Agreement with Coordinated Subjects in the World's Languages

`http://www.multivaluation.de/database.php`

Himmelreich, Anke, Melissa Jeckel, Katharina Hartmann, Johannes Mursell, Nelly Kerezova, and Peter W. Smith
Goethe-Universität Frankfurt am Main
Contact: `himmelreich@lingua.uni-frankfurt.de`

1 Introduction

Basics:

- According to the literature, resolution of multiple values for agreement is subject to variation across languages, although there is no explanation for why languages vary in the way that they do.

- Our aim in this project is to collect data on the variation of agreement resolution and to find generalizations about the factors that influence the choice of the resolution strategy.

- To do so, we want to systematically investigate subject-verb agreement constructions where the subject is coordinated. Coordinations can be conjunctions (coordination with *and*, see (1)) or disjunctions (coordination with *or*, see (2)).

(1)   CONJUNCTION:
        [*The boy and the man*] **are/*is** running to the village.

(2)   DISJUNCTION:
        [*The boy or the man*] **are/is** running to the village.

Goals of the project:

- Create a database that contains reliable information about verb agreement with coordinated subjects.

- Find empirical generalizations about strategies for agreement with coordinated subjects in the world's languages.

Difficulties:

- Information about agreement with coordinated subjects is rarely found in language grammars.

- Coordinated subjects (especially disjunctions), where both conjuncts have mismatching ϕ-features are not frequently used by speakers. Therefore, speakers are often insecure which form to use. This may lead to unreliable results.

Claims:

- In this talk, we summarize and present first results from our database project.

- We show that word order is an important factor for the agreement strategy: Concretely, we show that closest conjunct agreement is more likely to occur when the coordinated subject follows the verb, while resolved agreement is more probable when the coordinated subject occurs first.

- We found no evidence that neither ϕ-features ($\#,\gamma,\pi$) or type of coordination (disjunction, conjunction) differ with respect to the agreement strategy.

- Finally, also language variation does not seem to play as big a role as expected.

2 Background

2.1 Constructions

- For the database to be developed, we are looking into structures with coordinated subjects.

- A coordinated subject consists of two parts, which can be coordinated either by *and* (3-a) or by *or* (3-b). The parts are called *disjuncts* in coordinations with *or* (aka *disjunctions)* and conjuncts in coordinations with *and* (aka *conjunctions*).

  (3) a. [*The boy$_{\phi_1}$ and the man$_{\phi_2}$*] **run**$_{\phi_?}$.
       b. [*The boy$_{\phi_1}$ or the man$_{\phi_2}$*] **run**$_{\phi_?}$.

- Assuming that word order might be a factor that determines the agreement strategy, the subject can either follow the finite verb of the clause (3) or precede it (4).

  (4) a. **Run**$_{\phi_?}$ [*the boy$_{\phi_1}$ and the man$_{\phi_2}$*].
       b. **Run**$_{\phi_?}$ [*the boy$_{\phi_1}$ or the man$_{\phi_2}$*].

- Furthermore, the conjuncts can either match in ϕ-features (3) or mismatch (5).

  (5) a. [*The boy$_{\phi_1}$ and the man$_{\phi_2}$*] **run**$_{\phi_?}$.
       b. [*The boy$_{\phi_1}$ or the women$_{\phi_2}$*] **run**$_{\phi_?}$.

- **Question:** We would like to find out which ϕ-features different languages mark on the verb.

2.2 Features

- For this database, we are looking at the three ϕ-features *gender*, *number*, and *person*, as well the feature *noun class*, which is used in many Bantu languages and Caucasian languages. (*Note:* Currently, we do not have data from languages with a noun class system in our database.)

- Ideally, each feature is looked at separately, since there seem to be different resolution strategies for different features in the same construction in the same language.

