

C08 (NEW) FANSELOW: CONSEQUENCES OF HEAD ARGUMENT ORDER FOR SYNTAX

1.1. General information about Project C08

1.1.1. Project title: Consequences of Head Argument Order for Syntax (CHAOS)

1.1.2. Research area: 104 Linguistics

1.1.3. Project leader

Fanselow, Gisbert, Prof. Dr., June 20th 1959, German, University of Potsdam, Faculty of Human Sciences, Cognitive Sciences, Dept. Linguistics, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Phone: ++49-331-9772446; E-Mail: fanselow@uni-potsdam.de

Is the employment of the project leader(s) at the institution(s) indicated contractually secured for the duration of the proposed funding period? Yes

Do any of the above mentioned persons hold fixed-term positions? No

Funding for the position(s) of the project leader(s) at the institution(s) indicated is covered by core support (state funds or similar): Yes

1.1.4. Legal issues

This project includes

1.	research on human subjects or human material.	no
2.	clinical trials.	no
3.	experiments involving vertebrates.	no
4.	experiments involving recombinant DNA.	no
5.	research involving human embryonic stem cells.	no
6.	research concerning the Convention on Biological Diversity.	no
7.	investigations involving dual use research of concern.	no

1.2. Summary

Research on word order typology has shown strong correlations among the serialization of different heads relative to their arguments. More recently, the thesis has been formulated that other syntactic differences between languages, which go beyond serialization, can also be predicted or derived from the linear order of verb and object (or, rather, the structural factors determining linearization; cf. especially Haider, 2015, 2020).

The project *Consequences of Head Argument Order for Syntax (CHAOS)* proposed here is intended to investigate in a cross-linguistic comparison whether such predictions are correct, but also to identify those factors that can be simultaneously held responsible for the differences in the serialization of the object and the verb and for the corresponding syntactic properties going beyond linearization. For this purpose, different language samples will be examined: samples representative of the linguistic areas, languages with ‘striking’ serialization patterns (e.g., OVX), and language families or languages with considerable internal variability.

Thus, the project tries to test hypothesis-based predictions about the limits of linguistic variability (which properties are, for example, inadmissible or unlikely in OV-languages?) in syntactic areas for which such a connection has not been investigated thoroughly so far. It also addresses the plausibility of syntactic models that assume that different structural types exist within the group of OV- or VO-languages, and contributes to the topic “surface stability” of the proposed CRC in this respect.

1.3. Research rationale

1.3.1. Current state of research and preliminary work

Word order typology is one of the most visible and successful branches of comparative grammar research. In its research tradition, following the seminal work of Greenberg (1963), important correlations among the orders of various types of heads and their dependents have been uncovered, such as the strong tendency for a harmonic serialization of verbs and adpositions relative to their complements: object-verb (OV-) languages are predominantly postpositional, while the large majority of verb-object (VO-) languages have prepositions. The source of such correlations has mostly been traced back to the diachrony of languages, language learning, and to the functional-adaptive shaping of the form of languages (cf. Braquet &

Culbertson, 2017; Dryer, 2019; Hahn, Jurafsky, & Futrell, 2020, Haspelmath, 2019; Hawkins, 2004, 2014, among others).

Research in word order typology has so far focused on correlations among the ordering laws for heads and their respective dependents, so that relatively few results have been found for the impact of head-dependent order on further grammatical domains. One prominent example of the latter type of effect is the correlation between the order of the verb and the object, and the type of the marking of core arguments in a clause; in particular, there is a predominance of the SVO-⁹ (subject-verb-object) word order type among zero marking languages in which grammatical functions of arguments are not expressed by any morphological marking (Mallinson & Blake, 1981; Siewierska & Bakker, 1996; Sinnemäki, 2010). Ergative alignment does not appear to go well together with SVO-order (Roberts, 2019, p. 482ff.). The realization of indirect objects is influenced by verb-object order (Dryer with Gensler, 2013). Languages with VO-order are also more likely to move *wh*-phrases (*who*, *which proposal*) to the left periphery of the clause than subject-object-verb (SOV-) languages (Dryer, 2007; Hawkins, 2014), a fact that may stem from the frequent presence of a preverbal position for focus in the SOV-languages (e.g., Cypionka, 2007), into which *wh*-phrases fit in semantic/pragmatic terms. There is also an influence of the order of the verb and its object on the distribution of serial verb constructions, clause chaining (Schiller, 1990), and converbs (Haspelmath, 1995). The Final-over-Final Constraint (Biberauer, Holmberg, & Roberts, 2014) regulating the ordering possibilities of an embedding and an embedded head relative to their complements is a further important result.

The question the proposed research project CHAOS wants to address is whether there are *systematic overall* differences in the grammar of VO- and OV-languages that go beyond head-dependent ordering and are intrinsically related to or caused by the structures linked to the different linearizations of the verb and the object. There are at least three reasons for why such systematic correlates of the order of the verb and the object should exist.

First, some syntactic processes and relations have turned out to be inherently directional. Due to this directionality, their effects may be different in VO- and OV-languages. For the purposes of scope marking and clausal typing, *wh*-phrases that do not stay in their canonical position always move to the left periphery and never to the right periphery of the clause, at least in spoken languages (see Petronio & Lillo-Martin, 1997; Geraci & Cecchetto, 2013¹⁰). Linear order plays a role in quantificational binding (Jackendoff, 1990; Bruening, 2014, among others), selection and agreement in coordination (Bruening & Al Khalaf, 2020; Nevins & Weisser, 2019), and the licensing of negative polarity elements (Neeleman & van de Koot, 2020).

The inherent directionality of (A-) scrambling (the operation that is often used to describe, for example, free constituent order in the German midfield) has been made responsible for the possible correlation of preverbal objects and free constituent order (Reuland & Kosmeijer, 1988; Corver & van Riemsdijk, 1997). The authors assume that A-scrambling is inherently leftward bound. Further assumptions (such as that A-scrambling must not move phrases across the selecting head, Corver & Riemsdijk, 1997) render A-scrambling inapplicable in VO-languages, i.e., these languages are predicted to come with fixed constituent order. See Saito & Fukui (1998) for a related proposal predicting the distribution of scrambling and Heavy NP Shift in VO- and OV-languages. In other words, at least some grammatical processes may be directly sensitive to the ordering of the verb and its complement.

Second, the functional adaptive forces shaping grammars may yield different outcomes in VO- and OV-languages. Hawkins (2008), for instance, argues that the different circumstances in the identification of the immediate constituents of VP imply that the distinction between objects and adjuncts is less strict in OV-languages than in VO-languages. This leads to the expectation that this greater similarity of objects and adjuncts might extend to grammatical-function-related constraints on the extraction of phrases – one could argue that island effects should be “crisper” in VO-languages. Furthermore, one should also see major grammatical differences resulting from the fact that syntax processing in OV-languages may be less structure-dependent than in VO-languages.

Third, from a generative syntax perspective, the basic laws of structure building in (universal) grammar may be such that they imply substantial differences in the syntactic behaviour of OV- and VO-languages. This is particularly the case if, as assumed by Kayne (1994), Fukui and Takano (1998), and Haider (2013, 2014, 2015), the clausal structures related to one of the two orders are syntactically “more basic” (VO for Kayne and OV for Haider), so that the other serialization induces a more complex derivation for clauses and their structural representation. These differences in the derivation of VO- and OV-clauses can, for

⁹ In the tradition initiated by Dixon (e.g., Dixon, 1994), one would speak of AVO- rather than SVO-order.

¹⁰ They may optionally end up in the right periphery if the language places foci there; cf., e.g., Cain (2000) for Dhivehi.

example, have an impact on the degree of distinctness of subjects, objects, and adjuncts. If one starts out with a basic VO-order (Kayne, 1994), and derives OV-order by leftward movement of the phrases surfacing preverbally (e.g., Zwart, 1993), then all preverbal XPs end up in derived adjunct or specifier positions, which suggests a high degree of similarity for processes that refer to structural properties of a phrase. Again, the differences between subjects, objects, and adjuncts should be less pronounced in OV-languages than in their VO-counterparts. In Haider's proposal, (complex) verb phrases in VO languages involve an elaborate system of leftward head movements which yield structures in which objects occupy higher positions than they do in OV-languages (in line with the independently motivated proposals of Larson, 1988 and Johnson, 1991).

