# Remarkable Similarity of Clausal Coordinate Ellipsis in Russian Compared to Dutch, Estonian, German, and Hungarian

**Karin Harbusch** and **Denis Krusko**
Computer Science Department
University of Koblenz–Landau, Koblenz, Germany
harbusch, kruskod@uni-koblenz.de

## Abstract

Elliptical constructions can help to avoid repetition of identical constituents during natural-language generation. From grammar books, it is not easy to extract executable rules for ellipsis—in our case in Russian. Therefore we follow a different strategy. We test the accuracy of a rule set that has been evaluated for the two Germanic languages, Dutch and German, and the two Finno-Ugric languages, Estonian and Hungarian. For a Russian test corpus of about 100 syntactically annotated coordinated sentences that systematically vary the conditions of rule application, our Java program can automatically produce all elliptical variants. Over- and undergeneration in the resulting lists have been tested in two experiments with native speakers. Basically, the rules work very well for Russian. Within the four target languages, Russian works best with the Estonian amendments. Here we report two slight deviations partially known from the linguistic literature.

## 1 Introduction

In natural-language generation, ellipsis can help to avoid repetition of identical constituents. For instance, the conceptual structure 'eat(Peter, apples) & eat(Mary, apples)' where 'eat' and 'apples' occur two times can be formulated as *Peter eats apples and Mary too*, a case of Stripping. However, many other paraphrases can be produced such as the aggregation into one sentence with NP-coordination (*Peter and Mary eat apples*)—a case of reduction we do not address in the following as it works on the conceptual structure whereas we only deal with syntactic structures as input.

Ellipsis occurs frequently in written and spoken language. In the following, we study four types of clausal coordinate ellipsis (CCE): (1) Gapping (including Long Distance Gapping (LDG), Subgapping and Stripping), (2) Forward Conjunction Reduction (FCR), (3) Backward Conjunction Reduction (BCR), and (4) Subject Gap with Finite/Fronted Verb (SGF). In German written text, clausal coordination, i.e., the two conjuncts comprise verbal constructions (not necessarily finite), occurs in 14 and ellipsis in at least one of the two conjuncts in 7 percent of the investigated corpus (cf. Harbusch & Kempen, 2007). All these types of clausal coordinate ellipsis also emerge in spontaneous speech in German (cf. Harbusch & Kempen, 2009a). This observation is in line with English corpus studies (see, e.g., Greenbaum & Nelson, 1999) and Dutch (Harbusch, 2011).

For recent theoretical treatments of CCE in various linguistic frameworks see, e.g., Schwabe & Zhang, 2000; Frank, 2002; Beaver & Sag, 2004; te Velde, 2006; Haspelmath, 2007; Johnson, 2009; Kempen, 2009; Van Craenenbroeck & Merchant, 2013; Griffiths & Lipták, 2014. For Russian as target language, see, e.g., Kazenin, 2006 or Gribanova, 2013. Parsing elliptical constructions is a difficult problem partially due to the fact that both conjuncts may be grammatically incorrect when viewed in isolation (see, e.g., Kobele, 2012). In a natural-language generation-system, CCE is only one realization option (cf. Shaw 1998) out of many (e.g., Pronominalization also avoids repeating the same NP). The implemented CCE-generation component ELLEIPO, which embodies the CCE rule set we present below, can serve as a post-editing component for NLG systems that provide a syntactic structure annotated with co-referentiality tags (cf. Harbusch & Kempen, 2009b). ELLEIPO takes these non-elliptical (redundant) structures as input and provides all reduced to CCE options as output. ELLEIPO was originally developed for Dutch and German (see Harbusch & Kempen, 2006), but the implemented



*Proceedings of the 5th Workshop on Balto-Slavic Natural Language Processing*, pages 75–80,
Hissar, Bulgaria, 10–11 September 2015.

set of CCE rules was designed in a language-independent manner. This makes it possible to discover CCE rules in a new target language. For the Finno-Ugric language Estonian, Harbusch, Koit & Õim (2009) report high accuracy of the rule set, which suggests that the entire process is language independent. However, Estonian is suspected to be strongly influenced by language contact with Germanic languages. Nevertheless, Hungarian, another Finno-Ugric language, yields equally good results (cf. Harbusch & Bátori, 2013).