2.3 Agreement strategies

There are 8 agreement strategies which are logically possible
(*Note:* For lack of a better terminology, the term *conjunct* also means disjuncts):

1. **First conjunct agreement**: V agrees with the linearly first NP

    (6)   Aoun et al. (1994), (Munn 1999:650)
        a. **Mša**   [*ʕumar w   ʕali*].
            left.3SG Omar   and Ali
            'Omar and Ali left.'
        b. **mšitu**   [*ntuma w   ana*] məžmuin.
            left.2PL you.PL and I   together
            'You and I left together.'   *Moroccan Arabic*

2. **Last conjunct agreement**: V agrees with the linearly last conjunct

    (7)   [*Ich oder du*] **wirst**   krank.
           I   or   you get.2SG sick
           'I or you get sick.'   *German*

3. **Closest conjunct agreement**: V agrees with the linearly closest conjunct (the first conjunct under VS order, and the last conjunct under SV order)

    (8)   a. ?[*Yo o él*] **corre**.
               I   or he run.3SG
               'I or he runs.'
           b. ?**Corro**   [*yo o él*].
               run.1SG I   or he
               'I or he runs.'   *European Spanish*

4. **Ineffability throughout**: A coordinated subject with conflicting ϕ-features is not possible. Another construction is used to express this.

    (9)   Italu   **ilou**.
           we.1DU.INCL run.DU
           'You and I run.'/'I and you run.'/lit. 'We run.'

    (10)   a. [*Ita/ami*   **ilou**] me [*ila tee la*   **ilou**].
               we.INCL/we.EXCL run.PL and they also they run.PL
               'We and they run.'
           b. [*La* **ilou**] me [*ita/ami*   tee ita/ami   **ilou**].
               they run.PL and we.INCL/we.EXCL also we.INCL/we.EXCL run.PL
               'They and we run.'   *Mussau-Emira*

5. **Ineffability unless the forms are syncretic**: Mismatching ϕ-features are not allowed unless both forms are syncretic (an abstract example here illustrated with German data (= German′)).

(11)   a.   [*Ich oder mein Kollege*   ] ***habe/*hat**   gestern   einen Fehler   gemacht.
              I   or   my   colleague   have.1SG/3SG yesterday a   mistake   made
              'I or my colleague made a mistake yesterday.'
       b.   [*Ich oder mein Anwalt*] **soll**   morgen   dem Richter Bescheid sagen.
              I   or   my   lawyer   should.1SG~3SG tomorrow the   judge   notice   say
              'I or my lawyer should notify the judge tomorrow.'   *German′*

6. **Resolved agreement with hierarchy effects**: Mismatching gender and person features are resolved to the higher of the two values on a scale. The most common scales are shown in (12). Number is resolved to the value that reflects the sum of the numbers of the individual parts (12-c).

(12)   a.   GENDER:
              MASC > FEM
       b.   PERSON:
              1 > 2 > 3
       c.   NUMBER:
              SG + SG = DUAL/PL
              SG + NON-SG = PL
              NON-SG + NON-SG = PL   *non-sg ∈ {dual, pl}*

(13)   a.   [ʔ*anta wa-ʔana:*] **narkud^{ˤ}u**.
              You   and-I   run.1PL
              'You and I run.'
       b.   [ʔ*ana: wa-ʔanta*] **narkud^{ˤ}u**.
              I   and-you   run.1PL
              'I and you run.'   *Modern Standard Arabic*

7. **Default agreement**: Instead of computing new ϕ-features, a default agreement form is chosen.

(14)   **yarkud^{ˤ}u** [*alʔawla:du w-arrija:lu*].
        run.3SG   boy.PL   and-man.PL
        'The boys and the men run.'   *Modern Standard Arabic*

8. **Special agreement form**: Instead of computing new ϕ-features, a special agreement form is chosen that is not part of the regular agreement paradigm. (*Note*: It might be that this form is the default form. The special form can also be a form that does not mark agreement at all, as in Somali in (15).)

