Squib

Are Ellipses Important for Machine Translation?

Payal Khullar
IIIT Hyderabad
khullar.payal@gmail.com

This article describes an experiment to evaluate the impact of different types of ellipses discussed in theoretical linguistics on Neural Machine Translation (NMT), using English to Hindi/Telugu as source and target languages. Evaluation with manual methods shows that most of the errors made by Google NMT are located in the clause containing the ellipsis, the frequency of such errors is slightly more in Telugu than Hindi, and the translation adequacy shows improvement when ellipses are reconstructed with their antecedents. These findings not only confirm the importance of ellipses and their resolution for MT, but also hint toward a possible correlation between the translation of discourse devices like ellipses with the morphological incongruity of the source and target. We also observe that not all ellipses are translated poorly and benefit from reconstruction, advocating for a disparate treatment of different ellipses in MT research.

1. Introduction

Ellipsis is a linguistic phenomenon in which parts of a sentence are omitted, and have to be retrieved from discourse or real-world context. For example, in (1), the phrase *like apples* is deleted at the site marked by [e] and can be understood from the context.

1. Kim **likes apples**, but Alex does not [e].

Ellipsis is a form of anaphora that often functions to reduce redundancy in language and improve discourse cohesion (Menzel 2017; Mitkov 1999). Languages provide various mechanisms to elide information, based on which different ellipses are defined in linguistics. For our study, following the theory of ellipses in Halliday and Hasan (1976) and Miller and Pullum (2013), we classify ellipses as nominal, verbal, and clausal. Nominal ellipses correspond to the deletion of the head noun, as in (2), sometimes with their dependents, as in (3). They are also called *head noun ellipsis* (McShane, Nirenburg, and Babkin 2015) and *Noun Phrase Ellipsis* (Corver and van Koppen 2011). A more recent, theory neutral term for such constructions is *fused head NPs* since the phrasal head here is realized jointly with a dependent function (Huddleston and Pullum 2002).

2. My sister's two **boys** are wild, but John's two [e] are really quite well-behaved.

3. They adopted Mary's **analysis on data** as John's [e] was poorly structured.

Ellipses occur in the environment of certain syntactical structures or trigger words, known as the licensors of ellipses. The nominal ellipsis in (2) is licensed by the cardinal number *two* and in (3) by the genitive proper noun *John's*. Demonstrative determiners, quantifiers, and so forth, can also license nominal ellipses (Khullar, Majmundar, and Shrivastava 2020; Khullar, Anthony, and Shrivastava 2019; Menzel 2017; Halliday and Hasan 1976).

Verbal ellipsis, verb ellipsis, or verb phrase ellipsis (VPE) is the deletion of the main verb, as in (4).[1] We also have instances of post-auxiliary ellipsis (PAE), as in (5), where the ellipsis is licensed by a modal or auxiliary verb (Sag 1976; Hankamer 1978).

4. I **checked** the hall and he [e] the living room.

5. Mary can **write with a fountain pen** but Jack cannot [e].

Finally, when the entire clause in a sentence gets deleted to avoid repetition of information, it is called clausal ellipsis, such as in (6). This phenomenon, also known as sluicing, is licensed by *wh*-words. Predicate ellipses, such as in (7), have been loosely put together in the clausal ellipsis category by Halliday and Hasan (1976). However, we identify them with instances of PAE where the negation is contracted to the auxiliary.

6. **We have a linguistics exam**, but I do not remember when [e].

7. She was always talking about who was **good-looking** and who wasn't [e].

The context from which an ellipsis[2] gets its sense and/or reference is called the antecedent.[3] If the antecedent is present textually, the ellipsis is endophoric. However, if the ellipsis cannot be recovered from a co-text, it is exophoric (Miller and Pullum 2013) or situational ellipsis, such as in (8), where the interlocutors infer the missing information using situational cues and the knowledge of the grammar of the language.

8. I will take two [e].

For this study, we do not take into account closely related phenomenon like do-so anaphora, such as in (9), where the exact site of ellipsis is not evident and one-anaphora, such as in (10), where a noun gets replaced with a non-lexical proform entity *one*. They are usually discussed as cases of *substitution* rather than ellipsis.

