Symmetry and Structural Equivalence in Grammar

Thomas Holder

Symmetries deal with invariant properties of objects and have become in the guise of groups one of the key concepts of modern mathematics and physics. The first attempt to apply these concepts to syntax is the *Bare Grammar* framework of Keenan and Stabler ([12]). The aim is to study natural languages via syntactic invariants of their grammars.

Bare Grammar and Syntactic Automorphisms

Kobele(2002) [13] has pointed out that bare grammars are just finitely generated partial algebras and hence techniques from universal algebra apply[1)]. We endorse this view here.

Def. A *bare grammar* is a partial algebra $G = \langle L_G, Rule \rangle$ of finite signature, with $Rule$ a finite indexed set of partial generating functions over this signature, and G finitely generated from a finite set of lexical items $lex \subseteq L_G$.

Def. A function $f : L_{G_1} \to L_{G_2}$ between two bare grammars of the same signature has the *homomorphism property* with respect to a generating function $g_1^i \in Rule_1$ when $f(g_1^i(x,y)) = g_2^i(f(x), f(y))$ for all x, y in the domain of definition of g_1^i. [2)] f is called a *strong homomorphism* if f has the homomorphism property for all $g_1^i \in Rule_1$ and definedness of the lefthand side implies definedness of the righthand side in each case.

Prop.([12]) *Let G be a bare grammar. Then the set $Aut(L_G)$ of all strong automorphisms $\alpha : L_G \to L_G$ is a group, the automorphism group of L_G.*[3)]

Def. Two expressions $x, y \in L$ are *structurally equivalent* if there exists $\alpha \in \mathrm{Aut}(L)$ such that $\alpha(x) = y$.

The resulting equivalence relation on L is the central tool in the study of syntactic invariants.

[1)] For the relevant material on partial algebras see [10] or [8]. [15] and [9] contain the basic group theory. [2] is a good source for category theory. For Keenan and Stabler's original definition see their book [12].

[2)] Especially definedness of the righthand side implies definedness of the lefthand side. We mostly assume generating functions binary for ease of notation and write diacritics only when needed.

[3)] α is determined by its values on lex. When no lexical item x is in the range of a generating function then $\mathrm{Aut}(L_G)$ is finite. (See [12])

Syntactic Sorts

In Keenan and Stabler's original definition of a bare grammar a set of syntactic sorts forms part of the structure. We reconstruct sorts with the help of congruence relations[4].

Def. An equivalence relation $\cong$ on L is a *strong congruence* if for all $x, y \in L$ and all $g \in Rule$ such that $x \cong y$ and either $g^i(\ldots,x,\ldots)$ or $g^i(\ldots,y,\ldots)$ is defined, implies definedness of the other and $g^i(\ldots,x,\ldots) \cong g^i(\ldots,y,\ldots)$.

Strong congruence relations correspond to kernels of strong homorphisms and are partially ordered by inclusion. There exists a maximal strong congruence relation μ on L.[5]

Obs. *Let $\pi : L \to Sort$ be the function that assigns to an expression x its sort $\pi(x)$. Surjectivity and sort functionality in the sense that $\pi(g(x,y))$ is a function $\hat{g}$ of $\pi(x), \pi(y)$ suggests that π is a surjective homorphism to a "sort algebra".*

There is a universal such algebra namely the quotient algebra L/μ of L under the maximal strong congruence μ: it yields the biggest possible sorts hence we take L/μ and the canonical projection $\pi : L \to L/\mu$ as our syntactic sort structure.[6]

Prop. *Let $\varphi : L \to L'$ be any strong surjective homomorphism with domain L. Then there exist a unique strong surjective homomorphism $\psi : L' \to L/\mu$ such that $\pi = \psi \circ \varphi$.*

Def. A function $f : L \to L$ is *sort preserving* if $\pi \circ f = \pi$. The group of all sort preserving bijections is denoted $\mathrm{Aut}(\pi)$. The group of all sort preserving automorphisms is $\mathrm{Aut}_\pi(L) := \mathrm{Aut}(\pi) \cap \mathrm{Aut}(L)$.

Def. A function $f : L \to L'$ *preserves sortal equality* if $\pi(x) = \pi(y)$ implies $\pi'(f(x)) = \pi'(f(y))$. f *preserves sortal inequality* if $\pi(x) \neq \pi(y)$ implies $\pi'(f(x)) \neq \pi'(f(y))$. A function preserving both *respects sorts*.

Sort respecting functions can neither fission nor fusion sorts. The following proposition is a direct consequence of the universal property of the sort algebra.

Prop. *A surjective strong homomorphism $\eta : L \to L'$ preserves sortal equality.*

Prop. *A syntactic automorphism $\alpha \in Aut(L)$ respects sorts.*

This shows that a syntactic automorphism α can at most exchange a sort with another sort. In most cases α will just relate expressions of the same sort and derivational depth. Hence $\mathrm{Aut}(L)$ can only probe into sameness and similarity of distribution. This in turn suggests that it is unable to account for postdistributional regularities that are in the focus of linguistic research since the Chomskian turn. Distributional symmetries might nevertheless shed light on natural languages when supplemented with semantic considerations.[7]

Cor. *$Aut_\pi(L)$ is a normal subgroup of $Aut(L)$.*

The sort exchanging automorphisms form no subgroup as they lack the sort preserving identity and they are not multiplicatively closed in general. But we can form the quotient group under a normal subgroup and define $X_\pi(L) := \mathrm{Aut}(L)/\mathrm{Aut}_\pi(L)$ with canonical

[4] The affinity of congruences on L to syntactic sorts has been pointed out in [13].