(15)   a.   [*Wiilka ama ninka*]   **ayaa orda**.
              boy.SG or   man.SG DEF   run.DEF
              'The boy or the man run.'
       b.   [*Wiilasha ama ninka*]   **ayaa orda**.
              boy.PL   or   man.SG DEF   run.DEF
              'The boys or the man run.'

c. [*Wiilka ama nimanka*] **ayaa orda**.
    boy.SG or man.PL DEF run.DEF
    'The boy or the men run.'   *Somali*

2.4 Factors for the agreement strategy

We hypothesize that the agreement strategy can depend on four factors:

- **Language** (there are language-specific patterns):
  Given the Chomsky/Borer Conjecture in (16), it follows that syntactic rules, such as agreement should apply in all languages similarly.

  (16) *Chomsky/Borer Conjecture* (as formulated in Obata et al. (2015:3))
        Syntactic parameters are restricted to variation in the morphological features of functional syntactic heads. (Borer (1984), Chomsky (1995))

  Literature on agreement with coordinations has usually only looked at one or a handful of languages at once (e.g. Aoun et al. (1994), Munn (1999), Bošković (2009), Bhatt and Walkow (2013), Marušič et al. (2015), Murphy and Puškar (2018), Nevins and Weisser (2018)). Comparing previous language-specific studies, it seems obvious that languages do show variation. Thus, language must be a factor that determines the pattern of agreement, at least partially. What needs to be seen is whether related languages behave alike and whether the cross-linguistic variation in agreement patterns can be reconciled with the Chomsky/Borer Conjecture.

- **Agreement feature** (i.e. gender, number, person, noun class):
  Besides looking at the features separately, it is also worth looking at feature interactions: Marušič et al. (2015) have shown that there is a connection between gender and number: Gender depends on number. They showed for Slovenian that closest conjunct agreement, first conjunct agreement, resolved agreement, and default agreement are all possible depending on the ϕ-features.

- **Word order** (SV, VS): Aoun et al. (1994) were among the first to show that there are word order differences: Various Arabic dialects exhibit resolved agreement under SV order and CCA under VS order.

- **Coordination type** (*and*, *or*): Marušič and Shen (2020) showed that, in Slovenian, both coordination types behave the same when it comes to the range of agreement strategies (CCA, FCA and Resolved agreement). However, they found that disjunctions show a greater tendency for CCA than conjunctions.
  (*Note*: There is no only very little work on the syntactic and semantic differences between disjunctions and conjunctions. For some work see Payne 1985; Haspelmath 1985; Schmitt 2013.)

The database should be designed in a way that each of these factors is stored as information. This way, we can check which of the four factors play a role typologically.

3 The database

3.1 Structure

The database consists of two parts:

1. A csv-file (= *comma separated value* file) that stores all the agreement strategies found in the languages under investigation. See (17) for an illustration.

   (17)

Language Name	Lang. Family	Features	Orders	Coord.	Agreement Strategy
Arabic (Modern Standard)	Afro-Asiatic	Gender	SV	and	Resolved Agreement
Arabic (Modern Standard)	Afro-Asiatic	Gender	VS	and	Closest Conjunct Agreement
Arabic (Modern Standard)	Afro-Asiatic	Gender	SV	or	Resolved Agreement
Arabic (Modern Standard)	Afro-Asiatic	Gender	VS	or	Closest Conjunct Agreement
Arabic (Modern Standard)	Afro-Asiatic	Number	SV	and	Resolved Agreement
Arabic (Modern Standard)	Afro-Asiatic	Number	VS	and	Default Agreement
Arabic (Modern Standard)	Afro-Asiatic	Number	SV	or	Resolved Agreement
Arabic (Modern Standard)	Afro-Asiatic	Number	VS	or	Default Agreement
Arabic (Modern Standard)	Afro-Asiatic	Person	SV	and	Resolved Agreement
Arabic (Modern Standard)	Afro-Asiatic	Person	SV	or	Resolved Agreement

2. A set of language files which contain more details about the respective language and the agreement strategies, including language examples. This information is stored on a separate language site. The links to these language sites from the main site of the database are generated automatically.