The degree of the syntactic difference between OV- and VO-languages to be expected is, of course, a function of the mechanisms proposed for mapping VO to OV or vice versa. Mathew (2014), for instance, derives Malayalam OV-order from a VO-base by simply moving the verb to the right, and Fukui and Takano (1998) assume that VO-order arises by leftward verb movement from a verb-final base. In such models, the hierarchical relations among the dependents of the verb remain the same in OV- and VO-languages – but note that objects fail to form a constituent with the verb in OV-languages in Mathew's proposal! It is, however, fair to say that the empirical and theoretical foundation of the models assuming that differences in verb-object order are related to pronounced structural differences, viz., Kayne (1994) and Haider (2013), is much stronger. Furthermore, one also need not declare one of the two orders as the more basic one if one wants to capture at least some of the presumed correlates of the order of verb and object. Neeleman (2014) and Belk and Neeleman (2017), for example, derive the prediction that we find obligatory verb-object adjacency in VO- and not in OV-languages from a law of case theory in interaction with the basic order in a language. Summing up, we conclude that much of the pertinent generative literature supports the expectation of substantial grammatical differences between OV- and VO-languages.

It should be noted that there is some overlap between the system of correlates of the OV-VO distinction and the configurationality parameter formulated by Hale (1983) some decades ago. Hale argued for a principled division between languages that come with a rich hierarchical sentence structure (such as English) and those with a rather flat sentence structure (such as Warlpiri). If structural distinctions are less pronounced in OV-languages for syntactic or for processing reasons (as briefly discussed above), they will be more “non-configurational” than their VO-counterparts.

The idea that the order of the verb and the object shapes the grammar to a large extent has been most explicitly and most visibly formulated by Haider (2013, 2014, 2015). There are at least three types of challenges for this view.

First, the empirical basis for such a far-reaching claim is still somewhat small, and there are known counterexamples. Haider (2013) based his proposal on an in-depth comparison of Germanic languages. It is neither obvious that his results are not influenced by an areal or genetic bias, nor can we be sure that the grammatical characteristics of OV-languages are not strongly influenced by other factors usually accompanying OV-order such as morphological richness. In other words, a much broader empirical basis on which to assess the impact of verb-object order is called for in which more linguistic families and linguistic macroareas figure as well. That said, Haider's method of highlighting the comparison of languages belonging to the same family is quite fruitful because it allows us to clearly carve out what changes are mandatory or highly likely when languages change their word order type.

The identification of the factor that really causes some syntactic property constitutes the second challenge. Consider, for instance, the connection between free constituent order and OV-order. Corver and Riemsdijk (1997) already admitted that Slavic languages do not show such a correlation. Haider and Szucsich (in press) respond to this difficulty by denying that Slavic languages belong to the SVO-type and classify them as languages with undetermined verb-complement order.

There is, however, a different perspective on the Slavic languages that does not put them on a par with English without giving up their classification as VO-languages. As Dryer (1997, 2013a) has argued, complete clause patterns such as “SVO” or “SOV” do not really figure in comparative explanations, rather, the explanatory force resides in two binary parameters “VO vs OV” and “SV vs VS”. As Dryer 1997, 2013a) continues to argue, the Slavic language Polish may indeed be classified as VS (because of the ordering frequency in intransitive clauses) rather than SV. Consequently, the word order differences between Polish and English might be caused by the SV vs VS difference, which is in turn related to the “height” of the subject (is it a constituent of VP, or does it have to be placed into a higher functional projection?). If OV-languages place their subjects into low positions, the height of the subject rather than OV-order would be the key to an understanding of free constituent order, if one assumes that it is movement to high functional positions that makes word order inflexible (Fanselow, 2001). Other potential candidates for explaining the Slavic-English/Scandinavian contrast are the richness of morphology or verb-object adjacency as an

independent property. The sample of languages by which one tries to assess the extent of the influence of head-complement order on other aspects of syntax must therefore be composed in such a way that it allows the identification of the relative contribution of factors such as those just discussed.

The third challenge is the most difficult one to deal with, at least in practical terms. Syntacticians tend to think about grammatical distinctions between languages in terms of binary choices: a language either has OV-order or it does not, the auxiliary precedes or follows the main verb, etc. But for some (if not most) of the criteria proposed by, for example, Haider (2015), the syntactic differences are not necessarily categorical: with respect to superiority violations (such as *was hat wer gesehen* lit.: “what has who seen”) or extractions from subjects (*wen hat einladen zu müssen dich verwirrt*), the difference between VO-English and OV-German is gradient in nature if one compares acceptability ratings (Featherston, 2005; Häussler et al., 2015 for superiority, and Jurka, 2010 for extractions from subjects), and pertinent sentences can even be found in corpora of English, though much less frequently than in German. Gradience may even affect the presence of subject expletives that may be obligatory (English), optional (Dutch), or prohibited (German) and their occurrence may be influenced by many factors, among them the nature of the predicate (transitive vs intransitive vs unaccusative). Finally, while English certainly lacks VS sentences, other languages allow them in different frequencies and under different pragmatic contexts. The properties relevant for possible correlation of the order and V and O are therefore not necessarily subject to a categorical parametrization. Unfortunately, for most languages, the collection of such gradient data is beyond the scope of a project involving cross-family and cross-area comparison.

The proposed research project CHAOS will mainly address the first two challenges, and confine itself to a minimum for the third challenge for practical reasons: investigating this challenge properly would possibly generate work for a whole Collaborative Research Centre.

The applicant and his team have been concerned with the research question of this proposed project for quite some time. Fanselow (1987) is an attempt to explain away the syntactic differences between English and German. The argumentation in this book often involves the gradient nature of the English-German contrasts. In later work, Fanselow has stressed the importance of low/VP internal subjects for German (e.g., Fanselow, 1991) and argued for a movement analysis of fixed rather than free constituent order (Fanselow, 2001). Cross-linguistic comparisons involving verb position effects stemming from former research projects are, for instance, reported in Cypionka (2007) and Skopeteas and Fanselow (2010a,b) with respect to the placement of focus.

Andreas Schmidt addresses the major question of the first challenge from above by in-depth analyses of Uralic languages (e.g., Schmidt, 2016), largely following the lead of Haider (2013). He has collected a considerable amount of data related to the project topic for Uralic languages, and collaborates intensively with the group of Balasz Suranyi in Budapest (NKFIH research project #135958 on the postverbal domain in OV-languages). In many respects, relevant syntactic differences in Uralic can be correlated with the OV-VO distinction. At the same time, there are untypically rigid OV-languages (South Sami) and untypically flexible VO-languages (Komi-Syrjänic), which shows that the OV-VO distinction must at least be modulated by further factors in an account of variability in Uralic. Fanselow, Traoré, and Féry (submitted) examine the OVX-language Tagbana for correlates of OV-syntax and find that the language’s profile resembles VO-languages much more than OV-languages. In work in progress by Fanselow, Nguyen Thi Suu, and Thuan Tran the isolating VO-languages Vietnamese and Taoih are compared. They find a surprising number of “OV-properties” in Taoih.

Fanselow (2020) discusses the research issues arising in the proposed project from a theoretical point of view, and comes to the conclusion that the height of the subject is a central factor in the correlations of properties such as those defined below. He concludes that an understanding of the connection between verb-object adjacency and high subject positions (should it be confirmed by comparative research) is a major challenge for theoretical modelling. Häussler, Eythorsson, Fanselow, Šimik, and Vicente† have compared the behaviour of seven European languages with respect to superiority violations, and found gradient differences in the acceptability of crossing movement that can, apparently, be best related to different modes of identifying the grammatical functions of core arguments (Häussler et al., under revision).

The present research proposal would not have been possible without intense discussions with Balthasar Bickel, Walter Bisang, and Hubert Haider, which also led to a jointly organized workshop on the topic at the 2015 DGfS annual conference in Leipzig.

References (other than own “Project-related publications”, see below)

- Aikhenvald, A. (2003). *A grammar of Tariana, from North West Amazonia*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/S0047404503321013>
- Aoun, J & Li, Y.-H. (1993). *Syntax of Scope*. Cambridge, MA: MIT-Press.

