence pairs and ask MTurk workers to evaluate the translation quality. Each sentence pair is evaluated independently by two workers, i.e., we collect two ratings for each pair. Sentence pairs are rated on a scale from 1 to 5, with 1 being the worst and 5 being the best possible score.[4]

Each evaluation HIT consists of 11 sentence pairs. For quality control purposes, each HIT contains one manually selected example with a perfect (for Hindi–English) or almost perfect (for Tamil–English) translation. Annotators who do not give a rating of 5 (for Hindi–English) or a rating of at least 4 (for Tamil–English) do not pass this check. Their tasks are rejected and republished.

Results The average ratings given by the annotators are shown in Table 2. Sentence pairs collected via GIF-as-pivot obtain an average rating of 2.92 and 3.03 for Hindi–English and Tamil–English, respectively. Sentences from the image-as-pivot setting only obtain an average rating of 2.20 and 2.33 for Hindi–English and, respectively, Tamil–English. The rating obtained for M20 (Hindi only) is 2.63. As we can see, for both language pairs the GIF-as-pivot setting is rated consistently higher than the other two settings, thus showing the effectiveness of our data collection strategy. This is in line with our hypothesis that the movement displayed in GIFs is able to guide the sentence writer's attention.

We now explicitly investigate how many of the translations obtained via different strategies are

---

[4]The definitions of each score as given to the annotators can be found in the appendix.

Test Set	Training Set	500	1000	1500	1900	2500
		Direction: Hindi to English				
GIF	GIF	**6.41**	**13.06**	**14.39**	**14.81**	**16.09**
GIF	Image	5.71	8.17	9.5	9.7	10.49
GIF	M20	3.19	6.84	7.99	6.9	N/A
Image	GIF	2.93	8.18	9.11	8.84	9.24
Image	Image	**8.46**	**10.05**	**11.15**	**11.25**	**12.14**
Image	M20	1.27	5.79	6.76	6.68	N/A
M20	GIF	1.66	5.21	5.75	6.78	**6.69**
M20	Image	1.63	4.53	4.98	5.09	5.63
M20	M20	**5.08**	**6.96**	**7.23**	**8.23**	N/A
All	GIF	3.47	**8.46**	**9.35**	**9.81**	**10.28**
All	Image	**4.9**	7.28	8.19	8.32	9.04
All	M20	3.37	6.57	7.32	7.37	N/A
		Direction: English to Hindi				
GIF	GIF	0.63	1.68	2.01	1.72	**3.07**
GIF	Image	**0.81**	**2.18**	1.43	2.29	1.86
GIF	M20	0.42	2.09	**2.99**	**3.06**	N/A
Image	GIF	0.11	1.19	1.03	0.97	**1.42**
Image	Image	0.15	1.19	1.04	1.09	1.29
Image	M20	**0.22**	**1.23**	**1.95**	**1.68**	N/A
M20	GIF	1.15	2.75	4.25	4.52	4.88
M20	Image	1.32	3.09	4.41	4.1	**5.16**
M20	M20	**5.12**	**12.27**	**12.65**	**13.31**	N/A
All	GIF	0.68	1.96	2.61	2.62	**3.3**
All	Image	0.82	2.25	2.51	2.65	3.01
All	M20	**2.24**	**5.9**	**6.54**	**6.75**	N/A

Table 4: BLEU for different training and test sets; *All* denotes a weighted average over all test sets; all models are obtained by finetuning mBART; best scores for each training set size and test set in bold.

acceptable or good translations; this corresponds to a score of 3 or higher. Table 3 shows that 61.15% of the examples are rated 3 or above in the GIF-as-pivot setting for Hindi as compared to 39.0% and 51.43% for the image-as-pivot setting and M20, respectively. For Tamil, 67.5% of the sentences collected via GIFs are at least acceptable translations. The same is true for only 42.5% of the sentences obtained via images.

We show example sentence pairs with their ratings from the GIF-as-pivot and image-as-pivot settings for Hindi–English in Table 1.