In the present paper, we aim to further verify our claim that CCE can be generated by language-independent rules by testing ELLEIPO's rules for Russian. To this purpose, we built a test corpus of about 100 Russian syntactic structures of (unreduced) coordinated sentences in Russian varying the conditions for CCE-rule application. RUSSIAN-ELLEIPO produces all CCE reductions for the test corpus. In the first experiment, we let native speakers of Russian judge the quality of the output (*overgeneration* of the CCE rules). In the second, native speakers generated all reductions (inclusive Pronominalization etc.) for unreduced coordinated sentences in order to spot CCE realizations that ELLEIPO does not generate (*undergeneration*). In general, we observed a very high level of accuracy of the CCE rules.

The paper is organized as follows. In Section 2, we define the CCE phenomena ELLEIPO is able to generate. In Section 3, we describe the test corpus, and elaborate on ELLEIPO's output and on the user studies. In Section 4, we outline our results. In the final Section, we draw some conclusions and address future work.

## 2 Definition of the CCE Rules

We distinguish four types of CCE applicable to *binary and*-coordinations, and specify elision conditions on the first (*anterior*) and second (*posterior*) member of two conjoined clauses connected by the Russian equivalent *i*[1] of the coordinating conjunction *and* (cf. examples in Table 1).

The CCE rules of ELLEIPO are based on the psycholinguistically motivated definitions of CCE types proposed by Kempen (2009). They check the following conditions in syntactic trees whose inner nodes additionally provide 'referential identity features'.

[1] As in Russian, *a* 'but' is used for contrasts, we vary our examples in the Gapping test where contrast is mandatory.

| | |
|---|---|
| **GAPPING (g)** | (1) Ûrij **živët** v Tambove i ego synov´â **~~živut~~**$_{g}$<br>Ûrij live$_{3SG}$ in Tambov and his sons live$_{3PL}$<br>v Kaluge<br>in Kaluga<br>'Ûrij **lives** in Tambov and his sons **~~live~~**$_{g}$<br>in Kaluga.'<br><br>(2) Ûrij **živët** v Tambove i v Kaluge **~~živut~~**$_{g}$<br>Ûrij live$_{3SG}$ in Tambov and in Kaluga live$_{3PL}$<br>ego synov´â<br>his sons<br>'Ûrij **lives** in Tambov and in Kaluga, his sons **~~live~~**$_{g}$' |
| **LDG (g)$^{+}$g)** | (3) **Segodnâ dolžen** Pëtr svoû mašinu myt´ i<br>Today should Pëtr his car wash$_{INF}$ and<br>**~~segodnâ~~**$_{gg}$ **~~dolžna~~**$_{g}$ Maša svoj velosiped **~~myt´~~**$_{gg}$<br>today should Maša her bike wash$_{INF}$<br>'Today Pëtr **should wash** his car and **~~today~~**$_{gg}$<br>Maša **~~should~~**$_{g}$ **~~wash~~**$_{gg}$ her bike' |
| **SUBGAP-PING (g)** | (4) Ivan **hočet** spat´ a Pëtr **~~hočet~~**$_{g}$ mečtat´<br>Ivan want$_{3SG}$ sleep$_{INF}$ but Pëtr want$_{3SG}$ dream$_{INF}$<br>'Ivan **wants** to sleep but Pëtr **~~wants~~**$_{g}$ to<br>dream' |
| **STRIP-PING (str)** | (5) Â **splû** i ty **~~spiš~~**$_{str}$ tože<br>I sleep$_{1SG}$ and you sleep$_{2SG}$ too<br>'I **sleep** and you **~~sleep~~**$_{str}$ too' |
| **FCR (f)** | (6) **Cvetaevu** lûblû **â** i **~~Cvetaevu~~**$_{f}$<br>Cvetaeva$_{ACC}$ like$_{1SG}$ I and Cvetaeva$_{ACC}$<br>čitaû **~~â~~**$_{s}$ často<br>read$_{1SG}$ I often<br>'I like **Cvetaeva** and **~~Cvetaeva~~**$_{f}$ read **~~I~~**$_{s}$ often'<br><br>(7) Maša slyšala, **[čto Pëtr]** popal v<br>Mary hear$_{PST.SG.F}$ that Pëtr get$_{PST.SG.F}$ an<br>avariû i **[~~čto Pëtr~~]**$_{f}$ mog umeret´<br>accident and that Pëtr can$_{PST_{SG.M}}$ die$_{INF}$<br>'Mary heard **that Pëtr** had an accident and<br>**[~~that Pëtr~~]**$_{f}$ could die' |
| **BCR (b)** | (8) Maša **pridët** do trëh **~~časov~~**$_{b}$<br>Maša come$_{FUT.3SG}$ before three$_{ACC}$ o'clock<br>a Katâ **~~pridët~~**$_{g}$ posle četyrëh **čhasov**<br>but Katâ come$_{FUT.3SG}$ after four$_{ACC}$ o'clock<br>'Maša **will come** before three **~~o'clock~~**$_{b}$ and<br>Katâ **[~~will come~~]**$_{g}$ after four **o'clock**' |
| **SGF (s)** | (9) V les hodil **ohotnik**<br>Into forest$_{ACC.SG}$ go$_{PST.SG.M}$ the hunter<br>i podstrelil **~~ohotnik~~**$_{s}$ odnogo zajca<br>and shot$_{PST.SG.M}$ the hunter one$_{ACC.M}$ hare<br>'Into the forest went **the hunter** and<br>**[~~the hunter~~]**$_{s}$ shot a hare' |