- Baker, M. (2010). Formal Generative Typology. In B. Heine & H. Narrog (eds.) *The Oxford Handbook of Linguistic Analysis* (pp. 285-312). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199544004.013.0012>
- Belk, Z. & Neeleman, A. (2017). AP-adjacency as a Precedence Constraint. *Linguistic Inquiry*, 48, 1-45. https://doi.org/10.1162/LING_x_00262
- Bell, D. (2019). Chinese possesses Japanese style scrambling: the case of Xining Mandarin. *Journal of East Asian Linguistics*, 28, 143-178. <https://doi.org/10.1007/s10831-019-09192-6>
- Biberauer, T., Holmberg, A. & Roberts, I. (2014). A Syntactic Universal and Its Consequences. *Linguistic Inquiry*, 45, 169-225. https://doi.org/10.1162/LING_a_00153
- Bickel, B. (2010). Capturing particulars and universals in clause linkage: a multivariate analysis. In I. Brill (ed.), *Clause-hierarchy and clause-linking: the syntax and pragmatics interface* (pp. 51-102). Amsterdam: Benjamins. <https://doi.org/10.1075/slcs.121.03bic>
- Bickel, B. (2011). Multivariate typology and field linguistics: a case study on detransitivization in Kiranti (Sino-Tibetan). In P. Austin, O. Bond, D. Nathan, & L. Marten (eds.), *Proceedings of Conference on Language Documentation and Linguistic Theory 3* (pp. 3-13). London: SOAS. <https://doi.org/10.5167/uzh-57331>
- Bickel, B. (2013). Distributional biases in language families. In B. Bickel, L. Grenoble, D. Peterson, & A. Timberlake (eds.), *Language typology and historical contingency* (pp. 415-444). Amsterdam: Benjamins. <https://doi.org/10.1075/tsl.104.19bic>
- Bickel, B. (2017). Areas and Universals. In R. Hickey (ed.), *The Cambridge Handbook of Areal Linguistics* (pp.40-55). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781107279872.004>
- Bickel, B., Witzlack-Makarevich, A., Choudhary, K., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2015). The neurophysiology of language processing shapes the evolution of grammar: evidence from case marking. *PLOS ONE*, 10(8), e0132819. <https://doi.org/10.1371/journal.pone.0132819>
- Bickel, B., Nichols, J., Zakharko, T., Witzlack-Makarevich, A., Hildebrandt, K., Rießler, M., Bierkandt, L., Zúñiga, F., & Lowe, J. (2017). *The AUTOTYP typological databases*. Version 0.1.1. Retrieved from <https://github.com/autotyp/autotyp-data/tree/0.1.1>
- Bobaljik, J., & Wurmbrand, S. (2012). Word order and scope: Transparent interfaces and the $\frac{3}{4}$ signature. *Linguistic Inquiry*, 43, 371-421. https://doi.org/10.1162/LING_a_00094
- Bošković, Ž. (2009). More on the no-DP analysis of article-less languages. *Studia Linguistica*, 63, 187-203. <https://doi.org/10.1111/j.1467-9582.2009.01158.x>
- Braquet, G., & Culbertson, J. (2017). Harmony in a Non-Harmonic Language: Word Order Learning in French Children. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 156-161). Austin, TX: Cognitive Science Society. Retrieved from <https://cogsci.mindmodeling.org/2017/papers/0042/paper0042.pdf>
- Bruening, B. (2014). Precede-and-command revisited. *Language*, 90, 342-388. <https://doi.org/10.1353/lan.2014.0037>
- Bruening, B., & Al Khalaf, E. (2020). Category Mismatches in Coordination Revisited. *Linguistic Inquiry*, 51, 1-36. https://doi.org/10.1162/ling_a_00336
- Cain, B. (2000). Dhivehi (Maldivian): A synchronic and diachronic study. Doctoral dissertation, Cornell. Retrieved from <http://hdl.handle.net/11858/00-001M-0000-0012-973D-E>
- Chamoreau, C. (2017). Purepecha: A polysynthetic but predominantly dependent-marking language. In M. Fortescue, M. Mithun, & N. Evans (eds), *Handbook of Polysynthesis* (pp. 667-695). Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199683208.013.38>
- Constable, P. G. (1989). Basic clause structure in Veracruz Huastec. MA Thesis, University of North Dakota. Retrieved from <https://commons.und.edu/theses/829/>
- Corver, N., & van Riemsdijk, H. (1997). The Position of the Head and the Domain of Scrambling. In B. Palek (ed.), *Typology: Prototypes, Item Orderings and Universals* (pp. 59-90). Prague: Charles University Press.
- Creissels, D. (2005). A typology of subject marker and object marker systems in African languages. In E. F. K. Voeltz (ed.), *Studies in African Linguistic Typology* (pp. 445-459). Amsterdam: Benjamins.
- Cristofaro, S. (2013). 'When' Clauses. In M. Dryer & M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/126>
- Czypionka, A. (2007). Word Order and Focus Position in the World's Languages. *Linguistische Berichte*, 212, 439-454. Link to preprint.
- Dixon, R. (1994). *Ergativity*. Cambridge: Cambridge University Press.
- Dryer, M. (1992). The Greenbergian Word Order Correlations. *Language*, 68, 81-138. <https://doi.org/10.2307/416370>
- Dryer, M. (1997). On the 6-way Word Order Typology. *Studies in Language*, 21, 69-103. <https://doi.org/10.1075/sl.37.2.02dry>
- Dryer, M. (2000). Explanation of Data Base Format. Manuscript, UBuffalo. (Available online at <http://www.acsu.buffalo.edu/~dryer/wo.vals.html>, accessed on 2020-10-28.)
- Dryer, M. (2007). Word Order. In T. Shopen (ed.), *Language Typology and Syntactic Description* (pp. 61-131). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511619427.002>
- Dryer, M. (2011). The evidence for word order correlations: a response to Dunn, Greenhill, Levinson and Gray's paper in *Nature Linguistic Typology*, 15, 335-380. <https://doi.org/10.1515/lity.2011.024>
- Dryer, M. (2013a). On the six-way order typology, again. *Studies in Language*, 37, 267-301. <https://doi.org/10.1075/sl.37.2.02dry>
- Dryer, M. (2013b). Definite Articles. In M. Dryer & M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/37>
- Dryer, M. (2013c). Indefinite Articles. In M. Dryer & M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/38>
- Dryer, M. (2019). Grammaticalization accounts of word order correlations. In K. Schmidtke-Bode, N. Levshina, S. Michaelis, & I. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence* (pp. 63-95). Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.2583810>
- Dryer, M., with O. Gensler. (2013). Order of Object, Oblique, and Verb. In M. Dryer & M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/84>.
- Dryer, M. & Haspelmath, M. (eds.) (2013). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://wals.info>
- Fanselow, G., Philipp, M., & Zimmermann, M. (submitted). Accessing the availability of inverse scope in German in the covered box paradigm. *Glossa: a journal of general linguistics*.
- Fanselow, G., Traoré, Y., & Féry, C. (submitted). The Object Position in Tagbana. *Journal of African Languages and Linguistics*.
- Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: Some wh-constraints in German. *Lingua*, 115, 1525-1550. <https://doi.org/10.1016/j.lingua.2004.07.003>
- Fominyam, H. (in preparation, 2021). Aspects of Awing grammar and information structure. Doctoral dissertation. Potsdam.
- Fukui, N., & Takano, Y. (1998). Symmetry in syntax: Merge and demerge. *Journal of East Asian Linguistics*, 7, 27-86. <https://doi.org/10.1023/A:1008240710949>
- Geraci, C. & Cecchetto, A. (2013). Neglected cases of rightward movement. When wh-phrases and negative quantifiers go to right. In G. Weibelhuth, M. Sailer, & H. Walker (eds.), *Rightward Movement in a Comparative Perspective* (pp. 211-241). Amsterdam: Benjamins. <https://doi.org/10.1075/la.200.08ger>
- Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (ed.), *Universals of language* (2nd ed., pp. 63-113). Cambridge, MA: MIT Press.