4.2 Extrinsic Evaluation

We further extrinsically evaluate our data by training an MT model on it. Since, for reasons of practicality, we collect only 2,500 examples, we leverage a pretrained model instead of training from scratch. Specifically, we finetune an mBART model (Liu et al., 2020) on increasing amounts of data from all setting in both directions. mBART is

Test Set	Training Set	500	1000	1500	2000	2500
		Direction: Tamil to English				
GIF	GIF	**2.63**	**4.46**	**8.26**	**9.27**	**4.99**
GIF	Image	2.33	3.34	3.00	4.77	3.83
Image	GIF	0.95	2.42	3.15	3.67	2.74
Image	Image	**6.65**	**5.62**	**6.02**	**7.75**	**7.22**
All	GIF	1.79	3.44	**5.71**	**6.47**	3.87
All	Image	**4.49**	**4.48**	4.51	6.26	**5.53**
		Direction: English to Tamil				
GIF	GIF	0	**0.54**	**1.00**	**0.83**	**0.84**
GIF	Image	**0.5**	0.18	0.96	0.43	0.48
Image	GIF	0	0.31	0.36	**0.62**	**0.7**
Image	Image	**0.41**	**0.35**	**0.51**	0.36	0.29
All	GIF	0	**0.43**	0.68	**0.73**	**0.77**
All	Image	**0.46**	0.27	**0.74**	0.4	0.39

Table 5: BLEU for different training and test sets; *All* denotes a weighted average over all test sets; all models are obtained by finetuning mBART; best scores for each training set size and test set in bold.

a transformer-based sequence-to-sequence model which is pretrained on 25 monolingual raw text corpora. We finetune it with a learning rate of 3e-5 and a dropout of 0.3 for up to 100 epochs with a patience of 15.

Results The BLEU scores for all settings are shown in Tables 4 and 5 for Hindi–English and Tamil–English, respectively. We observe that increasing the dataset size mostly increases the performance for all data collection settings, which indicates that the obtained data is useful for training. Further, we observe that each model performs best on its own in-domain test set.

Looking at Hindi-to-English translation, we see that, on average, models trained on sentences collected via GIFs outperform sentences from images or M20 for all training set sizes, except for the 500-examples setting, where image-as-pivot is best. However, results are mixed for Tamil-to-English translation.

Considering English-to-Hindi translation, models trained on M20 data outperform models trained on sentences collected via GIFs or our images in nearly all settings. However, since the BLEU scores are low, we manually inspect the obtained outputs. We find that the translations into Hindi are poor and differences in BLEU scores are often due to shared individual words, even though the overall meaning of the translation is incorrect. Similarly, for English-to-Tamil translation,

all BLEU scores are below or equal to 1. We thus conclude that 2,500 examples are not enough to train an MT system for these directions, and, while we report all results here for completeness, we believe that the intrinsic evaluation paints a more complete picture.[5] We leave a scaling of our extrinsic evaluation to future work.

5 Conclusion

In this work, we made two assumptions: (1) that a non-textual modality can serve as a pivot for MT data collection, and (2) that humans tend to focus on moving objects. Based on this, we proposed to collect parallel sentences for MT using GIFs as pivots, eliminating the need for bilingual speakers and reducing annotation costs. We collected parallel sentences in English, Hindi and Tamil using our approach and conducted intrinsic and extrinsic evaluations of the obtained data, comparing our strategy to two baseline approaches which used images as pivots. According to the intrinsic evaluation, our approach resulted in parallel sentences of higher quality than either baseline.

Acknowledgments

References

Thomas D Albright and Gene R Stoner. 1995. Visual motion perception. *Proceedings of the National Academy of Sciences*, 92(7):2433–2440.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Marzieh Fadaee, Arianna Bisazza, and Christof Monz. 2017. Data augmentation for low-resource neural

---

[5]We also manually inspect the translations into English: in contrast to the Hindi translations, most sentences at least partially convey the same meaning as the reference.

machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 567–573, Vancouver, Canada. Association for Computational Linguistics.

Tahmid Hasan, Abhik Bhattacharjee, Kazi Samin, Masum Hasan, Madhusudan Basak, M. Sohel Rahman, and Rifat Shahriyar. 2020. Not low-resource anymore: Aligner ensembling, batch filtering, and new datasets for Bengali-English machine translation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2612–2623, Online. Association for Computational Linguistics.

Micah Hodosh, Peter Young, and Julia Hockenmaier. 2013. Framing image description as a ranking task: Data, models and evaluation metrics. *Journal of Artificial Intelligence Research*, 47:853–899.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT summit*, volume 5, pages 79–86. Citeseer.

