Why Find the Right *One*?

Payal Khullar
Language Technologies Research Centre
International Institute of Information Technology Hyderabad
Gachibowli, Hyderabad, India.
`payal.khullar@research.iiit.ac.in`

Abstract

The present paper investigates the impact of the anaphoric one words in English on the Neural Machine Translation (NMT) process using English-Hindi as source and target language pair. As expected, the experimental results show that the state-of-the-art Google English-Hindi NMT system achieves significantly poorly on sentences containing anaphoric *ones* as compared to the sentences containing regular, non-anaphoric *ones*. But, more importantly, we note that amongst the anaphoric words, the noun class is clearly much harder for NMT than the determinatives. This reaffirms the linguistic disparity of the two phenomenon in recent theoretical syntactic literature, despite the obvious surface similarities.

1 Introduction

English has three distinct lexemes spelled as one– the regular third person indefinite pronoun, such as in (1); the indefinite cardinal numeral (determinative), such as in (2); and regular common count noun, such as in (3).

1. One must obey the laws of the state at all times.

2. Could you pass me one one glass of water here.

3. It is important that we take care of our loved ones.

A visible difference in their orthographic base form is not observable. However, these can be totally differentiated on the basis of their morphological, syntactic, and semantic functions in the language. Note that the examples presented in (1), (2) and (3) are non-anaphoric one words. Coming to the anaphoric class of *one*– we have two subtypes. The first one belongs to the determinative category, as seen in (4); and the second one is a noun, as in (5).

4. I bought three red glasses, but she bought only one.

5. After looking at all the glasses, I decided to buy this small one.

As expected, the determinative anaphoric ones behave like a determiner, and the one-anaphora behave like nouns in a sentence. Note that the plural form of the determinative one in example (4) is *some*, but that of one-anaphora in (5) is *ones*. They are also different with respect to the kind of antecedents they take. The constituent whose repetition the determinative anaphora avoids is the whole NP, *a glass*. But in case of one-anaphora, it is the noun head optionally with one or more of its modifiers *red glass*, but never the whole NP (Payne et al., 2013).

Like other cohesive devices like pronouns and ellipsis, anaphoric ones make language less redundant and more engaging (Menzel, 2017; Mitkov, 1999; Halliday and Hasan, 1976). Resolving the information encoded in such structures is not hard for humans as they can easily disambiguate meanings from linguistic or extralinguistic context, cognitive commonsense extension as well as logical reasoning (Chen, 2016). However, all of this is not that straightforward for a machine. In fact, anaphoric ones can potentially present a special challenge for Machine Translation (MT) since the meaning of the word does not come from its most frequent usage as a cardinal number, but instead relies on its context, thereby becoming unavailable overtly at the surface syntax for text processing.

2 Previous Work

The determinative anaphoric ones have been discussed majorly as an instance of noun ellipsis, nominal ellipsis or noun phrase ellipsis (NPE) in linguistics (Halliday and Hasan, 1976; Dalrymple et al., 1991b; Lobeck, 1995; Lappin, 1996; Hobbs and Kehler, 1997; Hardt, 1999; Johnson,

2001; Wijnen et al., 2003; Merchant, 2004; Frazier, 2008; Chung et al., 2010; Merchant, 2010; Goksun et al., 2010; Gunther, 2011; Rouveret, 2012; Lindenbergh et al., 2015; van Craenenbroeck and Merchant, 2013; Park, 2017; Hyams et al., 2017; Kim et al., 2019). One-anaphora, on the other hand, has been referred to as noun anaphora, one-insertion, one-substitution and pronominalization (Menzel, 2017, 2014; Kayne, 2015; Hankamer and Sag, 2015; Payne et al., 2013; Corver and van Koppen, 2011; Gunther, 2011; Culicover and Jackendoff, 2005; Akhtar et al., 2004; Cowper, 1992; Luperfoy, 1991; Dalrymple et al., 1991a; Dahl, 1985; Radford, 1981; Baker, 1978; Halliday and Hasan, 1976; Bresnan, 1971).