- Hahn, S., Jurafsky, D., & Futrell, R. (2020). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences*, Jan 2020, 201910923. <https://doi.org/10.1073/pnas.1910923117>
- Haider, H. (2013). *Symmetry breaking in Syntax*. Cambridge: Cambridge University Press.
- Haider, H. (2014). The VO-OV split of Germanic languages –a T3 & V2 production. *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis*, 19, 57-79.
- Haider, H. (2015). Head directionality –in syntax and morphology. In A. Fábregas, J. Mateu, & M. Putnam (eds.), *Contemporary linguistic parameters* (pp. 73-97). London: Bloomsbury Academic.
- Haider, H. (2019). On absent, expletive, and non-referential subjects. In P. Herbeck, B. Pöll, & A. Wolfsgruber (eds.), *Semantic and syntactic aspects of impersonality* (pp. 11-46). Hamburg: Helmut Buske Verlag.
- Haider, H. (2020). VO/OV-base ordering. In R. Page & M. Putnam (eds.), *The Cambridge Handbook of Germanic Linguistics* (Chapter 15, pp. 339-364). Cambridge: Cambridge University Press.
- Haider, H., & Szucsich, L. (in press). Slavic languages –"SVO" languages without SVO qualities? *Theoretical Linguistics*. Link to preprint.
- Hale, K. (1983). Warlpiri and the grammar of non-configurational languages. *Natural Language & Linguistic Theory*, 1(1), 5-47. <https://doi.org/10.1007/BF00210374>
- Haspelmath, M. (1995). The converb as a cross-linguistically valid category. In M. Haspelmath & E. König (eds.), *Converbs in Cross-Linguistic Perspective* (pp. 1-55). Berlin: Mouton de Gruyter.
- Haspelmath, M. (2019). Can cross-linguistic regularities be explained by constraints on change? In K. Schmidtke-Bode, N. Levshina, S. Michaelis, & I. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence* (pp. 1-23). Berlin: Language Science Press. <http://doi.org/10.5281/zenodo.2583804>
- Häussler, J., Eythórsson, T., Fanselow, G., Šimik, R., & Vicente, L. (under revision). Crossing movement paths: multiple wh-questions in seven languages. *Journal of Linguistics*.
- Havas, F., Csepregi, M., Gulyás, N. F., & Németh, S. (2015). Typological Database of the Ugric Languages. Budapest: ELTE Finnugor Tanszék. Available at <http://utdb.elte.hu/>
- Hawkins, J. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Hawkins, J. (2008). An asymmetry between VO and OV languages: The ordering of obliques. In G. Corbett & M. Noonan (eds.), *Case and Grammatical Relations: Essays in Honour of Bernard Comrie* (pp. 167-190). Amsterdam: Benjamins. <https://doi.org/10.1075/tsl.81.08.ana>
- Hawkins, J. (2014). *Cross-linguistic Variation and Efficiency*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199664993.001.0001>
- Huang, C.-T. (1982). *Logical relations in Chinese and the theory of grammar*. Doctoral dissertation, MIT. Retrieved from <http://www.ai.mit.edu/projects/dm/theses/huang82.pdf>
- Jackendoff, R. (1990). On Larson's analysis of the double object construction. *Linguistic Inquiry*, 21(3), 427-456. Retrieved from <https://www.jstor.org/stable/4178683>
- Janssen, D., Bickel, B., & Zúñiga, F. (2006). Randomization tests in language typology. *Linguistic Typology*, 10, 419-440. <https://doi.org/10.1515/LINGTY.2006.013>
- Jelinek, E. (1984). Empty Categories, Case and Configurationality. *Natural Language & Linguistic Theory*, 2, 39-76. <https://doi.org/10.1007/BF00233713>
- Jing, Y., Blasi, D., & Bickel, B. (under review, personal communication). Dependency length minimization and its limits: a possible role for a probabilistic version of the Final-Over-Final Condition.
- Johnson, K. (1991). Object Position. *Natural Language & Linguistic Theory*, 9, 577-637. <https://doi.org/10.1007/BF00134751>
- Jurka, J. (2010). The importance of being a complement: the CED effects revisited. Doctoral dissertation, University of Maryland. Retrieved from <https://drum.lib.umd.edu/handle/1903/10835>
- Kayne, R. (1994). *The Antisymmetry of Syntax*. Cambridge, MA: MIT-Press.
- König, C. (2008). Khoisan languages. *Language and Linguistics Compass*, 2, 996-1012. <https://doi.org/10.1111/j.1749-818X.2008.00084.x>
- Larson, R. (1988). On the double object construction. *Linguistic Inquiry*, 35, 319-391. Retrieved from <https://www.jstor.org/stable/25164901>
- Liu, D. (2015). Wuyu he xibei fangyan shoushi qianzhi yuxu de leixing bijiao. [Comparison of patient fronting word orders in the Wu dialect and Northwestern dialects]. *Fangyan*, 2, 97-110.
- Mahajan, A. (1990). *The A/A-bar distinction and movement theory*. Doctoral dissertation, MIT. Retrieved from <http://hdl.handle.net/1721.1/13650>
- Mallinson, G., & Blake, B. (1981). *Language typology: crosslinguistic studies in syntax*. Amsterdam: North-Holland.
- Mathew, R. (2014). *The Syntactic Effect of Head Movement: Wh and Verb movement in Malayalam*. Doctoral dissertation, Tromsø. Retrieved from <https://munin.uit.no/handle/10037/6177>
- Morey, S. (2006). Constituent Order Change in the Tai Languages of Assam. *Linguistic Typology*, 10(3), 327-367. <https://doi.org/10.1515/LINGTY.2006.011>
- Neeleman, A. (2014). Two Asymmetries between Pre- and Post-head Order and their Implications for Syntactic Theory. Retrieved from <https://ling.auf.net/lingbuzz/002207>
- Neeleman, A. & van der Koot, H. (2020). PP Extraposition and Precedence. *Linguistic Inquiry* https://doi.org/10.1162/ling_a_00399
- Nevis, A. & Weisser, P. (2019). Closest Conjoint Agreement. *Annual Review of Linguistics*, 5, 219-241. <https://doi.org/10.1146/annurev-linguistics-011718-012708>
- Nichols, J. (1985). Head-marking and dependent-marking grammar. *Language*, 62, 56-119. <https://doi.org/10.2307/415601>
- Nichols, J. (2016). Polysynthesis and head marking. In M. Fortescue, M. Mithun, & N. Evans (eds.), *The Oxford Handbook of Polysynthesis* (pp. 59-69). Oxford: Oxford University Press <https://doi.org/10.1093/oxfordhb/9780199683208.013.2>
- Osborne, T., & Gross, T. (2016). The do-so diagnostic: Against finite VPs and for flat non-finite VPs. *Folia Linguistica*, 50(1), 97-135. <https://doi.org/10.1515/flin-2016-0004>
- Petronio, K., & Lillo-Martin, D. (1997). Wh-movement and the position of Spec-CP: Evidence from American Sign Language. *Language*, 73, 18-57. <https://doi.org/10.2307/416592>
- Reuland, E., & Kosmeijer, W. (1988). Projecting Inflected verbs. *GAGL*, 29, 88-113. Retrieved from <https://ugp.rug.nl/GAGL/article/download/31074/28380>
- Roberts, I. (2019). *Parameter Hierarchies and Universal Grammar*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198804635.001.0001>
- Saito, M., & Fukui, N. (1998). Order in Phrase Structure and Movement. *Linguistic Inquiry*, 29(3), 439-474. <https://doi.org/10.1162/002438998553815>
- Schäfer, L. (2019). Möglichkeiten und Grenzen der Fragebogenmaterialien des 'Language and Culture Archive of Ashkenazic Jewry' am Beispiel der Wortstellung von Verbparkeln. *Colloquia Germanica Stetinensia*, 28, 5-31. Retrieved from <https://www.ceeol.com/search/viewpdf?id=832540>
- Schiller, E. (1990). On the Definition and Distribution of Serial Verb Constructions. In B. D. Joseph & A. M. Zwicky (eds.), *Ohio State University Working Papers in Linguistics*, 39, 34-64. Retrieved from <http://hdl.handle.net/1811/82138>
- Schmidt, A. (2016). *Udmurt as an OV language and Finnish as a VO language*. MSc Thesis, Potsdam. Retrieved from <urn:nbn:de:kobv:517-opus4-89465>