Table 1: CCE examples in Russian (using the *ISO 9* transliteration standard for better readability). Crossed-out text represents elisions. Subscripts indicate CCE type. Elided constituents and their overt counterparts are marked in bold font.

1. *Gapping* ignores word order (compare example (1) with the marked word order in (2). It only requires lemma-identity of the two Verbs in the two conjuncts & contrastiveness[2] of the remnants (non-elided constituents). For *lemma-identity* only the stems need to coincide. However, morphological properties such as Number or Person of a Verb may differ in the two conjuncts (e.g., in example (1), *žívët* and *žívut* are lemma-identical). The Gapping variant *Long-Distance Gapping (LDG)* recursively applies the general Gapping conditions top-down to corresponding Verb pairs in the structure, provided they are in the range of a so-called superclause[3] in both conjuncts. *Subgapping* works as LDG but a Nonfinite Verb structure happens to be not identical; this yields a Nonfinite clausal remnant in the second conjunct. *Stripping* is applied after any form of Gapping: during read-out of Gapping results it inspects whether there is no more than one non-Verb remnant; in that case read-out adds a language-specific stripping particle.

2. *Forward-Conjunction Reduction (FCR)* requires wordform-identity, i.e., in addition to lemma- and grammatical-function identity, identity of the morphological features is needed in the left-periphery of major clausal constituents, i.e., both clausal conjuncts should start with a wordform-identical sequence of FULL constituents.

3. *Backward-Conjunction Reduction (BCR)* is licensed by lemma-identity in the right-periphery, that is, both clausal conjuncts end with the same sequence of wordform and grammatical-function identical WORDS (e.g., in example (8), *o'clock* is such a sequence). Note that FCR and BCR are not complete mirror images because only BCR is allowed to disregarding major constituent boundaries.

4. *Subject Gap with Finite/Fronted Verb (SGF)* requires wordform-identical Subjects where the first conjunct starts with Verb/Modifier/Adjunct or where the first conjunct is a Conditional Subordinate Clause (Subject-Verb-Inversion) & FCR is applied if licensed.