To the best of our knowledge, the earliest computational approach to one-anaphora comes from Gardiner (2003), who presents several linguistically-motivated heuristics to distinguish one-anaphora from other non-anaphoric uses of one in English, and later from Ng (2005) that uses Gardiner's heuristics as features to train a simple Machine Learning (ML) model. Another seminal work on the anaphoric *one* comes from Recasens et al. (2016) where it has been treated as one of the several sense anaphoric relations in English. The authors create sAnaNotes corpus where they annotate one third of the OntoNotes corpus for sense Anaphora. They use a Support Vector Machine (SVM) classifier - LIBLINEAR implementation (Fan et al., 2008) along with 31 lexical and syntactic features, to distinguish between the anaphoric and the non-anaphoric class. Trained and tested on one-third of the OntoNotes dataset annotated as the SAnaNotes corpus, their system achieves 61.80% F1 score on the detection of all anaphoric relations, including one-anaphora. The detection and resolution of the determinative one anaphor, on the other hand, has been carried out as a part of computational research on noun ellipsis (Khullar et al., 2020b, 2019).

Recent research shows that discourse devices such as pronominal anaphora, ellipsis, deixis and lexical cohesion create inconsistencies in MT output (Voita et al., 2019; Mitkov, 2004). Unlike these discourse devices, however, the exact role of anaphoric ones in NLP tasks such as MT has not been studied. In the present paper, we conduct a data-driven study to study this extent and nature of this impact, using English and Hindi as source and target language pairs.

Point	Fluency	Adequacy
4	Flawless	Perfect/Ideal
3	Few errors	Mostly correct
2	Many errors	Somewhat correct
1	Unacceptable	Unrelated to source

Table 1: 4-Point Numeric scale for judging the fluency and adequacy of the translations.

3 Experiment

3.1 Curating Test sets

We prepare three test sets– the first containing sentences with determinative anaphoric ones; the second containing one-anaphora; and the third containing regular non-anaphoric one words. For the first test set, we randomly choose 750 sentences from the NoEl corpus (Khullar et al., 2020b), the curated dataset prepared by (Khullar et al., 2019) and the sAnaNotes corpus (Recasens et al., 2016); for the second, we take 750 sentences from (Khullar et al., 2020a) and (Recasens et al., 2016); and for the third, pick 750 sentences each from Cornel movie dialogs dataset (Danescu-Niculescu-Mizil and Lee, 2011) and The British National Corpus (2001), manually checked to contain non-anaphoric ones. We also undertake translation of these 2,250 sentences to assist automatic evaluation. The translation is carried manually by a professional translator, who is bilingual in English and Hindi. We get up to three translations for each sentence, which are then verified by a native Hindi speaker.

3.2 Obtaining Translations

To get the English sentences translated into Hindi, we use Google NMT (GNMT). The system comprises a deep LSTM network with 8 encoder and 8 decoder layers with attention and residual connections (Wu et al., 2016). It serves us well for our experiment as its performance is at par with the current state-of-the-art NMT systems and is also freely available for translations between English and Hindi. This system is run on the three test sets and the translations are saved for analysis.

3.3 Evaluation

In automatic evaluation, we get a BLEU (Bilingual Evaluation Understudy Score) (Papineni et al., 2002) score of 39.72 for the sentences in the first test set, 38.21 in the second and 41.46 in the third. We also try manual evaluation, where four evalua-

tors rate the translations of all sentences from the three test sets for their *fluency*) or syntactic correctness and (*adequacy* or translation accuracy. The evaluation of all the metrics is done on a 4-point Likert scale, see Table 1 for reference. The assigned scores by different raters are totalled and averaged for all the given sentences. We use the Fleiss's Kappa coefficient (Fleiss, 1971) to calculate the inter-annotator agreement between multiple evaluators. We get a score of 0.83 for fluency and 0.77 for adequacy that confirms reliability of the evaluation.

4 Results and Discussion

As can be seen, BLEU is the lower for the sentences containing anaphoric ones as compared to non-anaphoric ones. However, this may not be indicative of a trend as the test set is small and the difference observed is not that huge. Coming to manual evaluation, a total of 389 sentences from the first set, 601 from the second test set and 202 sentences from the third set get a rating of either 1 or 2 in the adequacy evaluation perspective. See Table 2. This shows that a majority of the sentences containing anaphoric one words are either poorly translated or have major translation quality errors, although they are grammatically still acceptable.

English (source text)	*The name is the same as the original one.*
Translation (Hindi)	नाम मूल एक के समान है
Transliteration	naam mool ek ke samaan hai
Gloss	नाम/name मूल/original एक/CRD(1) के/ADP समान/equal है/AUX
Meaning	The name is the same as original number one.

Figure 1: Translation of an English sentence containing one-anaphora to Hindi. The one-anaphora gets translated as non-anaphoric cardinal numeral one in the target language.