- Siewierska, A., & Bakker, D. (1996). The distribution of subject and object agreement and word order type. *Studies in Language*, 20(1), 115-161. <https://doi.org/10.1075/sl.20.1.06sie>
- Simpson, J. (1991). *Warlpiri morpho-syntax: a lexicalist approach*. Dordrecht, Boston: Kluwer Academic.
- Sinnemäki, K. (2010). Word order in zero-marking languages. *Studies in Language*, 34, 869-912. <https://doi.org/10.1075/sl.34.4.04sin>
- Skopeteas, S. (2016). *The Caucasian Urums and the Urum language/Kafkasya Urumları ve Urum Dili*. In S. Eker & Ü. Çelik (eds), *Handbook of Endangered Turkic Languages*. Ankara. Retrieved from <urn:nbn:de:0070-pub-29006178>
- Smeets, I. (2008). *A Grammar of Mapuche*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110211795>
- Sposato, A. (2014). Word Order in Miao-Yao (Hmong-Mien). *Linguistic Typology*, 18(1), 83-140. <https://doi.org/10.1515/lingty-2014-0004>
- Szeto, P. Y., Ansaldo, U., & Matthews, S. (2018). Typological variation across Mandarin dialects: An areal perspective with a quantitative approach. *Linguistic Typology*, 22, 223-275. <https://doi.org/10.1515/lingty-2018-0009>
- Vasisht, S., Nicenboim, B., Beckman, M. E., Li, F., & Jong, K. E. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, 71, 147-161. <https://doi.org/10.1016/j.wocn.2018.07.008>
- Wurmbrand, S. (2015). Restructuring cross-linguistically. In T. Bui & D. Özyıldız (eds.), *Proceedings of the North Eastern Linguistics Society Annual Meeting 45* (pp. 227-240). Amherst, MA: GLSA.
- Yiu, C. Y-m. (2014). Typology of Word Order in Chinese Dialects: Revisiting the Classification of Min. *Language and Linguistics*, 15(4), 539-573. <https://doi.org/10.1177/1606822X14532052>
- Zwart, J-W. (1993). *Dutch syntax: a minimalist approach*. Doctoral dissertation. Rijksuniversiteit Groningen. [Link to pdf](#).

1.3.2. Project-related publications by participating researchers

a) Peer-reviewed publications and books

- Fanselow, G. (1987). *Konfiguralität*. Tübingen: Narr.
- Fanselow, G. (2001). Features, θ -roles, and free constituent order. *Linguistic Inquiry*, 32(3), 405-437. <https://doi.org/10.1162/002438901750372513>
- Fanselow, G. (2020). Is the OV-VO distinction due to a macroparameter? In M. Tanaka, T. TsuTsu, & M. Hashimoto (eds), *Linguistic Research as an Interdisciplinary Science* (pp. 1-26). Tokyo: Hituzi Publishers.
- Häussler, J., Grant, M., Fanselow, G., & Frazier, L. (2015). Superiority in English and German: Cross-language grammatical differences? *Syntax*, 18, 235-265. <https://doi.org/10.1111/synt.12030>
- Skopeteas, S., & Fanselow, G. (2010a). Focus types and argument asymmetries: a cross-linguistic study in language production. In C. Breul & E. Göbbel, (eds), *Contrastive information structure* (pp. 169-197). Amsterdam: Benjamins. <https://doi.org/10.1075/la.165.07sko>
- Skopeteas, S., & Fanselow, G. (2010b). Focus in Georgian and the expression of contrast, *Lingua*, 120(6), 1370-1391. <https://doi.org/10.1016/j.lingua.2008.10.012>

b) Other publications

- Fanselow, G. (1991). *Minimale Syntax*. Habilitation thesis. Passau.

1.4. Project plan

With the CHAOS project, we intend to help clarify whether the order of verb and object is a good systematic predictor of grammatical characteristics that do not themselves relate to the order of heads and dependents. In this respect, CHAOS fits into the general prediction testing approach of the second phase of the CRC. To achieve this goal, we want to collect highly comparable data for languages from different areas and language families. On the basis of this data collection, we will find out what influence areal and diachronic factors have on the correlates of the verb-object-position, and, above all, we will identify other factors that interact with the verb-object-position in determining the grammatical profile of a language, and determine the relative strength these factors have in explaining possible correlations. Ideally, the results will enable us to draw conclusions as to how (and whether) such correlates can be predicted in a specific grammatical theory. A further goal is to establish whether the notion of an “underlying” head-complement order is useful for comparative research. This would imply that both OV-order and VO-order can have different underlying sources in different languages, so that CHAOS can also contribute to the “surface stability” idea of the CRC.

We first describe the set of grammatical properties to be investigated and then discuss the choice of languages. Methodological issues of data collection and data analysis are addressed in the third subsection, and the implications for grammar theory in the fourth. Finally, we describe two specific substudies that bring further methods into the project.

1.4.1. Properties to be investigated

The selection of the properties we want to check for a correlation with the position of the verb relative to the object is primarily determined by the question of whether a systematic grammar-theoretical connection between the criteria and the verb's position can be made plausible. In the second place, we want to consider criteria for which a processing-theoretical reason cannot easily be translated into a syntactic law. Finally, however, criteria for which a correlation already appears to have been established should be added because they can serve as a control factor. The number of criteria to be checked should not exceed 20, and, ideally, should be lower.

For obvious reasons, the criteria list proposed by Haider (2013) and Haider and Szucsich (in press) can be taken as a starting point. However, the list given below has been augmented by further criteria based on results from the literature. As it stands now, our construction list is biased towards properties that figure in European languages, so we intend to integrate into the final list further criteria that may be useful for languages from other families and areas (and consequently drop items from the current list). In its first phase, the project will discuss the final composition of the list with its expert partners for the linguistic areas and families (cf. 1.4.3.).

1. Position of the main verb relative to the object

In addition to determining the normal order of the object and the main verb in main (a) and embedded clauses (b), possibly as a function of the main verb's finiteness, we will collect information concerning possible adjacency requirements for the object-verb pair (c). Here, it may be necessary to determine the number and order of elements that can intervene between the verb and the object in order to control for possible parallels with Universal 20 effects (Neeleman, 2014). Furthermore, it must be determined (d) whether it is possible or necessary to place certain elements such as PPs or secondary objects behind the verb in OV-languages (as in Dutch or the OVX-languages, but not in Malayalam) and in front of the verb in VO-languages (XVO-languages as one finds them in the Hmong Mien family), and in what order they appear there (Neeleman & van de Koot, 2020).

The combined inspection of finite and nonfinite verbs in main and embedded contexts will allow an evaluation of the predictions of the Final-over-Final Constraint (Biberauer, Holmberg, & Roberts, 2014), which we will look at as a control because its validity has already been established by various methods, for example, by Jing, Blasi, and Bickel (under review, personal communication).

2. Height of the subject

This criterion is difficult to implement but data must be collected for it because of its theoretical relevance as an alternative source for the clustering of properties we are looking at.

When a language allows low subjects, they can stay in the VP. For VO-languages, the lowest licit position of a subject can thus be determined (or at least estimated) on the basis of the availability of VS-order. Does the subject precede or follow the main verb in intransitive/ergative constructions (a)? Is this ordering obligatory or optional, and are there factors that determine the choice if both orders are fine?

For OV-languages, the structural "height" of subjects will not normally affect the order of V and S, but the height of the subject/its possible location in the VP can be determined by checking whether subjects can be placed between the verb and a clear constituent of the VP, such as an indirect object for the case of a passivized or unaccusative construction, or a relatively low adverb (b).

There may be VO-languages in which the verb occupies a very low position in the VP, and may thus be unable to precede a subject even if the subject stays in VP. This problem can be addressed by determining the ordering of subjects and indirect objects and adverbs in VO-languages as well.

Quite in general, it is advisable to collect the data for both correlates of low subjects for both VO- and OV-languages, and analyse their impact independently first. The two subcases can then be collapsed into one at a later stage of the interpretation of the results.

3. Obligatoriness of the subject position

For languages with unstressed subject pronouns, it needs to be determined whether expletive subject pronouns exist and what the conditions are under which they are inserted (e.g., construction type: transitive, intransitive, unaccusative, passivized, extraposition). The criterion is an indication of a high subject position, or rather of the obligatoriness thereof.