Guokun Lai, Zihang Dai, and Yiming Yang. 2020. Unsupervised parallel corpus mining on web data. *arXiv preprint arXiv:2009.08595*.

Guillaume Lample, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2017. Unsupervised machine translation using monolingual corpora only. *arXiv preprint arXiv:1711.00043*.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

Yuncheng Li, Yale Song, Liangliang Cao, Joel Tetreault, Larry Goldberg, Alejandro Jaimes, and Jiebo Luo. 2016. TGIF: A new dataset and benchmark on animated GIF description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4641–4650.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.

Aman Madaan, Shruti Rijhwani, Antonios Anastasopoulos, Yiming Yang, and Graham Neubig. 2020. Practical comparable data collection for low-resource languages via images. *arXiv preprint arXiv:2004.11954*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia,

Pennsylvania, USA. Association for Computational Linguistics.

P. Resnik, M. Olsen, and Mona T. Diab. 1999. The bible as a parallel corpus: Annotating the 'book of 2000 tongues'. *Computers and the Humanities*, 33:129–153.

Philip Resnik and Noah A. Smith. 2003. The web as a parallel corpus. *Computational Linguistics*, 29(3):349–380.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96, Berlin, Germany. Association for Computational Linguistics.

Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. 2018. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 353–355, Brussels, Belgium. Association for Computational Linguistics.

Mengzhou Xia, Xiang Kong, Antonios Anastasopoulos, and Graham Neubig. 2019. Generalized data augmentation for low-resource translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5786–5796, Florence, Italy. Association for Computational Linguistics.

Boliang Zhang, Ajay Nagesh, and Kevin Knight. 2020. Parallel corpus filtering via pre-trained language models. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8545–8554, Online. Association for Computational Linguistics.

Shikun Zhang, Wang Ling, and Chris Dyer. 2014. Dual subtitles as parallel corpora. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1869–1874, Reykjavik, Iceland. European Language Resources Association (ELRA).

A Sentence Rating Instructions

Score	Title	Description
1	Not a translation	There is no relation whatsoever between the source and the target sentence
2	Bad	Some word overlap, but the meaning isn't the same
3	Acceptable	The translation conveys the meaning to some degree but is a bad translation
4	Good	The translation is missing a few words but conveys most of the meaning adequately
5	Perfect	The translation is perfect or close to perfect

Table 6: Description of the ratings for the manual evaluation of translations.

B MTurk Instructions

Instructions for English image task

Below you will see five images. Your task is to describe each image in one English sentence. You should focus solely on the visual content presented in the image. Each sentence should be grammatically correct. It should describe the main characters and their actions, but NOT your opinions, guesses or interpretations.

- DOs
  - Please use only English words. No digits allowed (spell them out, e.g., three).
  - Sentences should neither be too short nor too long. Try to be concise.
  - Each sentence must contain a verb.
  - If possible, include adjectives that describe colors, size, emotions, or quantity.
  - Please pay attention to grammar and spelling.
  - Each sentence must express a complete idea, and make sense by itself.
  - The sentence should describe the main characters, actions, setting, and relationship between the objects.
- DONTs
  - The sentence should NOT contain any digits.
  - The sentence should NOT mention the name of a movie, film, and character.
  - The sentence should NOT mention invisible objects and actions.
  - The sentence should NOT make subjective judgments about the image.

Remember, please describe only the visual content presented in the images. Focus on the main characters and their actions.

निर्देश (Instructions for GIF Task in Hindi)

नीचे आपको पांच गिफ (GIF) दिखाई देंगे. आपको हर गिफ को एक वाक्य में हिंदी में समझाना है. आपको सिर्फ गिफ में जो हो रहा है उसपर ध्यान देना है. आपके वाक्य की व्याकरण सही होनी चाहिए. आपको मुख्य पात्रों और उनके कार्यों का वर्णन करना है और आपको अपनी राय नहीं देनी है.