About 90% of the sentences containing the non-anaphoric instances of *one* are translated rather well by the system. Most of the errors observed are due to the incorrect translation of named entities and incorrect subject-verb agreement for gender marking. We do not encounter any errors that are caused due to incorrect translation of the word one in the target language.

English (source text)	*Because you have a hypothesis, an Important one.*
Translation (Hindi)	क्योंकि आप एक परिकल्पना है, एक महत्वपूर्ण है
Transliteration	Kyunkii aap ek parikalpanaa hai, ek mahatvapoorna hai
Gloss	kyunkii/because aap/you ek/one parikalpanaa/hypothesis hai/AUX ek/one mahatvapoorna/important hai/AUX
Meaning	Because you are a hypothesis, an Important one.

Figure 2: Translation of an English sentence containing determinative anaphoric *one* to Hindi. The one-anaphora gets translated incorrectly as a pronoun in the target language.

In comparison to the sentences contaning the non-anaphoric one words, the sentences containing anaphoric one words are translated much poorly. Within the latter, we note that the highest number of wrong translations are for the sentences with one-anaphora. The errors observed in such incorrect translations can be categorized into three types. In the first type, the anaphoric one words are translated into non-anaphoric *one* expressions, specifically as the cardinal numeral, in the target language. For example in Figure 1, the one-anaphora in the English sentence, which means *name* as seen from preceding context, gets translated as cardinal numeral one in the target language. Out of 750 sentences, a total of 232 sentences exhibit this error. One possible reason for this error could be the most common occurrence of the word *one* in English as a cardinal number (Gardiner, 2003). Hence, in case of ambiguity, the word *one* is more likely to be treated as a cardinal number by the MT system. The second type of errors are where the anaphoric one gets translated as a pronoun in the target language. Such errors occur very few times–only 25 from all sentences in our test set. See Figure 2 for one such example. Finally, in the third type of errors, the one-anaphora gets completely disregarded by the translation system and the translated sentence shows no equivalent lexeme to the anaphor. Note that these errors result into poor translation adequacy, but a majority of the translated sentence are more or less grammatically acceptable as per the rules of the target language, as seen in Figure 1 and Figure 2. They can, however, also become

Test set	Evaluation Perspective	1	2	3	4
Determinative Anaphora	Fluency	102	159	**291**	198
(750)	Adequacy	188	**210**	199	153
One-anaphora	Fluency	94	188	**354**	114
(750)	Adequacy	**309**	292	85	66
Non-anaphoric ones	Fluency	59	98	263	**330**
(750)	Adequacy	94	101	**304**	251
&					

Table 2: Evaluation scores of the sentences in the test sets containing determinative anaphoric *ones*, one-anaphora and non-anaphoric one words. There are 750 sentences in each test set. The highest values in each row are highlighted.

totally absurd in meaning in some cases, as can be seen in Figure 2.

As compared to one-anaphora, the severity of wrong translations for determinative anaphoric *ones* is slightly less. Hindi is morphologically richer as compared to English. We observe that the error in the translations come from copying of wrong agreement morphology on the verb in the absence of the noun whose repetition the determinative anaphoric *one* avoids. See Figure 3 for one such example. This also implies that although such sentences get a lower rating for fluency, they rate higher for translation adequacy.

From a long time in traditional syntactic literature, right from Baker (1978), one-anaphora and determinative anaphoric one words have been clubbed together, with frequent interchangeable uses of them in discussions and analysis. It is only recently (Payne et al., 2013) that the morphological, syntactic and semantic differences between the two anaphoric forms have been extensively discussed. Note that although recent work by Kayne (2015) aims to render all instances of the word *one* a homogeneous internal structure, comprising a classifier merged with an indefintive article through a variety of examples, he too identifies subtypes within this class and points out how they behave differently than one another. Our simple experiment highlights the differences between these two forms, restating their linguistic analysis and advocating for a disparate treatment for them in future Computational Linguistics and NLP research.

Finally, in the sentences that are correctly translated, we observe that a majority of the one-anaphora and the determinative anaphoric ones get translated exactly into their antecedent. This means

English (source text)	*She bought two baskets from the store, so we only took one.*
Translation (Hindi)	उसने स्टोर से दो बास्केट खरीदी, इसलिए हमने केवल एक ही लिया।
Transliteration	usne store se do basket khariidii, isliye hamne keval ek hii liyaa
Gloss	us/she ne/ERG store se/ADP do/ CRD(2) basket khariidii/bought.**Fem**, isliye/so ham/we ne/ERG keval/only ek/ CRD(1) hii/PRT liyaa/took.**Mas**
Meaning	*She bought two baskets from the store, so we only took one.*
Notes	Wrong gender agreement on the verb.