Languages with subject pro-drop do not allow the obligatoriness of the subject position to be tested easily, but if Haider (2019) is correct that "null expletives" do not exist, the presence of impersonal passives (and other impersonal constructions) can be counted as evidence against the obligatoriness of the subject position. As suggested for the previous criterion, later analyses can work with the two subcases collapsed into a single criterion and with the two subcases figuring as independent criteria.

4. Superiority

Does the order of subject and object wh-pronouns in a multiple question correspond to the canonical serialization of subjects and objects (as in English), or is the reverse order possible as well (as in German), or is inversion even mandatory (as in Awing, Fominyam, dissertation to be submitted in 2021)? If the language can but need not cluster wh-elements at the left periphery (as, e.g., in Yiddish), superiority data must be collected for both clustering and non-clustering sequences.

5. Islands for leftward displacement

Subject-object asymmetries characterize leftward movement in English: only direct objects are fully transparent for movement (a Constraints on Extraction Domain (CED) effect, Huang, 1982). Can a subconstituent of X be moved out of X at all, and if so, is this possible independently of whether X is a subject, a direct object, an indirect object, or an adjunct, or is transparency confined to certain grammatical functions? Are there positional constraints (extractions only out of a preverbal/postverbal position)?

Relevant displacement operations that can be tested are *wh*-movement, focus movement, the formation of discontinuous noun phrases (“split topicalization”), and quantifier floating.

6. Free omission of objects

Hawkins (2008) argues that transitive constructions with two lexical arguments come with greater processing difficulty in OV- than in VO-languages, and suggests that the omission of contextually inferable arguments is a way of circumventing this problem. One may thus expect to find more languages that allow the omission of object arguments in OV-languages. The omissibility of both subject and object positions was also proposed as a criterion for non-configurationality by Hale (1983), but the status of the criterion was already relativized by Jelinek (1984). There are also strong areal factors in the distribution of radical object drop (Creissels, 2005).

7. Type of marking of core arguments

Is the language head- or dependent-marking (Nichols, 1985)? How are the core arguments “agent of a transitive verb”, “object of a transitive verb”, and “subject of an intransitive” verb marked morphologically, if at all? Can the two arguments of a transitive verb be distinguished morphologically? And if so, under which pragmatic and syntactic conditions? What is the alignment pattern?

8. Constituent order

Is the order of the constituents in a clause fixed or free? In particular, can discourse-given objects and adverbs swap their positions (as in Dutch), and does the same hold for discourse-given objects and subjects in a transitive clause (A-scrambling, as in German)? Can an object precede a subject when the former is a discourse-topic, a contrast, or a focus (A-bar-scrambling)? See, for example, Corver and Riemsdijk (1997) for the importance of the distinction.

9. VP constituency

The various accounts of VO- and OV-order described above make different predictions concerning the presence of a verb phrase constituent. Can a constituent that comprises the verb, the object, and possibly further material dependent on the verb be preposed/moved in the language? Does this constituent necessarily exclude the subject (as in English, but unlike what holds in German)? Can the constituent that moves be smaller than the maximal VP (a so-called remnant VP as in German)? Is there VP-ellipsis? (But see Osborne & Gross, 2016 for a critical view.)

10. Locative and time expressions

The canonical order of temporal and locative phrases allows inferences concerning the basic branching direction of a language for its adjuncts since temporal expressions must be merged later into the structure than locative ones for semantic reasons.

11. Determiners and nouns

OV-languages turned out to be more likely to lack any article (86 of 187 languages) than SVO-languages (54 of 186) in the survey of Dryer (2013b,c). It has been claimed that the absence of articles (the absence of a DP layer, in generative terms) clusters with many further properties that we are also investigating here; cf. Boškovič (2009). In order to get a better assessment of the internal makeup of complex nominal arguments, one should also register whether the language allows recursive adjectival modification of a noun. Whenever easily available from the literature, the normal order of the heads in the noun phrase should also be registered.

12. Discontinuous constructions

Discontinuity of noun phrases is one of Hale’s (1983) criteria for non-configurationality. Together with Caroline Féry, we have collected pertinent data for some 300 languages that can be easily added to our database. So far, our data suggest that, syntactically, verb-initiality, verb-complement order, the nature of

case marking (phrasality of case), and the presence of A-scrambling play a role in determining if a language allows discontinuous noun phrases.

13. Quantification/pronouns

Bobaljik and Wurmbrand (2012) (among others) have developed a model which predicts that languages with free constituent order should not show scope ambiguity for nominal quantification in the case of sentences with canonical constituent order. Languages with a low subject position (such as Chinese) should also be less likely to display scope ambiguities than fixed order languages with high subjects (cf. Aoun & Li, 1993 for pertinent empirical observations) such as English. It is not clear whether there is a practicable way of checking scope ambiguities in an enterprise like ours, and whether the differences between free and fixed constituent order languages are crisp enough for appearing as a criterion (cf., e.g., Fanselow, Philipp, & Zimmermann, submitted). Together with Project C02, we will carry out pilot studies comparing English, German, Dutch, and Russian in the first half of 2021 in order to have a basis for an informed decision on this possible criterion.

In many OV-languages, objects can bind reflexive possessors of subjects to their right (e.g., in Indian languages, cf. Mahajan, 1990), which is a clear indicator of the kind of free constituent order (A-movement) that one expects to be correlated with OV-order. In some OV-languages, overt possessors of subjects cannot bind object pronouns (e.g., in Warlpiri, cf. Simpson, 1991). This is an indication of the (lack of a) hierarchical organization in the clause that one would expect for VO-languages only.

14. Converbs

Does the language allow converbs (cf. Haspelmath, 1995)?

15. Constraints on finiteness of subordinate clauses

At least for certain types of subordinate clauses, OV-languages are more likely to prescribe the choice of non-finite (“deranked”) verb forms than VO-languages. This holds, for instance, for when-clauses (Cristofaro, 2013). Hence, we will try to establish whether finite subordinate clauses are possible with adverbial function and as subjects and objects, and whether there are restrictions on their placement. Is the nominalization of propositional constituents prescribed in certain constellations?

16. Restructuring and compactness

If a clause contains more than one verb/auxiliary, do these have to be strictly adjacent? Does the language allow restructuring/clause union phenomena? Haider (2013), among many others, observed that in the Germanic family, verbal complexes occur in OV-languages only, and a correlation of clause union/restructuring and verb-object order is also visible in the survey of Wurmbrand (2015): it is only VO-languages that lack restructuring altogether, or that show only restricted types of restructuring.

17. Serial verbs vs clause chaining

Does the language allow serial verbs? If so, is the order of the verbs strictly determined? Does the language allow clause chaining? (See Schiller, 1990.)

18. Tense/modals

Where are tense/modal verbs placed in a main/embedded clause? Can they be combined (as in German, but not in English)? Is their order relative to lexical verbs fixed?

19. Marking of information structure

Do subjects and objects differ with respect to the marking of topic and focus in terms of obligatoriness and the means employed?

20. Wh-scope marking

Constructions such as *was denkst du wann es regnet* with an object wh-word marking matrix scope for a wh-phrase embedded in a complement clause are rare, but there seems to be a tendency for languages possessing the construction to have OV-order (German, Dutch, Sorbian, Japanese, Latin, many Indoaryan languages) rather than VO-order (Serbian and Macedonian, Romani).

There are further properties that can easily be extracted from grammars and can inform us about possible correlates of verb-object ordering such as polysynthesis and tonality (WALS counts 72 SVO-languages with tone systems and 51 without, while 50 SOV tone languages contrast with 101 toneless languages,

perhaps an areal effect). We will add information on such additional properties provided it can be collected as easily as we believe. The placement of adpositions and genitives belongs to this easily available data.

1.4.2. Selection of languages

While Haider (2013) presents an impressive case for the important role played by verb-object order in the syntax of the Germanic languages, the book has not adduced sufficient evidence for establishing this point for the world's languages in general, as mentioned above. The comparison of OV- and VO- languages must be placed on a broader foundation in a more representative way, so that the outcome is less likely to be influenced by a bias of area or language family. Below, five strategies for the selection of the language sample are described.