## 3 Set-up of RUSSIAN-ELLEIPO

In order to use ELLEIPO for any new target language, the existing Java implementation of ELLEIPO has to be changed only minimally because the rule set works target-language independently. We added the Russian Conjunction and the Russian Stripping particle along with its position (leading or trailing) in the language-specific part of the existing Java code.

In order to test the accuracy in a new target language, an appropriate *test corpus* of ELLEIPO should contain structures that trigger ALL constraints in the rule set, i.e., lemma- and wordform-identity, contrastiveness, grammatical-function and word-order variation in the left- and right periphery. A blueprint of such a collection is ELLEIPO's test corpus of about 100 sentences for German and Dutch (see Harbusch & Kempen, 2006). All these sentences have been translated into Russian. In order to avoid biases, preserving the meaning was not essential but trying to keep the varying constraints active in the Russian syntactic trees, i.e., natural constructions in Russian have been set up (cf. clues for rule application/failure of all phenomena in Table 2; N.B. that several phenomena can occur in one test sentence). The large number of Gapping examples represents the great variety of word ordering to be ignored, contrastiveness to be obeyed, differing superclause-boundary violation-options (relevant for LDG and Subgapping), grammatical-function and lemma and wordform variation. The larger number of FCR tests compared to (the near mirror image) BCR results from more variation options for major constituents in the frontfield of a sentence compared to the limited word variation in the right periphery in BCR. For SGF, the range of options is also restricted.

Processing the Russian test corpus, ELLEIPO provides a condensed list of all reductions

[2] *Contrastiveness* constraints rule out elisions such as *I eat apples and you* ~~*eat*~~ *in the car*—which is comprehensible but not grammatical.

[3] A *superclause* is defined as a hierarchy of Finite or Nonfinite Clauses that—with the possible exception of the topmost clause—do not include a Subordinating Conjunction. In (3), the Subjects *Pëtr* and *Maša* each belong to a Main Clause headed by the Verb *dolžen* 'should' whereas *segodnâ* 'today' and *svoû mašinu/svoj velosiped* 'his car/her bike' belong to the Nonfinite Complement Clause headed by the Infinitive *myt´* 'wash'. Nevertheless, they form one superclause in each of the conjuncts. Example (7)—actually, a case of FCR where no superclause test is elicited—contains two superclauses in each conjunct, due to the Subordinating Conjunction *čto* 'that'.

| CCE Rule | Number of inclusions |
|---|---|
| GAPPING | 91 |
| STRIPPING | 17 |
| FCR | 72 |
| BCR | 25 |
| SGF | 17 |

Table 2: Phenomena in the Russian test corpus.

with/without subscripts—slightly more elaborate than indicated in Table 1. ELLEIPO adds unique numbers to each CCE token so that an elided constituent and its remnant directly correspond. For instance, ELLEIPO's output for example (3) spells out the sentence variant depicting Subgapping along with Backward-Conjunction Reduction (cf. subscript number #4 for BCR in **Segodnâ**$_{2\ 3}$ **dolžen**$_{1}$ Pëtr svoû mašinu **myt´**$_{4\text{-b}\,2}$ i **~~segodnâ~~**$_{2\text{-gg}\ \ 3\text{-f}}$ **~~dolžna~~**$_{1\text{-g}}$ Maša svoj velosiped **~~myt´~~**$_{2\text{-gg}}$—also notice subscript #3 licensing *segodnâ* for FCR). ELLEIPO's read-out component can spell out all possible combinations of elisions (with or without elaborate subscripts).

The complete lists of unreduced and reduced sentences form the text materials we presented to the participants in the two experiments that we carried out to calculate the accuracy of our CCE rules, specifically, the amount of *overgeneration* and *undergeneration* of the rule set in Russian. In experiment 1, we targeted overgeneration. We had native speakers judge the acceptability of the elliptical structures proposed by the CCE generator for the test corpus. In experiment 2, aiming to detect undergeneration, we tried to elicit yet undiscovered elision types for a standard corpus of unreduced test sentences (i.e., sentences without CCE). Obviously, the scope of the latter experiment is restricted, due to the limited number and variability of the sentences presented to the participants.