9. She **swims really fast** and so do I.

10. The upper **room** is smaller than the lower one.

---

1 It is also known as gapping in some linguistic textbooks.
2 We mark the site of ellipsis by [e] throughout this article.
3 Following the standard linguistic notation, we denote the antecedent of the ellipsis **like this**.

Ellipses are not very frequent in text,[4] but for improving the accuracy of Natural Language Processing (NLP) systems that handle data with ellipses, they are important (Zhang et al. 2019; Dean, Cheung, and Precup 2016). One such NLP application could be Machine Translation (MT). This is because the elided parts of the text are unavailable overtly at the surface syntax for text processing, and their meaning may come from context that is often present outside the current sentence, or, in some cases, may not be endophorically available at all. Thus, the MT process would involve representation of this missing information from the source correctly into the target, which could become more challenging when the former and the latter exhibit different strategies to elide information. The empirical evidence to confirm the extent of this impact is sparse. In this article, we conduct a data-driven study to gauge the size of the research problem we are addressing for different ellipsis types using English–Hindi/Telugu as source and target language pairs. Both English and Hindi belong to the Indo-European language family and, hence, show some linguistic similarities, although Hindi is more inflectional and morphologically richer than English. Telugu, on the other hand, is an agglutinative language from the Dravidian language family, and is rather unrelated to English. Selecting these language pairs allows us to assess the errors in relation to the degree of morphological dissimilarity between the source and the target languages.

2. Previous Work

Ellipsis has been thoroughly studied in theoretical linguistics (Halliday and Hasan 1976; Hankamer 1978; Lobeck 1995; Merchant 2004, 2010; Gunther 2011; van Craenenbroeck and Merchant 2013; Miller and Pullum 2013; Park 2017), in cognitive linguistics (Kim, Brehm, and Yoshida 2019), and in language acquisition studies (Hyams, Mateu, and Winans 2017; Lindenbergh, van Hout, and Hollebrandse 2015; Goksun et al. 2007; Wijnen, Roeper, and van der Meulen 2003). Previous computational work on ellipsis resolution has mostly focused on VPE, gapping, and sluicing; for instance, the detection of VPE in the Penn Treebank using pattern match (Hardt 1992), a transformation learning-based approach to generated patterns for VPE resolution (Hardt 1998), the domain independent VPE detection and resolution using machine learning (Nielsen 2003), automatically parsed text (Nielsen 2004), sentence trimming methods (McShane, Nirenburg, and Babkin 2015), linguistic principles (McShane and Babkin 2016), improved parsing techniques that encode elided material dependencies for reconstruction of sentences containing gapping (Schuster, Nivre, and Manning 2018), discriminative and margin infused algorithms (Dean, Cheung, and Precup 2016), and Multilayer Perceptrons and Transformers (Zhang et al. 2019). Computational work on noun ellipsis is comparatively sparse, comprising a simple rule-based system (Khullar, Anthony, and Shrivastava 2019), an annotated corpus for noun ellipsis in movie dialogues (Khullar, Majmundar, and Shrivastava 2020), and end-to-end resolution pipeline experiments with statistical and neural model experiments (Khullar 2020).

The main source of inspiration for this empirical study comes from the recent work on MT for English–Russian by Voita, Sennrich, and Titov (2019), where VPE has been identified as one of the linguistic phenomena that cause inconsistencies in translation output, along with discourse structures like deixis and lexical cohesion. Using their

---

4 The reported frequency of noun ellipses is 1.99% (Khullar, Majmundar, and Shrivastava 2020); and that of VPE along with related phenomenon is 1% (Bos and Spenader 2011).

findings as the starting point, we conduct an empirical study to determine the impact of different ellipses on an existing NMT system for English to Hindi/Telugu.