(i) Our goal is to have each of the macro-areas (Dryer, 1992), viz., North America, South America, Africa, Eurasia without South East Asia, South East Asia and Oceania, and Australia-New Guinea, represented by at least 4 languages in the sample, of which 2 should be OV, and 2 VO. The identification of these 24 languages representing the macro-areas is a crucial first step in the project program: the criteria list will have to be adjusted to the areas on the basis of the interaction with experts for the 24 languages so that a definite criteria list can be compiled with which the data for the 24 language can be collected, making a first comparative analysis possible. Following Dryer (2011), we will attempt to select languages from larger families in this broad comparison. Our approach is reminiscent of the model of typology proposed by Baker (2010).

We will refrain from integrating verb-initial languages into the database in order to reduce the number of factors that could interact with the order of verb and object in determining grammatical properties.

(ii) The items on the criteria list differ in their status: some of them are candidates for alternative explanations for the clustering of properties, unlike others. This holds for the height of the subject, the type of the marking of core arguments, and perhaps also for the freedom of constituent order. The set of languages described under (i) should be complemented by further languages in such a way that there is also enough variability for the three alternative explanatory properties. For example, not all VO-languages in the sample should disallow VS order in addition to SV, or lack morphological marking of grammatical functions.

Ideally, these additional requirements can be factored in by selecting languages that also lead to a broader representativity of language families. At least for morphological marking, we will try to make the set of languages correspond to its frequency in VO- and OV-languages. We expect the size of the language set to at most double in this process of selecting languages for a broader representativity of language families.

(iii) Haider's method of focusing on a language family with variability concerning verb-object order (non-one-type families in the sense of Dryer, 2011) is particularly fruitful, since we can assume that normally there will be fewer further grammatical differences potentially blurring the impact of the OV-VO distinction for pairs of languages from the same family than for pairs of languages from different families. In addition, assuming that the language family was originally uniform with respect to the order of the verb and the object, a detailed analysis of correlations within that family can inform us about changes that are frequently concomitant with a shift from VO to OV or vice versa. Of course, this is true only to the extent that the changes are not due to intensive contact with a language from another family.

Consequently, we will attempt to have OV- and VO-languages from such families represented in our sample. Andreas Schmidt has made some considerable progress in collecting relevant data for the Uralic family, and it is fairly easy to fill gaps that may still exist for Germanic. We plan to investigate (at least) two further non-uniform language families. Candidate families are given in the following paragraph.

Gagauz and Urum (see Skopeteas, 2016) are easily accessible VO-members of the Turkic language family, which is otherwise OV, but to the extent that Gagauz and Urum have been influenced by Russian, a language that does not have the 'typical' VO-profile, their language contact situation would not lead one to expect marked grammatical differences from the other Turkic languages. They should best be contrasted with other European Turkic languages. Within the Afro-Asiatic family, the Semitic subfamily allows a comparison of VO- and OV-languages (Amharic, Tigre, Tigrinya). Kordofanian also lacks uniformity with respect to verb-object order (Moro is VO while Talodi languages are OV). Dogon, Gur, and Ijad are OV-subfamilies within the largely VO-dominated Niger Congo family. Nilo-Saharan languages would allow the comparison of Kanuri (OV) with Luo (VO) and Nuer and Dinka (VO or undetermined). The Khoisan languages would be particularly interesting because of the well-known differences between Khoe and Non-Khoe languages (König, 2008), which are reminiscent of the breadth of variation we have in Germanic: Non-Khoe languages are isolating and come with VO-order, while Khoe languages are inflectional and show OV-order. The latter lack serial verb constructions, while the former have them, as expected. Tai languages invite a comparison of, for instance, VO-Thai with OV-Khamti (but the latter

language may also be classified as having no fixed verb order, Morey, 2006). Munda languages differ from Mon Khmer languages with respect to verb placement within the Austroasiatic language family. The Austronesian Sepik family has languages with different verb orders, and so do the Iroquoian, Hokan, Uto-Aztecan, Arawak, Tupi-Guarani, and Carib language families. There are a few Papuan and Australian languages that have VO-order, such as Thayorre and Wangkumara.

(iv) West African OVX-languages (e.g., Mande, Senufo families) differ from “normal” OV-languages in many ways, so that they seem to resemble VO-languages in many respects. Such languages should be subject to a separate investigation, together with Chibchan languages, many of which also belong to the OVX-type. XVO-languages are to a certain extent the mirror image of the OVX-type, but it is not known whether they show comparable marked deviations from the standard VO-pattern. We plan to collect data for the Hmong Mien family, in which a number of XVO-languages can be found (Sposato, 2014). The minimum of languages to be studied would be 2 per family.

(v) Finally, there are individual languages that are of particular interest and that should figure in our final account. First, some languages have marked differences between transitive and intransitive sentences in serialization, for example, Mapuche (Smeets, 2008, p.20) with VO- and VS-order or Huastec (Constable, 1989, p. 43), which also shows VO combined with a VS/SV-serialization.

Second, languages can display considerable internal variability with respect to the placement of the verb. For example, several dialects of Mandarin such as Huangshui (Szeto, Ansaldo, & Matthews, 2018), Hezhou (Liu, 2015), Wu (Yiu, 2014), and Xining (Szeto et al., 2018) do not conform to the VO-pattern of standard Mandarin, and show OV-like properties such as scrambling (Xining, cf. Bell, 2019) and case marking (Huangshui, cf. Szeto et al., 2018).

Third, languages with a recent shift in verb placement such as Gagauz or Purepecha (Chamoreau, 2017) or a recent shift from Slavic-like patterns to patterns similar to English such as Hassidic Yiddish can shed light on the effects of verb placement in a particular way.

(vi) As long as it is compatible with the project’s progress relative to its time frame, we will opportunistically add languages to the sample beyond the strategies (i)–(iii) whenever possible in order to augment the database. This can also include languages with verb-initial orders (VSO, VOS), which we have deliberately excluded from our data collection in (i).

1.4.3. Method of data collection

A broad-based survey on the languages of the world requires cooperation with a network of scholars. Many of the grammatical characteristics we want to survey can certainly be found in the rich treasure of grammars or other syntactic descriptions. In this respect, we will try to focus on languages for which sufficient linguistic analyses are already available, so that the basics, such as the normal order of O and V or the type of alignment of the arguments, need not be established by the project. But even in these languages, some of the characteristics, such as order of the *wh*-words in multiple questions or the pragmatics of placing objects in front of the subject, can often not be taken from existing literature.

Therefore, we are in the process of building a network of experts who will help the project add missing data points. Depending on the particular case, these gaps may be filled on the basis of the expertise of the colleagues, but often they may require work contracts for surveys of judgements of native speakers on a limited scale. For this purpose, we are applying for a few work contracts (mainly) for students of the experts. Work contracts will be necessary in particular for the study of OVX- and XVO-languages and for Mandarin dialects, but also for the systematic comparisons of languages in the same family. Such work contracts include the costs of local travel, remuneration of the informants, and a honorarium for the contractor. We set a sum of 800 Euros per contract, and estimate the need to issue 30 such contracts over the course of the project work (=24.000€).

In addition, we are applying for funding for remunerating informants who work with us or directly with our experts in Germany or abroad (outside the context of the work contracts described above). Nowadays, native speakers of many languages are accessible in the major cities, and can be interviewed directly or via the internet. Including all possible costs of the informants, we calculate with 20€ per hour and plan 200h of interviews by Potsdam project members and 100h of interviews by external partners. We therefore request 6000€ and a sum of 3000€ for possible regional trips to informants.

In the pilot data collection for Tagbana (Senufo) and Taoih (Austroasiatic), two working days of interviews were sufficient to collect the information for more than half of the criteria listed above, but the work was, of course, preceded by a consultation of grammars for related languages. Thus, we expect that roughly a week will have to go into the collection of data for one language. Hence, we consider the amount of data collection described in (i)-(iv) of 1.4.2 to be quite feasible.

We would like to achieve different degrees of intensities of cooperation with our partners, from the simple contribution of data to the joint writing of cross-language overviews. Interviews of informants will be carried out over the internet, so that no air travel will be necessary for this. This makes the project programme, by and large, non-sensitive to possible pandemic-related restrictions and contributes to science's climate neutrality.