Figure 3: Translation of an English sentence containing determinative anaphoric one to Hindi. Although the translation is fine, the wrong agreement morphology on the verb makes it grammatically incorrect.

that the anaphoric expression *per se* is lost in the target language. For instance, the corresponding expression for one-anaphora in Hindi is *vaala* (singular, masculine). We see only 69 out of 750 translated sentences actually containing this lexeme. It is not surprising that 66 out of such sentences are rated 4 in the evaluation.

It is debatable, however, to claim that a translation that contains an anaphoric expression similar to the source is of better quality as compared to the translation that only copies the antecedent and replaces the anaphor with it. While both achieves nearly the same meaning and are grammatically acceptable, in our experiment, the former type were rated higher. It could be, then, argued that the latter added redundant information which might not be desirable in most cases.

5 Conclusion

In the present paper, we performed a simple experiment to investigate the impact of anaphoric and non-anaphoric one words on Neural Machine Translation process using English and Hindi as source and target language pair. Evaluation by manual methods revealed that anaphoric instances of the word one are much harder to translate as compared to the non-anaphoric one words. We also conclude that within the anaphoric class, one-anaphora are harder to translate than determinative anaphors, which reaffirms the linguistic disparity between the two phenomenon as shown in recent syntactic research. The long term goal of such a study is to improve the quality of translation of discourse structures such as anaphoric ones.

References

Nameera Akhtar, Maureen Callanan, Geoffrey K Pullum, and Barbara C Scholz. 2004. Learning antecedents for anaphoric one. *Cognition*, 4:141â145.

Carl Lee Baker. 1978. Introduction to generative transformational syntax. *Englewood Cliffs, NJ:: Prentice-Hal*.

Joan Bresnan. 1971. A note on the notion âidentity of sense anaphoraâ. *Linguistic Inquiry*, 2:589â597.

Wei Chen. 2016. The motivation of ellipsis. *Theory and Practice in Language Studies*, 6(11):2134–2139.

Sandra Chung, William Ladusaw, and James Mc-Closkey. 2010. Sluicing (:) between structure and inference. *In Representing language: Essays in honor of Judith Aissen*.

The British National Corpus. 2001. *Oxford University Computing Services on behalf of the BNC Consortium*, (2).

Norbert Corver and Marjo van Koppen. 2011. Np-ellipsis with adjectival remnants: a micro-comparative perspective. *Natural Language & Linguistic Theory*, 29(2):371–421.

Elizabeth A. Cowper. 1992. A concise introduction to syntactic theory. *Chicago, IL: University of Chicago Press*.

Jeroen van Craenenbroeck and Jason Merchant. 2013. Ellipsis phenomena. In *The Cambridge Handbook of Generative Syntax*, pages 701–745.

Peter W Culicover and Ray Jackendoff. 2005. Simpler syntax. *Oxford, England: Oxford University Press*.

Deborah Anna Dahl. 1985. The structure and function of one-anaphora in english. *Ph.D. thesis, University of Minnesota*.

Mary Dalrymple, Stuart M. Shieber, and Fernando C.N. 1991a. Ellipsis and higher order unification. *Linguistics and Philosophy*, 14:399â452.

Mary Dalrymple, Stuart M. Shieber, and Fernando C. N. Pereira. 1991b. Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14(4):399–452.

Cristian Danescu-Niculescu-Mizil and Lillian Lee. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, ACL 2011*.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. Liblinear: A library for large linear classification. volume 9, page 1871â1874.

Joseph Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76.

Lyn Frazier. 2008. Processing ellipsis: A processing solution to the undergeneration problem? In *Proceedings of the 26th West Coast Conference on Formal Linguistics*.

Mary Gardiner. 2003. *Identifying and resolving one-anaphora*. Department of Computing, Division of ICS, Macquarie University.

Tilbe Goksun, Tom W. Roeper, Kathy Hirsh-Pasek, and Roberta Michnick Golinkoff. 2010. From noun-phrase ellipsis to verbphrase ellipsis: The acquisition path from context to abstract reconstruction.

Christine Gunther. 2011. Noun ellipsis in english: adjectival modifiers and the role of context. *The structure of the noun phrase in English: synchronic and diachronic explorations*, 15(2):279–301.

Michael Alexander Kirkwood Halliday and Ruqaiya Hasan. 1976. Cohesion in english. *Longman London*, page 76.

Jorge Hankamer and Ivan Sag. 2015. Deep and surface anaphora. *Linguistic Inquiry*, 7:391–428.