We used a rating scale specifying three levels of acceptability of a reduced sentence (*good, dubious, unacceptable*) in order to avoid overtaxing and exhausting the test subjects—in contrast to the very fine-grained method for grammaticality rating used by Keller (2000). In case of dubious acceptability, more details have been asked. Basically, a more fine-grained tendency for more/less acceptability as well as insights in misinterpretations have been traced. This type of comments was obtained in an *interview situation with a moderator*.[4] The moderator should be a linguist speaking the target language to bring up follow-up questions. Such digression does not spoil the study—compared to a standardized experiment as in Psychology. Another deviation from standardized testing (originally proposed for Usability (UX) Testing and verified with a meta-study on case studies by Nielsen (2012)) works very well here. Few test subjects—Jakob Nielsen suggests five in UX, although, user behavior varies more than in grammaticality rating—suffice to point out the majority of all problems.

In experiment 1, we let three native speakers of Russian evaluate ELLEIPO's output. The participants always saw unreduced sentences together with the reduced ones. This setup is necessary because it is known that, although some reductions are acceptable in themselves, they do not express the same meaning as its unreduced counterpart (cf. example (12) in next section). We counted a match as successful if at least one participant judged it acceptable.[5]

In experiment 2, we tried to identify *undergeneration* with the CCE rule set, i.e., judged acceptable by native speakers but failing to be produced by ELLEIPO. For this purpose we presented unreduced sentences only and let the participants freely produce any kind of reduction crossing their mind. In the list of answers we first identified Pronominalization, One-anaphora and other non-CCE forms of ellipsis as they do not count in our study (however, the participants cannot know this). Given the high amount of different linguistic constructions the participants produced, the motivation of the participants during the experiments can be judged to be high (so to speak playful in a positive sense). The experiment unveiled great similarities of Russian with Estonian and Hungarian which allow weaker word-ordering conditions for SGF and FCR (e.g., Ditransitive Verbs allow for non-peripheral elision of wordform and grammatical-function identical constituents).

[4]Further options are unmoderated tests conducted in an observation lab or (internet) questionnaires. The user studies for Estonian, Hungarian and Russian were conducted as face-to-face interviews to make test subjects try considerably harder (cf. Schulte-Mecklenbeck and Huber, 2003). Moreover, all kinds of misinterpretation can be discussed on the spot given that the moderator remains neutral in order to minimize unwelcome influence on the results of the test.

[5]This weak acceptability criterion was prompted by the fact that CCE acceptability ratings can give rise to wide inter-rater variability. In German, grammar books (and ELLEIPO) license BCR for constituents that are lemma-identical but not grammatical-function identical. However, many German native speakers rule out *Hilf* ***~~dem Mann~~]***$_{DAT}$ *und reanimier* ***[den Mann]***$_{ACC}$ 'Help and reanimate the man'.

## 4 Results

In experiment 1 (on overgeneration), 79 % of the sentences produced by ELLEIPO were judged acceptable. At a first glance this sounds meager. However, one should realize that if ELLEIPO wrongly applies a CCE rule, it does so for all sentences embodying the same trigger condition. Second experiment accomplished 97 % accuracy. The number of identified CCE tokens along with over- and undergeneration cases by type of CCE rules is shown in Figure 1. Note: the columns show absolute numbers, not percentages.

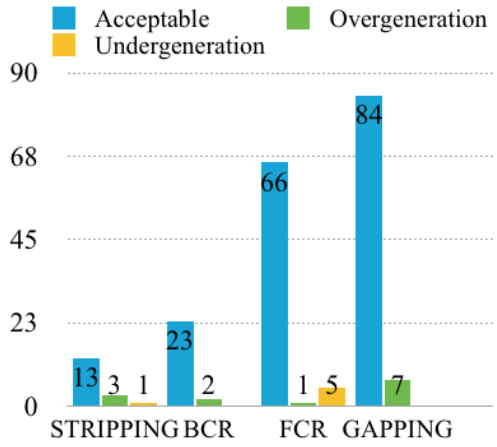


Figure 1: Numbers of cases revealing over- and undergeneration in the two experiments.