We are grateful to the following colleagues who have declared their willingness to collaborate with us: Walter Bisang (Mainz; Sinitic languages, Austronesian, Mande languages), Dubi Nanda Dakal (Kathmandu; Tibeto-Burman and Indoaryan languages), Christian Döhler (Berlin and Potsdam; Papuan languages), Ayesha Kidwai (Delhi; languages of India), Gerson Klumpp (Tartu; Uralic languages), Mathias Jenny (Zürich, Austroasiatic), Pieter Muysken (Utrecht; Southern America), Nguyen Ti Suu (Vietnamese languages), Masha Polinsky (Harvard; Caucasian languages), Uli Reich (Berlin; Quechua, South America), Catherine Rudin (Wayne State; Siouan), Matthias Schlesewsky (Adelaide; Australian languages), Shi Defu (Beijing; Hmong Mien), K. V. Subbarao (Delhi; languages of India), Balasz Suranyi (Budapest; Uralic), Lisa Verhoeven (Berlin; Mesoamerica), Wang Bei (Beijing Institute of Technology; Sinitic languages, Hmong Mien languages). For languages of Sub-Saharan Africa, we intend to base our work on the networks of our fellow CRC-members Doreen Georgi and Malte Zimmermann. Our gratefulness also extends to Hubert Haider, Salzburg, Martin Haspelmath, Leipzig, Ad Neeleman, London, and Barbara Stiebels, Leipzig, who have agreed to consult with us, and to the proposed Mercator Fellow Kriszta Szendrői, London. We will get advice on possible prosodic issues from our long-term scientific partner Caroline Féry.

We plan to invite six of the consultants for a longer stay in Potsdam to jointly modify the list of criteria in accordance with the requirements of specific macro-areas or language families and to optimize the choice of languages. These visits will take place during the first 9–12 months of the project. We hope to be able to integrate more experts into our cooperation network in the course of the coming weeks.

1.4.4. Data analysis

While the intended interpretation of the project's data as described in 1.4.6 will be largely qualitative in nature, a quantitative data analysis is necessary for a firmer support of the qualitative interpretation. The data gathered in the project will mostly consist of feature–value pairs with binary, categorical, and ordinal data types as in other typological databases such as Matthew Dryer's original word order database (Dryer, 2000), the *World Atlas of Language Structures* (Dryer & Haspelmath, 2013), the *AUTOTYP typological databases* (Bickel et al., 2017), and the *Typological Database of the Ugric Languages* (Havas et al., 2015). Most of the criteria in 1.4.1 are *multivariate* in the sense of Bickel (2010, 2011) in that, for example, criterion 7 *Type of marking of core arguments* also requires information about the conditions under which arguments of the verb are morphologically distinguished.

There is no consensus on how typological data ought to be analysed and how correlations between features are to be statistically established. Two difficulties are that diachronic (genetic and areal) factors need to be accounted for, and that languages cannot be randomly sampled (cf. Janssen, Bickel, & Zúñiga, 2006; Bickel, 2013). The statistical methods employed reach from binomial tests (Bickel et al., 2015) and Fisher's exact tests for significance testing (e.g., Nichols, 2016), to linear models for assessing the relative impact of different factors (e.g., Sinnemäki, 2010; Bickel, 2017) and Bayesian methods (also Bickel et al., 2015). We will employ linear regressions as frequentist measures. Additionally, we will employ Bayesian models to account for the fact that typological surveys are non-repeatable events. This approach is also in line with current developments in the field advanced by Balthasar Bickel. We will draw on expertise from the Q-project, which has considerable experience in the application of Bayesian data analysis to linguistic data (Vasishth et al., 2018).

1.4.5. Pursuit of additional research perspectives

We propose to complement the core of the comparative work in this research project in two ways.

i. The fact that the characteristics of the languages that we will investigate involve categorical distinctions obviously involves an idealization, as described above: languages may, for instance, prescribe the use of expletives in certain contexts only. For the purpose of establishing implicational relations between the criteria, one can map the values of the criteria to binary values by varying a “threshold” above which a property is considered to be present. For example, for the characteristic “subject expletive”, we register the contexts in which expletives occur, and can map these to a binary value in various ways (Is the existence of a single context sufficient? Or are several contexts required for assuming a positive value?). By comparing different conversions, the most predictive possibility can be identified.

Besides such conversions to a binary interpretation of the properties, we also want to test different options of a gradient approach. One possibility is to assign numerical values to the characteristics in such a way that they represent the linguistically systematic degree of expression of a characteristic. For example, it is possible to determine how many categories can occur behind the verb in an SOV-language (none, only CPs, only CPs and PPs, all categories), defining thereby a degree of “strictness” of verb-finality.

A more direct approach would try to read the degree of the expression of a certain characteristic off linguistic data, for example, from corpus frequency. However, such frequency data are hardly available for many of the properties we are interested in. Often, this is due to the fact that such corpora would have to be very large to allow the extraction of meaningful information. This is, for instance, the case for word order in multiple questions. Corpus frequencies, however, correlate with acceptability data. Therefore, we will try to determine the mean acceptability for the variants of some of our criteria for a number of languages from the same family with non-uniform verb placement through rating experiments, and compare the results with those of the binary interpretation. In this context, we will use the platform L-Rex (<https://www.l-rex.de/>) for the remote collection of acceptability ratings. Members of our group have already applied the platform in the course of collecting acceptability ratings for Awing directly in Cameroon.

We plan to run such a study with 6 languages of a single family. The costs of 632€ per experiment are composed of the remuneration for the participants (calculating 36 subjects per experiment who receive a 12€ honorarium) and the costs for having the material translated by a native speaker. We are also applying for the sum of 2000€ for two work contracts to adapt the platform to the languages we will be testing.

ii The interaction with our Mercator Fellow Kriszta Szendrői will allow us to directly observe a shift of a language from one of the grammatical systems we are investigating to another. While Ashkenazi Yiddish is grammatically close to the Slavic languages, and will therefore belong to the languages with undetermined verb serialization in Haider’s terminology (see also Schäfer, 2019 for the branching direction in the VP), Hassidic Yiddish in London has developed into a language with fixed word order. The interaction with Prof. Szendrői opens up access to the characteristics of a very recent VO-language in the narrower sense.

1.4.6. Interpretation of the results

The main objective of the project is to examine the relationship between the placement of verbs and other syntactic properties that do not themselves belong to the domain of the arrangement of heads and complements/adjuncts. We hope to be able to identify clear relationships between the at most 20 syntactic properties we are examining and the relative importance of factors such as “height of subject” or the influence of areal factors.

At the same time, we will try to evaluate the results of the survey in terms of grammar theory, i.e., we will try to see how well they fit in with the predictions of approaches such as Haider (2013), and to decide what changes would have to be made to such approaches so that they are more compatible with the findings.

In addition, we also expect to be able to draw conclusions on the relationship between theory and empirical foundations in language comparison. Some of the characteristics we will investigate can in principle be read directly from the linguistic surface (e.g., the order of subject and object in a multiple question), others are more indirect like the height of the subject, i.e., they are determined on the basis of a cluster of word order relations which need not be identical in different languages.

Furthermore, the question has been discussed whether typological generalization can or even must refer to “underlying” sequences (Haider, 2015) since, for example, the superficial order of the verb and the object can be the result of movements that map an underlying order to some different surface linearization. This surface order might be irrelevant for the functioning of syntax if one assumes (as Haider does) that underlying verb-object order informs us about the branching direction in a clause, which is the real cause of the correlations we are investigating. Concretely, the verb is not always placed in final position in German because of the finite second property, and Tariana shows crisp verb-finality only in the subordinate clause, while verb placement in the main clause is apparently free (Aikhenvald, 2003). Still, the grammars of German and Tariana strongly resemble the grammars of OV-languages. Conversely, OVX-languages such as Tagbana have many characteristics of VO-languages, and one could explain this as Fanselow, Traoré, and Féry (submitted) do by assuming that a (typologically relevant) VO-base structure is transformed into OV by a later object movement.

If such reflections are valid, the typological generalizations would (sometimes) refer to hidden properties (height of subject position) or abstract properties (basic position of the verb). However, alternative explanations could, for example, refer to verb-object adjacency as the driving force of the typological relationship, which distinguishes OVX-languages from ‘standard’ OV-languages such as Japanese or Hindi. We hope that our data will allow us to estimate which of the tested characteristics have the strongest