Daniel Hardt. 1999. Dynamic interpretation of verb phrase ellipsis. *Linguistics and Philosophy*, 22(2):187–221.

Jerry R. Hobbs and Andrew Kehler. 1997. A theory of parallelism and the case of vp ellipsis. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, ACL '98/EACL '98, pages 394–401, Stroudsburg, PA, USA. Association for Computational Linguistics.

Nina Hyams, Victoria Mateu, and Lauren Winans. 2017. Ellipsis meets wh-movement: sluicing in early grammar.

Kyle Johnson. 2001. What vp ellipsis can do, and what it canât, but not why. pages 439–479. In The Handbook of Contemporary Syntactic Theory, Mark Baltin and Chris Collins (eds.). Blackwell Publishers.

Richard S Kayne. 2015. English one and ones as complex determiners. *New York University.*

Payal Khullar, Allen Anthony, and Manish Shrivastava. 2019. Using syntax to resolve npe in english. In *Proceedings of Recent Advances in Natural Language Processing*, pages 535–541.

Payal Khullar, Arghya Bhattacharya, and Manish Shrivastava. 2020a. Finding the right one and resolving it. In *Asian chapter of Association of Computational Linguistics.*

Payal Khullar, Kushal Majmundar, and Manish Shrivastava. 2020b. Noel: An annotated corpus for noun ellipsis in english. In *Language Resources Evaluation Conference.*

Nayoun Kim, Laurel Brehm, and Masaya Yoshida. 2019. The online processing of noun phrase ellipsis and mechanisms of antecedent retrieval. *Language, Cognition and Neuroscience*, 34(2):190–213.

Shalom Lappin. 1996. The interpretatin of ellipsis. In *The Handbook of Contemporary Semantic Theory*, pages 145–176. Blackwell Publishers.

Charlotte Lindenbergh, Angeliek van Hout, and Bart Hollebrandse. 2015. Extending ellipsis research: The acquisition of sluicing in dutch. *BUCLD 39 Online Proceedings Supplement*, 39.

Anne Lobeck. 1995. *Functional Heads, Licensing, and Identification*. Oxford University Press.

Susann Luperfoy. 1991. Discourse pegs: A computational analysis of context-dependent referring expressions. *Ph.D. thesis, University of Texas at Austin.*

Katrin Menzel. 2014. A corpus linguistic study of ellipsis as a cohesive device1. *Proceedings of Corpus Linguistics.*

Katrin Menzel. 2017. *Understanding English-German contrasts: a corpus-based comparative analysis of ellipses as cohesive devices*. Ph.D. thesis, Universitat des Saar- Â¨ landes, Saarbrucken.

Jason Merchant. 2004. Fragments and ellipsis. *Linguistics and Philosophy*, 27(6):661–738.

Jason Merchant. 2010. *Three Kinds of Ellipsis: Syntactic, Semantic, Pragmatic?*

R. Mitkov. 2004. Introduction: Special issue on anaphora resolution in machine translation and multilingual nlp. *Machine Translation*, 14:159–161.

Ruslan Mitkov. 1999. *Anaphora Resolution*. Oxford University Press.

Hwee Tou Ng, Yu Zhou, Rober Dale, and Mary Gardiner. 2005. A machine learning approach to identification and resolution of one-anaphora. pages 1105–1110.

Kishore Papineni, S. Roukos, T. Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *ACL.*

Dongwoo Park. 2017. *When does ellipsis occur, and what is elided?* PhD dissertation, University of Maryland.

John Payne, Geoffrey K. Pullum, Barbara C. Scholz, and Eva Berlage. 2013. Anaphoric one and its implications. *Language*, 4:794â829.

Andrew Radford. 1981. Transformational syntax: A studentâs guide to chomskyâs extended standard theory. *Cambridge, UK: Cambridge University Press.*

Marta Recasens, Zhichao Hu, and Olivia Rhinehart. 2016. Sense anaphoric pronouns: Am i one? page 1â6.

Alain Rouveret. 2012. Vp ellipsis, phases and the syntax of morphology. *Natural Language & Linguistic Theory*, 30(3):897â963.

Elena Voita, Rico Sennrich, and Ivan Titov. 2019. When a good translation is wrong in context: Context-aware machine translation improves on deixis, ellipsis, and lexical cohesion.

Frank Wijnen, Tom W. Roeper, and Hiske van der Meulen. 2003. Discourse binding: Does it begin with nominal ellipsis?

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Åukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144.