Comparison of our Russian data with those obtained in previous work for the two Finno-Ugric languages revealed interesting similarities. In Estonian and Hungarian, the left-periphery constraint is less strict compared to Dutch and German. In SGF, Estonian and Hungarian allow for more freedom in the frontfield whereas this is not possible in Dutch and German. In Russian, Arguments in the frontfield also license FCR (cf. example (10) with a Complement Clause in the frontfield).

(10) **Examen** **sdat´** hočet **on/student** i
The-exam$_{ACC}$ pass$_{INF}$ will he/student and

~~**examen**~~$_{f}$ ~~**sdat´**~~$_{f}$ možet ~~**on/student**~~$_{s}$ tože
the-exam$_{ACC}$ pass$_{INF}$ can he/student also
'The exam, he/the student wants to pass and he is also able to'

The typical superclause violation identified as acceptable in Hungarian for the subordinating conjunction *hogy* 'that' was not obtained in Russian (cf. example (11)). However, some informants indicate they might use it in colloquial spoken Russian.

(11) Maša **nadeetsâ** čto Pëtr **ujdet** i
Maša hope$_{3SG}$ that Pëtr leave$_{3SG}$ and

Katâ ~~**nadeetsâ**~~$_{g}$ that Jan ~~**ujdet**~~$_{gg}$
Katâ hope$_{3SG}$ that Jan leave$_{3SG}$
'Mary hopes that Peter leaves and Cathrine hopes that Jan leaves'

The acceptability judgments suggest two rule amendments that could help avoiding overgeneration and serious misunderstandings of the reduced sentences. In Long-Distance and Subgapping, exactly two constituents may remain in the second conjunct (see example (12) where the participants interpret the reduced sentence as 'you are in the bus'). Obviously, in Russian any inflected form of 'to be' is assumed to be left out for two remaining remnants instead of taking into account the Verbal constituents in the anterior conjunct. Notice, that we expected this reaction as this Russian-specific CCE phenomenon is discussed in the literature (see, e.g., Kazenin, 2006).

(12) * Â **[vižu Petra kotoryj spit]** v mašine i
I see$_{1SG}$ Pëtr$_{ACC}$ who sleep$_{3SG}$ in car$_{DAT}$ and

ty ~~**[videš´ Petra kotoryj spit]**~~$_{g}$ v avtobuse
you see$_{2SG}$ Pëtr$_{ACC}$ who sleep$_{3SG}$ in bus$_{DAT}$
'I see Pëtr that sleeps in the car and you in the bus'

Another remarkable difference that we could not trace in the linguistic literature is the fact that Russian speakers do not allow violation of *co-referentiality* of elided constituents (cf. example (13)). In this sentence, the constituents *svoj velosiped* 'his bike' cannot be elided by Gapping because the two constituents refer to two different referential objects (4 % of the reduced corpus sentences were rejected due to this fact).

(13) Maša slyšala, **čto** Ûrij **svoj velosiped pomyl**
Maša hear$_{PST.SG.F}$ that Ûrij his bike wash$_{PST.SG.M}$

i ~~**čto**~~$_{f}$ Pëtr **svoj velosiped pomyl**
and that Pëtr his bike wash$_{PST.SG.M}$
'Maša heard, that Ûrij and Pëtr washed their bikes'

## 5 Conclusions

We have identified remarkable similarity of the language-independent CCE rules in Russian compared to Dutch, Estonian, German, and Hungarian. Russian ellipsis reveals the highest similarity to Estonian if written text quality is considered.

As for future work, we plan to conduct a corpus study into Russian treebanks of spoken and written language in order to find additional subtle deviations that go beyond our studies (cf. Harbusch & Kempen, 2007).

## Acknowledgements

We thank Gerard Kempen for helpful discussions on earlier versions of the paper, and three reviewers for constructive criticism.