3. Experimental Setup

We prepare two test sets—the first one contains sentences with the three aforementioned ellipsis types and the second the same sentences with resolved ellipses. We gather the sentences from various popular annotated corpora, such as the VPE corpus by Bos and Spenader (2011), the NoEl corpus (Khullar, Majmundar, and Shrivastava 2020), a curated ellipses dataset by Khullar, Anthony, and Shrivastava (2019), and the GECCo Corpus (Menzel and Lapshinova-Koltunski 2014). For consistency and fair analysis, we randomly pick 500 sentences for each ellipsis type, which results in a total of 1,500 sentences. The antecedent of the ellipsis is frequently present in the same sentence as the ellipsis. But it can also be present in the previous or following sentence, although the latter is comparatively rare (Khullar, Majmundar, and Shrivastava 2020). For our study, we pick sentences where the ellipsis and its antecedent occur in the same sentence. Hence, they can be handled by MT systems that only operate sentence by sentence.

For the second test set, we manually reconstruct the ellipsis with their resolution marked in the respective ellipsis corpora. Reconstruction is one of the acceptable ways to resolve ellipses in some linguistic theories that consider resolution as involving searching for some antecedent that could be substituted at the ellipsis site to produce a well-formed string with the same meaning the elided string provides (Lappin and Shih 1996; Chomsky 1995; Wasow 1972). Thus, for a sentence like (11), the ellipsis reconstruction procedure leads to (12).

11. You definitely saved Kendall's life today, but not Pike's.

12. You definitely saved Kendall's life today, but not Pike's life.

Because the antecedents do not always occur strictly under identity with the elided material, reconstructed sentences such as (13) and (14) become grammatically incorrect. In the ellipses test set, we find 38 such sentences involving VPE or PAE and 12 nominal ellipses. All these errors are caused by a mismatch in agreement morphology. We manually correct the sentences and verify them with a native English speaker.

13. *I gave him one **pencil**, but he wanted three [pencil].

14. *John **lives with his grandparents**, but Bill does not [lives with his grandparents].

We create the second test set to check if reconstruction offers any advantage in MT research. We use manually reconstructed, gold sentences to analyze the exact impact of this procedure on MT. In practice, the text can be fed into an ellipses resolution system (Khullar 2020; Zhang et al. 2019) as a preprocessing step. This means that the accuracy of such a system will also contribute to the final translation quality.

To obtain the translations, we use the Google NMT, which comprises a deep LSTM network with 8 encoder and 8 decoder layers with attention and residual connections (Wu et al. 2016). It is freely available for translations between English and nine Indian languages, including Hindi and Telugu. It is competitive to the state-of-the-art MT and allows us to use the same underlying model for both of the language pairs.

Table 1
Evaluation categories for the translated sentences.

Category	Summary
A	Acceptable translation. Source & target have similar ellipsis strategy.
B	Acceptable translation. Source & target have different ellipsis strategy.
C	Acceptable translation. Source has ellipsis, target does not.
D	Small grammatical error(s), but meaning comprehensible.
E	Significant grammatical error(s), questionable interpretation.
F	Grammatically acceptable, but meaning slightly changed/ambiguous.
G	Grammatically acceptable, but meaning completely lost.

4. Results and Discussion

We opt for manual evaluation[5] to focus on the translation of parts of the sentence containing the ellipsis and its antecedent. For each language pair, two linguists[6] assign a category to a translated sentence, from a list of 7 proposed categories, summarized in Table 1. The sentences for which both the linguists assign the same category are separated out directly for analysis. The inter-annotator agreement is high (0.89 for Hindi and 0.91 for Telugu), which indicates reliability of our evaluation efforts. For the sentences where the category labels mismatch, the linguists discuss the dispute and check whether they can agree to assign the same category. There is no unresolved disagreement for any sentence at the end. Therefore, no sample is disregarded from analysis.

The correctly translated English sentences are further analyzed for the representation of the ellipses. When the source and target show a similar ellipsis strategy, as in (15), they are assigned the category A. For each source **(S)** and target **(T)** pair, we present a gloss **(G)** of the latter as per the Leipzig Glossing Rules. To avoid repetitiveness, we add the meaning **(M)** of the target only when it is different from the source.

15. **S** She bought a car but I don't know when.
    T *Āme kāru konnad-i   kānī eppuu uundō nāku teliya-du*
    G she car   bought-a but when be   I   know-NEG

The assigned category is B when the target sentence has a different ellipsis strategy than the source; however, the meaning remains unchanged. For example, in (16), the target has noun modifier coordination and not ellipsis.