- क्या करें -
  - कृपया केवल हिंदी शब्दों और हिंदी लिपि (देवनागरी) का उपयोग करें। किसी भी अंक को पूरी तरह लिखे (उदहारण - तीन लिखे नाकि ३)।
  - वाक्य न तो बहुत छोटे होने चाहिए और न ही बहुत लंबे। संक्षिप्त होने का प्रयास करें। वाक्य कम से कम चार शब्दों का होना चाहिए|
  - प्रत्येक वाक्य में एक क्रिया होनी चाहिए।
  - यदि संभव हो तो विशेषणों का इस्तेमाल करें जो की रंगो, आकार व भावनाओं को अच्छे से समझा सके।
  - कृपया व्याकरण और स्पेलिंग पर ध्यान दें।
  - प्रत्येक वाक्य को एक पूर्ण विचार व्यक्त करना चाहिए, और खुद से समझ में आना चाहिए।
  - आपके वाक्य को मुख्य अतिथिओं, वस्तुओं और उनके साथ हो रही चीज़ो को समझाना है।
- क्या न करें -
  - वाक्य में कोई अंक नहीं होना चाहिए।
  - वाक्य में किसी फिल्म या एक्टर का नाम नहीं होना चाहिए।
  - वाक्य में अदृश्य वस्तुओं और कार्यों का उल्लेख नहीं होना चाहिए।
  - वाक्य में अपने व्यक्तिगत राय न डालें।

याद रखें, गिफ में जो दिख रहा है उसी के बारे में लिखे। मुख्य पात्रों और उनके कार्यों पर ध्यान दें।

வழிமுறைகள் (Instructions for GIF task in Tamil)

கீழே ஐந்து அனிமேஷன் செய்யப்பட்ட கிப் (GIF) காட்டப்பட்டுள்ளன. ஒவ்வொரு GIF ஐ ஒரு தமிழ் வாக்கியத்தில் விவரிப்பதே உங்கள் பணி. GIF இல் உள்ள காட்சியில் மட்டுமே நீங்கள் கவனம் செலுத்த வேண்டும். GIF இல் உள்ள முக்கிய கதாபாத்திரங்களையும் அவற்றின் செயல்களையும் வர்ணிக்க வேண்டும், ஆனால் உங்கள் கருத்துக்கள், யூகங்கள் அல்லது விளக்கங்கள் அல்ல. ஒவ்வொரு வாக்கியமும் இலக்கணப்படி சரியாக இருக்க வேண்டும்.

- செய்க:
  - தமிழ் வார்த்தைகளை மட்டும் பயன்படுத்தவும். எண்களை வார்த்தையில் எழுதவும் (3 -> மூன்று)
  - வாக்கியங்கள் மிகக் குறுகியதாகவோ அல்லது நீண்டதாகவோ இருக்கக்கூடாது.
  - ஒவ்வொரு வாக்கியத்திலும் ஒரு வினை (செயலை குறிக்கும் வார்த்தை) இருக்க வேண்டும்.
  - முடிந்தால், வண்ணங்கள், அளவு, உணர்ச்சிகளை விவரிக்கும் வார்த்தைகளை சேர்க்கவும்.
  - இலக்கணம் மற்றும் எழுத்துப்பிழைகள் இல்லாதப்படி எழுதவும்.
  - ஒவ்வொரு வாக்கியமும் முழுமையாக இருக்க வேண்டும், மேலும் வாக்கியத்தை தனியாகப் படித்தால் அர்த்தம் புரிய வேண்டும்.।
  - வாக்கியம் GIFஇல் உள்ள முக்கிய கதாபாத்திரங்கள், செயல்கள், அமைப்பு, மற்றும் பொருள்களுக்கு இடையிலான உறவை விவரிக்க வேண்டும்.
- செய்யாதீர்:
  - வாக்கியத்தில் எந்த எண்களும் இருக்கக்கூடாது.
  - வாக்கியத்தில் எந்த திரைப்படம், மற்றும் நடிகர் அல்லது கதாபாத்திரத்தின் பெயரைக் குறிப்பிடக்கூடாது.
  - வாக்கியத்தில் கண்ணுக்கு தெரியாத பொருள்கள் மற்றும் செயல்களைக் குறிப்பிடக்கூடாது.
  - வாக்கியத்தில் தங்களின் எண்ணங்களோ தீர்மானங்களோ இருக்கக்கூடாது.

கவனிக்கவும்: அனிமேஷன் செய்யப்பட்ட GIF இல் காணும் காட்சியை மட்டும் வர்ணிக்கவும். அதில் இருக்கும் முக்கிய கதாப்பாத்திரங்கள் மற்றும் செயல்களில் கவனம் செலுத்தவும்.

Figure 2: Instructions for the data collection via images in English, via GIFs in Hindi and Tamil.