3.2 How to use the database

Accessing the database and retrieving information

1. Go to `http://www.multivaluation.de/database.php`

2. Click on the button "**Load Data**". You will receive a summary of your search result as well as a table which lists all the results found. The summary gives you the number of items found as well as the relative frequencies for the languages, language families, features, word orders, coordination types, and agreement strategies found. This way you can check for correlations between different factors.

3. **Searching for specific information**:
   In order to look for a specific language, agreement feature, word order, coordination type, or agreeement strategy, you can use the filters above the button "**Load Data**". If you do not use any filters, you will get a list of all the information currently stored in the database. You can use multiple filters at the same time.

4. **Find details and examples for each language**:
   Click on the language names in the list that you are interested in. Alternatively, you can use the sidebar menu. On the language site you will find a pdf which you can read and download. If you want to use the information, please follow the note on how to cite our work following the remarks in the beginning of the pdf.

5. **Starting a new search after visiting a language site**:
   Use the sidebar menu and click on "**Database**".

Downloading information

- **Download the summary of your search**:
  Click on the link "**Download summary**" right below the summary. The summary will be converted into a plain txt-file.

- **Download the language table**:
  Click on the link "**Download table**" right below the summary (and above the table). The table will be converted to a csv-file, which you can open in any spreadsheet program (e.g. *Microsoft Excel*, *LibreOffice*, *Google Tables*; select a comma as the delimiter).

- **Download language details and examples**:
  Click on the language you are interested in either in the table or in the sidebar. On the language site you will find a pdf which you can read and download.

Example 1:
You want to find out what agreement strategies you get in Turkish for person agreement.

1. Search for "Turkish" in the filter "Language".

2. Select "Person" in the filter "Agreement Feature".

3. Click on "Load Data".

4. Click on "Turkish" in the list.

Example 2:
You want to find out which languages show closest conjunct agreement for number.

1. Select "Person" in the filter "Agreement Feature".

2. Select "Closest Conjunct Agreement" in the filter "Agreement Strategy".

3. Click on "Load Data".

4. Click on links in the list for more details.

3.3 Methods of data collection

Main problem:
The relevant constructions are rarely found in language grammars.
Solution:

- We started eliciting data in various ways via online surveys and standard elicitation sessions.

- Currently, we use a new questionnaire containing translation and rating tasks:
  `http://multivaluation.de/questionnaire.php`

- The questionnaire consists of two parts:
  1. In the first step, speakers translate 18 simple sentences from English into their language. The goal is to elicit the complete agreement paradigm as well as all word forms necessary to construct coordinated subjects. We also try to find out whether languages allow VS orders. Based on the results, we semi-automatically generate sentences in the respective target language and generate part two of the questionnaire

2. In the second step, speakers are asked to rate sentences from their language on a scale from 1-5 (5 being the best possible rating):

   – The sentences vary regarding the coordinated noun phrases, the coordination type (*and* or *or*), verb agreement, and word order (VS or SV).

   – The sentences are presented in blocks, each block has the same sentence with different agreement options.

   – The blocks are presented in random order, which reduces the problem of the speakers seeing a sequence of minimal pairs.

   – The rating results are automatically saved once a block is finished. That means that we get results, even if the speaker does not finish the questionnaire. The rating results are then analyzed manually and the ratings are mapped to categories of grammaticality. The language files including the transliteration and glossing must also be generated manually.

4 Preliminary Results

Notes:

- The database currently has 150 entries.

- In this section, the percentages show how many percent the results for a language, a language family, a feature, a coordination type, a word order, and an agreement strategy constitute from the results found in total.

4.1 Overview and Disclaimers

- We can see that, currently, the database is not typologically balanced. The majority of entries are from Indo-European languages and Afro-Asiatic (specifically Semitic) languages.