16. **S** I will just take a minute or two.
    T *mujh-e   bas ek - do   minat   lagenge*
    G 1SG-OBL just one - two   minute   take-FT

---

5 We recognize that the empirical evaluation of this work is limited. Because we only examine low resource language pairs, we cannot say with certainty how much of the problem disappears with increasing amounts of training data, and how much it is a fundamental problem that requires different models.

6 The linguists are proficient bilinguals in English and the respective target language and also have translation/localization experience.

Table 2
Evaluation categories assigned to the translated Hindi and Telugu sentences.

Target	Test Set	Ellipses	Categories						
			A	**B**	**C**	**D**	**E**	**F**	**G**
Hindi	Ellipses	Noun	25	58	33	84	65	153	82
		Verbal	3	80	11	68	79	166	93
		Clausal	224	0	0	0	0	175	101
Telugu	Ellipses	Noun	19	27	19	56	121	157	88
		Verbal	11	23	14	63	99	173	117
		Clausal	186	0	0	0	0	159	125

The category C is for samples in which the target does not have the ellipsis seen in the source, such as in (17), although the meaning is perfectly localized.

17. **S** But you knew that, didn't you?
    T *Kānī mīku adi telusu, kādā?*
    G but you that knew right
    M 'But you knew that, right?'

4.1 Error Analysis

Out of the 1,500 sentences from the first test set, the Hindi translations of 1,066 sentences and the Telugu translations of 1,201 sentences receive a label from D–G categories (see Table 2). Hence, over 70% of the translations in both languages are poor. The higher frequency of errors in Telugu could hint toward a possible relation between the translation of ellipses with the degree of morphological dissimilarities between the source and the target.

Among the incorrect translations, the number of sentences assigned F/G categories is far greater than the number of sentences assigned D/E. This implies that despite the translation errors, most of these sentences are grammatically still acceptable. In other words, the translation adheres to the target language (*fluency*), but does not capture the source text well (*adequacy*). This is in line with the observation made in Voita, Sennrich, and Titov (2019) that the translation of a sentence containing a discourse structure such as ellipsis often looks correct when read independently but not in context. We also note that most of the errors are located in the phrase containing the ellipsis, indicating that translating elided parts of a sentence is indeed hard. We now analyze the errors for each ellipsis type.

4.1.1 Noun Ellipsis. In the first type of error from category D, the translated sentence is fairly comprehensible, but has small grammatical errors. These errors are contributed by wrong agreement morphology between the elided noun (and/or the noun modifiers) and the verb. For example, in (18), the Hindi word for *gave* has masculine gender, which is incorrect as the elided noun *baskets* in Hindi bears feminine gender, resulting in a gender agreement mismatch between the subject and verb. We display the errors in red.

18. **S** She brought three baskets, and gave us one.
    T *vah teen tokariyaan laee, aur ham-en ek di-ya
    G she three baskets(F) brought, and 1PL-ACC one give-M.PERF

The phrase containing the ellipsis is sometimes translated literally from the source into the target, even though the latter does not have the same ellipsis strategy. This results in a grammatically weird construction, as in (19). The translators agreed that there should have been clausal coordination in this sentence, as without it the adjective *scary* does not necessarily modify the elided noun *costume* in the target.

19. **S** We are looking for the funniest costume, and the scariest.
    T ham sabase majedaar poshaak-ki talaash kar rahe hai aur sabase daraavane
    G 1PL most funny costume-ACC search do PROG PRS and most   scary

When the meaning of the elided noun is slightly changed or ambiguous in the target, it results in the errors from category F. For example, the target sentence in (20) reads weirdly due to the incorrect translation of the intended meaning of the NP *mine*.

20. **S** I drove my friends' car today as mine was in a workshop.
    T gani varkāplō unnanduna nēnu īrōju nā snēhitu-la kārunu naipānu
    G thing workshop because   i   today my friend-GEN car   drove
    M 'Something is in the workshop because I drove my friend's car.'