## References

John Beavers and Ivan A Sag. 2004. Coordinate ellipsis and apparent non-constituent coordination. In *Proceedings of the HPSG04 Conference*, pages 48–69.

Anette Frank. 2002. A (discourse) functional analysis of asymmetric coordination. In *Procs. of the LFG02 Conference*, pages 174–196.

Vera Gribanova. 2013. Verb-stranding verb phrase ellipsis and the structure of the Russian verbal complex. *Natural Language & Linguistic Theory*, 31(1):91–136.

James Griffiths and Anikó Lipták. 2014. Contrast and island sensitivity in clausal ellipsis. *Syntax*, 17(3):189–234.

Karin Harbusch and István Bátori. 2013. Clausal Coordinate Ellipsis (CCE) in Hungarian compared to cce in Dutch, German, and Estonian. In *Approaches to Hungarian: Volume 13: Papers from the 2011 Lund conference*, volume 13, page 45. John Benjamins Publishing.

Karin Harbusch and Gerard Kempen. 2006. Elleipo: A module that computes coordinative ellipsis for language generators that don't. In *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations*, pages 115–118. Association for Computational Linguistics.

Karin Harbusch and Gerard Kempen. 2009a. Clausal coordinate ellipsis and its varieties in spoken German: A study with the TüBa-D/S treebank of the Verbmobil corpus. In *Eighth International Workshop on Treebanks and Linguistic Theories*, pages 83–94. EDUCatt.

Karin Harbusch and Gerard Kempen. 2009b. Generating clausal coordinate ellipsis multilingually: A uniform approach based on postediting. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 138–145. Association for Computational Linguistics.

Karin Harbusch, Mare Koit, and Haldur Õim. 2009. A comparison of clausal coordinate ellipsis in Estonian and German. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics: Demonstrations Session*, pages 25–28. Association for Computational Linguistics.

Karin Harbusch. 2011. Incremental sentence production inhibits Clausal Coordinate Ellipsis: A treebank study into Dutch and German. *Schlangen, D. & Rieser, H. (Eds.). Dialogue and Discourse*, 2(1):313—-332.

Martin Haspelmath. 2007. Coordination. In *Language Typology and Linguistic Description*, volume 2, pages 1–51. Cambridge University Press, Cambridge, UK, 2 edition.

Kyle Johnson. 2009. Gapping is not (VP-) ellipsis. *Linguistic Inquiry*, 40(2):289–328.

Konstantin Kazenin. 2006. Polarity in Russian and the typology of predicate ellipsis. *Ms., Moscow State University*.

Frank Keller. 2000. *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. Ph.D. thesis, University of Edinburgh, UK.

Gerard Kempen. 2009. Clausal coordination and coordinative ellipsis in a model of the speaker. *Linguistics*, 47(3):653–696.

Gregory M Kobele. 2012. Eliding the derivation: a minimalist formalization of ellipsis. In *Proceedings of the HPSG 2012 conference*.

Jakob Nielsen. 2012. How many test users in a usability study? Online article with date: June 4, 2012, see: http://www.nngroup.com/articles/how-many-test-users.

Michael Schulte-Mecklenbeck and Oswald Huber. 2003. Information search in the laboratory and on the web: With or without an experimenter. *Behavior Research Methods, Instruments, & Computers*, 35(2):227–235.

Kerstin Schwabe and Ning Zhang. 2000. *Ellipsis in conjunction*, volume 418. Max Niemeyer Verlag.

James Shaw. 1998. Segregatory coordination and ellipsis in text generation. In *Proceedings of the 17th international conference on Computational linguistics-Volume 2*, pages 1220–1226. Association for Computational Linguistics.

John R Te Velde. 2006. *Deriving coordinate symmetries: A phase-based approach integrating Select, Merge, Copy and Match*, volume 89. John Benjamins Publishing.

Jeroen Van Craenenbroeck and Jason Merchant, 2013. *Ellipsis phenomena*, pages 701–745. Cambridge University Press, Cambridge, in Marcel den Dikken (ed.), The Cambridge handbook of generative syntax edition.