(18)   **7 language families:**

Indo-European	59.33%	Afro-Asiatic	26.00%
Turkic	5.33%	Isolate	2.67%
Uralic	2.67%	Panoan	2.00%
Austronesian	1.33%		

- Since gender agreement is far less common than number and person agreement, it is expected to show up less.

(19)   **3 agreement features:**

Number	46.00%
Person	44.67%
Gender	8.67%

- Since the majority of the data are from the first survey where we did not check for VS order, SV order is dominant right now. We hope to overcome this problem with the new questionnaire.

(20)   **2 word orders:**

SV	72.00%
VS	27.33%

- As we have checked for both coordination types in all languages, there are nearly as much data on conjunctions as there are on disjunctions.

(21)   **2 coordination types:**

and	51.33%
or	48.00%

- So far, we can see that resolved agreement is the most common strategy for agreement. (This generalization also holds if we only look at Indo-European or only at Afro-Asiatic languages.)

(22)   **7 agreement strategies:**

Resolved Agreement	73.33%
Closest Conjunct Agreement	9.33%
Resolved Agreement/Closest Conjunct Agreement	7.33%
Resolved Agreement/Ineffability	4.67%
Resolved Agreement/Default Agreement	2.00%
Default Agreement	1.33%
Ineffability	1.33%

4.2 Factors for the agreement strategy

Note: In this section, we take a closer look at the two agreement strategies resolved agreement and closest conjunct agreement. We check whether the four hypothesized factors in section 2.4 play a role for the choice of the agreement strategy.

4.2.1 Effects of language

- There is some cross-linguistic variation: Languages differ in which strategies they show (more).

- There are intra-linguistic variation: One language can show multiple strategies.

- However, according to a Fisher's exact test, the factor language did not reach significance ($p=0.15$).

(23)   *Comparison of four languages*

	North Levantine Arabic	**Turkish**	**Hebrew**	**BCMS**
Resolved Agreement	75.00%	50.00%	83.33%	100.00%
Closest Conjunct Agreement	25.00%	12.50%	0.00%	0.00%

4.2.2 Effects of agreement features

- To test for agreement feature as a factor we only looked at SV word order.

- A Fisher's exact test revealed that agreement feature under SV order is not a significant factor ($p=0.78$)

(24)

	number	**gender**	**person**
Resolved Agreement	94.00%	85.71%	78.43%
Closest Conjunct Agreement	2.00%	0.00%	1.96%

4.2.3 Effects of word order

- It is noticeable that resolved agreement has a much higher frequency under SV-order than under VS-order.

- Closest conjunct agreement is much more prevalent under VS-order.

- A Fisher's exact test revealed that word order is a significant factor for the choice of the agreement strategy ($p<0.001$).

(25)

	SV	VS
Resolved Agreement	86.11%	41.46%
Closest Conjunct Agreement	1.85%	29.27%

4.2.4 Effects of coordination type

- To disentangle the factors word order and coordination type, we only looked at SV order.

- Overall, we can see that resolved agreement is the dominant pattern for both coordination types.

- Nevertheless, disjunction has a slightly bigger tendency for closest conjunct agreement. However, this difference did not reach statistical significance (Fisher's exact test, $p=0.10$).

(26)

	and	or
Resolved Agreement	81.82%	65.28%
Closest Conjunct Agreement	5.19%	13.89%

5 Summary and Outlook

- In this talk we have presented first results on our research about agreement with coordinated subjects in the world's languages.

- We determined four factors that possibly influence the agreement strategy according to the literature:
    - Language
    - Agreement feature
    - Word order
    - Coordination type

- Of these four factors, we can currently confirm that word order is a significant factor: Under VS order, closest conjunct agreement is more probable than under SV order.

- More data (testing more factor combinations and coming from a more balanced language sample) is needed to see whether or not the other factors play a role as well.