Finally, when the meaning of the elided noun is completely lost in the target, it results in the errors from category G, such as in (21), where the ellipsis is so poorly translated that the intended meaning *his story* is not present at all in the target.

21. **S** Everyone believed her story as his wasn't all that dramatically told.
    T Nāakīyagā ceppabainadi antā āme katha kādani andarū viśvasincāru
    G dramatically being-said everything her story not   everyone   believed
    M 'Not everything that was said dramatically was her story, everyone believed.'

4.1.2 Verbal Ellipsis. We find small grammatical errors like agreement feature mismatch in the sentences with verbal ellipses as well. For example in (22), the subject in the clause containing the ellipsis misses the ergative marker, making the sentence grammatically incorrect. The meaning, however, is still comprehensible.

22. **S** Mr. Wilson taught chemistry and his wife physics.
    T *shree vilsan ne rasaayan vigyaan aur una-kee patnee bhautikee padhaaya
    G Mr. Wilson ERG chemical science and 3PL-GEN wife physics teach-PERF

The most frequently observed error is the addition of an auxiliary or a *do* verb in the clause containing the elided verb. See example in (23) for Hindi and (24) for Telugu.

23. **S**: You either believe Seymour can do it again or you don't.
    T aap ya to   maanate hain ki semur ise   phirse kar–
    G you either PART believe PRS COMP Semur DEM-ACC again do
       – sakata hai ya aap nahin karate hain
         can   PRS or you not do-PERF PRS
    M 'You either believe Seymour can do it again or you do not do it.'

Table 3
Manual evaluation scores after reconstructed ellipses. Numbers are colored for emphasis.

Target	Fluency			Adequacy		
	−1	0	1	−1	0	1
Hindi	**656**	347	497	**0**	438	**1,062**
Telugu	**719**	233	548	**0**	286	**1,214**

24. **S** He has not changed, but those around him have.
    T atanu   māralēdu,   kānī atani cuū   unnavāru   unnāru
    G he not.have.changed, but his around those.who.have PST
    M 'He didn't change, but everyone around him are.'

This happens because the main verb cannot be completely dropped off in both the languages. The substitutions, although grammatically acceptable, often make the sentences weird and incomprehensible. We, thus, add such sentences to category F and G in our evaluation, depending upon the degree of meaning loss.

4.1.3 Clausal Ellipsis. Most sentences with clausal ellipsis are translated well. We do not find any grammatically incorrect translations related to ellipses, and so there are no samples in the D/E categories. The most common error observed in translation of these ellipses is the *wh*-word being incorrectly followed by an auxiliary, as in (25).

25. **S** Someone in the class was drawing a flower, but I couldn't see who.
    T kaksha mein koee vyakti ek phool kheench raha tha lekin main yah –
    G class LOC some person one flower sketch PERF PST but I this
       –nahin dekh sakata tha ki kaun hai
        not see can PST COMP what PRS
    M 'Someone in the class was drawing a flower, but I couldn't see who is there.'

Rarely, the *wh*-word is translated incorrectly, as in (26). Note that this sentence and the one in (25) are grammatically acceptable, although the meaning is somewhat altered.

26. **S** Ranjit is looking at someone, can you see who.
    T ranjeet kisee-ko   dekh raha hai, kya aap dekh sakate hain.
    G Ranjit someone-ACC look PERF PRS what you look can PRS
    M 'Ranjit is looking at someone, can you see?'

4.2 Reconstruction

We rate the reconstructed sentences in comparison to their counterparts from the first test set containing ellipsis. For both grammatical fluency and meaning adequacy perspectives, a 0 is assigned if the translation remains more or less the same (good or bad) as before, 1 if it shows improvement, and −1 if it becomes worse. See Table 3 for scores.
    Reconstruction corrects agreement mismatches in most of the sentences from Hindi and Telugu containing noun and verb ellipses from the D/E categories. For the

sentences in the F/G categories, this overt information improves meaning and relatedness to the source. For example, the sentence in (24) reconstructed as (27) is translated well.

27. **S** He has not changed, but those around him have **changed**.
    T atanu maraledu,   kanī atani cuu   unnavaru   mararu
    G he not.have.changed, but   his around those.who.have changed
    M 'He didn't change, but everyone around him have changed.'

If the error is not due to the ellipsis, like in (28), the reconstruction of the ellipsis, as in (29), does not improve the translation. However, it also does not make it worse.

28. **S** Some students in the class like physics and some don't.
    T *kaksha mein kuchh chhaatr jaise bhautikee aur kuchh nahin
       class   LOC some students like physics   and some   not
    M 'Some students in the class are like physics and some are not'.

29. **S** Some students in the class like physics and some don't **like physics**.
    T *kaksha mein kuchh chhaatr bhautikee kee   tarah hain aur kuchh–
    G class   LOC some students physics ACC similar PRS and some
       –bhautikee kee tarah nahin hain
       physics   ACC similar not   PRS
    M 'Some students in the class are like physics and some are not like physics'.

All in all, since the elided material gets overtly represented, this procedure is successful in improving the adequacy. In no sample does it make the meaning worse. A drawback, as discussed previously, is that it adds redundant information that lowers the fluency. Since the sentences with clausal ellipsis require repetition of an entire clause, their fluency is most negatively impacted. More importantly, since this procedure does not impact the translation of the wh-word, it is not of much use for clausal ellipsis.

5. Conclusion

Translating missing information that can be retrieved from elsewhere in the context poses an attractive goal for MT. We carried out an experiment to test the impact of different ellipses discussed in linguistics on NMT for English to Hindi/Telugu. The experimental results confirmed that ellipsis is hard for MT. We also found that ellipsis reconstruction is useful, mostly for sentences with noun and verb ellipses to improve their translation adequacy, although at the cost of their fluency.

References

Bos, Johan and Jennifer Spenader. 2011. An annotated corpus for the analysis of VP ellipsis. *Language Resources and Evaluation*, 45(4):463–494. https://doi.org/10.1007/s10579-011-9142-3

Chomsky, Noam. 1995. The minimalist program. *Current Studies in Linguistics 28*, pages 219–394. MIT Press.

Corver, Norbert and Marjo van Koppen. 2011. NP-ellipsis with adjectival remnants: A micro-comparative perspective. *Natural Language & Linguistic Theory*, 29(2):371–421. https://doi.org/10.1007/s11049-011-9140-6

Dean, Kian Kenyon, Jackie Chi Kit Cheung, and Doina Precup. 2016. Verb phrase ellipsis resolution using discriminative

and margin-infused algorithms. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1734–1743, Texas.

Goksun, Tilbe, Tom W. Roeper, Kathy Hirsh-Pasek, and Roberta Michnick Golinkoff. 2007. From noun phrase ellipsis to verb phrase ellipsis: The acquisition path from context to abstract reconstruction. In Jesse Harris and Margaret Grant, editors, *University of Massachusetts Occasional Working Papers in Linguistics: Processing Linguistic Structure*, Massachusetts, United States, pages 53–74.

Gunther, Christine. 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. https://doi.org/10.1017/S1360674311000049

Halliday, Michael, Alexander Kirkwood, and Ruqaiya Hasan. 1976. *Cohesion in English*, Longman London.

Hankamer, Jorge. 1978. On the non-transformational derivations of some null VP anaphors. *Linguistic Inquiry 9*, pages 66–74.

Hardt, Daniel. 1992. An algorithm for VP ellipsis. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, pages 9–14, Newark, DE. https://doi.org/10.3115/981967.981969

Hardt, Daniel. 1998. Improving ellipsis resolution with transformation-based learning. In *AAAI Fall Symposium*, pages 41–43, Orlando, FL.

Huddleston, Rodney D. and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press.

Hyams, Nina, Victoria Mateu, and Lauren Winans. 2017. Ellipsis meets wh-movement: Sluicing in early grammar. In LaCara, N., K. Moulton, and A.-M. Tessier, editors. *A Schrift to Fest Kyle Johnson*, Linguistics Open Access Publications, University of Massachusetts.

Khullar, Payal. 2020. Exploring statistical and neural models for noun ellipsis in English. In *Asian Chapter of Association for Computational Linguistics*, pages 34–43.