Accounting for word order effects?:

- At this point, it is unclear how theories of agreement can account for the word order effect.

- In many constructions the final word order (VS vs. SV) is fixed after agreement has applied.

- Additionally, one has to keep in mind that there often is some sort of optionality between agreement strategies for speakers.

- In fact, during elicitation, some speakers experienced some insecurity regarding their judgments.

Outlook:

- In the last years, the interest in agreement with coordinations has increased.

- Researchers can access our database to
    1. Find data of a specific language
    2. Find generalizations for language families and typological tendencies.

- Hopefully, the data can help to understand how coordinations resolve ϕ-features and eventually shed some more light on the specifics of agreement.

- There also are other database projects that collect language data to help find universal tendencies and cross-linguistic differences: see WALS (Dryer and Haspelmath 2013), TerraLing (Collins et al. 2013).

- As for coordinations specifically, Haslinger et al. (2019) have developed a database on TerraLing.

- We hope that we can connect our data with other projects in the future.

References

Aoun, Joseph, Elabbas Benmamoun and Dominique Sportiche (1994): 'Agreement, Word Order, and Conjunction in Some Varieties of Arabic', *Linguistic Inquiry* **25**, 195–220.

Bhatt, Rajesh and Martin Walkow (2013): 'Locating agreement in grammar: an argument from agreement in conjunctions', *Natural Language and Linguistic Theory* **31**, 951–1013.

Borer, Hagit (1984): *Parametric syntax: case studies in Semitic and Romance languages*. Foris, Dordrecht.

Bošković, Željko (2009): 'Unifying first and last conjunct agreement', *Natural Language and Linguistic Theory* **27**, 455–496.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Massachusetts.

Collins, Chris, Hilda Koopman, Richard Kayne, Dennis Shasha, Ross Kaffenberger, Alex Lobascio, Marco Liberati, Oleg Grishin, Lingliang Zhang, Hannan Butt and Andrea Olivieri (2013): TerraLing. New York: Available online at `https://terraling.com/`, Accessed on 2021-01-23.

Dryer, Matthew S. and Martin Haspelmath (2013): The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at `http://wals.info`, Accessed on 2021-01-23.).

Haslinger, Nina, Valentin Panzirsch, Eva Rosina, Magdalena Roszkowski, Viola Schmitt and Valerie Wurm (2019): A plural analysis of distributive conjunctions: Evidence from two cross-linguistic asymmetries. Ms., University of Vienna.

Haspelmath, Martin (1985): Coordination. *In:* T. Shopen, ed., *Language Typology and Syntactic Description*. Cambridge University Press, Cambridge, pp. 1–51.

Marušič, Franc, Andrew Nevins and Bill Badecker (2015): 'The grammars of conjunction agreement in Slovenian', *Syntax* **18**(1), 39–77.

Marušič, Franc and Zheng Shen (2020): Gender Agreement with Exclusive Disjunction in Slovenian. Ms., University of Nova Gorica/National University of Singapore.

Munn, Alan (1999): 'First Conjunct Agreement: Against a Clausal Analysis', *Linguistic Inquiry* **30**(4), 643–668.

Murphy, Andrew and Zorica Puškar (2018): 'Closest conjunct agreement is an illusion', *Natural Language and Linguistic Theory* **36**, 1207–1261.

Nevins, Andrew and Philipp Weisser (2018): 'Closest conjunct agreement', *Annual Reviews of Linguistics* pp. 1–25.

Obata, Miki, Samuel Epstein and Marlyse Baptista (2015): 'Can crosslinguistically variant grammars be formally identical? Third factor underspecification and the possible elimination of parameters of UG', *Lingua* **156**, 1–16.

Payne, John (1985): Complex phrases and complex sentences. *In:* T. Shopen, ed., *Language Typology and Syntactic Description*. Cambridge University Press, Cambridge.

Schmitt, Viola (2013): More pluralities. PhD thesis, Universtität Wien.