Khullar, Payal, 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. https://doi.org/10.26615/978-954-452-056-4_063

Khullar, Payal, Kushal Majmundar, and Manish Shrivastava. 2020. NoEl: An annotated corpus for noun ellipsis in English. In *Language Resources Evaluation Conference*, pages 34–43.

Kim, Nayoun, 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. https://doi.org/10.1080/23273798.2018.1513542

Lappin, Shalom and Hsue-Hueh Shih. 1996. A generalized reconstruction algorithm for ellipsis resolution. *COLING*, pages 687–692. https://doi.org/10.3115/993268.993288

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

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

McShane, Marjorie and Petr Babkin. 2016. Detection and resolution of verb phrase ellipsis. *Linguistic Issues in Language Technology*, 13(1):1–34. https://doi.org/10.33011/lilt.v13i.1385

McShane, Marjorie, Sergei Nirenburg, and Petr Babkin. 2015. Sentence trimming in service of verb phrase ellipsis resolution. In *EAPCogSci*, pages 228–233. https://doi.org/10.33011/lilt.v13i.1385

Menzel, Katrin. 2017. *Understanding English-German Contrasts: A Corpus-Based Comparative Analysis of Ellipses as Cohesive Devices*. Ph.D. thesis, Universitat des Saarlandes, Saarbrucken.

Menzel, Katrin and Ekaterina Lapshinova-Koltunski. 2014. Kontrastive analyse deutscher und englischer kohäsionsmittel in verschiedenen diskurstypen. *tekst i dyskurs - Text und Diskurs. Zeitschrift der Abteilung für germanistische Sprachwissenschaft des Germanistischen Instituts Warschau*.

Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy*, 27(6):661–738. https://doi.org/10.1007/s10988-005-7378-3

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

Miller, Philip and Geoffrey Keith Pullum. 2013. Exophoric VP ellipsis. In Philip Hofmeister and Elisabeth Norcliffe,

editors. *The Core and the Periphery: Data-Driven Perspectives on Syntax Inspired by Ivan A. Sag*, CSLI Publications, pages 167–220.

Mitkov, Ruslan. 1999. *Anaphora Resolution*, Oxford University Press. `https://doi.org/10.1075/scl.3.07mit`

Nielsen, Leif Arda. 2003. Using machine learning techniques for VPE detection. *Proceedings of RANLP*, pages 339–346.

Nielsen, Leif Arda. 2004. Verb phrase ellipsis detection using automatically parsed text. *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, pages 1093–1099. `https://doi.org/10.3115/1220355.1220512`

Park, Dongwoo. 2017. *When Does Ellipsis Occur, and What is Elided?* Ph.D. dissertation, University of Maryland.

Sag, Ivan Andrew. 1976. A note on verb phrase deletion. *Linguistic Inquiry*, 7:664–671.

Schuster, Sebastian, Joakim Nivre, and Christopher D. Manning. 2018. Sentences with gapping: Parsing and reconstructing elided predicates. In *Proceedings of North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1156–1168, New Orleans, LA. `https://doi.org/10.18653/v1/N18-1105`

van Craenenbroeck, Jeroen and Jason Merchant. 2013. Ellipsis phenomena. In *The Cambridge Handbook of Generative Syntax*. Cambridge University Press, pages 701–745. `https://doi.org/10.1017/CBO9780511804571.025`

Voita, Elena, 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. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1198–1212. `https://doi.org/10.18653/v1/P19-1116`

Wasow, Thomas. 1972. Anaphoric relations in English. Doctoral dissertation, MIT, pages 88–101.

Wijnen, Frank, Tom W. Roeper, and Hiske van der Meulen. 2003. Discourse binding: Does it begin with nominal ellipsis? *Lot Occasional Series*, pages 505–516.

Wu, Yonghui, 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, Łukasz 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.

Zhang, Wei Nan, Yue Zhang, Yuanxing Liu, Donglin Di, and Ting Liu. 2019. A neural network approach to verb phrase ellipsis resolution. *The Thirty-Third AAAI Conference on Artificial Intelligence*. `https://doi.org/10.1609/aaai.v33i01.33017468`