[5] See [10].

[6] This is akin to the construction of the minimal automaton in formal language theory ([9]).

[7] Keenan and Stabler ([12]) advance interesting proposals on the relation between syntax and semantics.

projection p. This group does not act on L as it consists of the left cosets $\alpha \cdot \mathrm{Aut}(L)$ of automorphisms $\alpha \in \mathrm{Aut}_\pi(L)$ but it is as close to a group of sort eXchanges as we can get.

Prop. *$Aut(L)$ is a group extension of $X_\pi(L)$ by $Aut_\pi(L)$ i.e. the following is a short exact sequence of groups*[8]*:* $1 \to Aut_\pi(L) \hookrightarrow Aut(L) \xrightarrow{p} X_\pi(L) \to 1$.

There is a parallel to the occurrence of fiber bundles $F \to E \to B$ in gauge field theory where the base B corresponds to the symmetries of space-time and F to internal symmetries: the symmetries of the total space arise by combining the internal symmetries with the spatial symmetries over every space-time point (See [1, 16]). This suggests to view $X_\pi(L)$ as the symmetry group of a base grammar from which L results by extension. The points of this base grammar have $\mathrm{Aut}_\pi(L)$ as internal symmetry group.

Language Extensions

Let $L \subseteq \overline{L}$ be a language extension and H be a subgroup of $\mathrm{Aut}(L)$. We are interested in extending H to a subgroup of $\mathrm{Aut}(\overline{L})$.

Def. The *lift* of H to $\overline{L}$ is defined as $\overline{H} := \{\overline{\alpha} \in \mathrm{Aut}(\overline{L}) : \overline{\alpha}\,|_L = \alpha \in \mathrm{Aut}(L)\}$.

Prop. *$\overline{H}$ is a subgroup of $Aut(\overline{L})$ and the restriction $\rho : \overline{H} \to H, \overline{\alpha} \mapsto \overline{\alpha}\,|_L$ is a group homomorphism.*

Def. $\mathrm{Aut}(\overline{L} : L) := \rho^{-1}(\mathrm{id}_L)$ is called the *relative automorphism group* of the extension $L \subseteq \overline{L}$.

$\mathrm{Aut}(\overline{L} : L)$ consists of all extensions of the identity of L i.e. all automorphisms of $\overline{L}$ that fix L pointwise.

Obs. *The sequence* $1 \to Aut(\overline{L} : L) \hookrightarrow \overline{H} \xrightarrow{\rho} H$ *is exact.*

There is no guarantuee that the sequence can be extended on the right to a short exact sequence i.e. ρ is not surjective in general and there exist $\alpha \in H$ that fail to extend to $\overline{L}$. Unfortunately this happens for $H = \mathrm{Aut}(L)$ in cases where the language extension results from the addition of lexical items: sort exchanging automorphisms depend on the lexical sort cardinality as they have to effect a bijection between sorts ([12]). If ρ is surjective we have a group extension and $\overline{H}/\mathrm{Aut}(\overline{L} : L) \cong H$. $\mathrm{Aut}(\overline{L} : L)$ is hence a normal subgroup of any surjective lift.

Def. $L \subseteq \overline{L}$ is called a *Galois extension* if $i : L \hookrightarrow \overline{L}$ preserves sortal equality and induces an injective homomorphism $i_* : L/\mu \to \overline{L}/\overline{\mu}$. In this case $\mathbf{Gal}(\overline{L} : L) := \mathrm{Aut}(\overline{L} : L)$ is called the *Galois group* of the extension.

The abusive terminology seeks to point to the parallels with the Galois Theory of field extensions ([5, 6]). When i_* is an isomorphism then $\mathbf{Gal}(\overline{L} : L)$ cannot contain sort interchanging automorphisms: $\mathbf{Gal}(\overline{L} : L) \subseteq \mathrm{Aut}_\pi(\overline{L})$. This case corresponds to purely lexical extensions.

Conjecture. $\overline{Aut_\pi(L)}/\mathbf{Gal}(\overline{L} : L) \cong Aut_\pi(L)$.

This basically says that sort preserving automorphisms α have an extension $\overline{\alpha}$ provided

[8] For the terminology see [15]. We reversed extending and extended group as is customary in some of the literature in order to comply with the fiber bundle perspective below.

the language extension $L \subseteq \overline{L}$ is sufficiently nice.
In linguistics implicational relations between grammatical constructions in the sense of Greenberg can be cast as problems of language extensions. This suggests that the *relative symmetry* provided by the Galois group turns out to be linguistically more interesting than the absolute symmetry provided by $\mathrm{Aut}(L)$.

Stable Structural Equivalence

The structural equivalence relation resulting from the action of $\mathrm{Aut}(L)$ is not stable under lexical extensions ([12]). Here we briefly discuss two ways to remedy this situation.

Rem. The cheapest way out is by retreat to the sort algebra L/μ. Consider two expressions $x, y \in L$ as structurally equivalent when $\pi(x) \cong \pi(y)$ in L/μ under the action of $\mathrm{Aut}(L/\mu)$. But this has the drawback that all expressions of the same sort come out as being structurally equivalent e.g. a conjunction of nouns to a noun. Hence the resulting equivalence relation expresses similarity of distributional structure rather than similarity of structure.

Rem. $\mathrm{Aut}_\pi(L)$ is sensitive to the derivational structure of expressions of the same sort and so $L/\mathrm{Aut}_\pi(L)$ makes for a second candidate to negotiate the equivalence of $x, y \in L$ upon. Constructing this quotient in the context of categorical syntax[9)] reveals another grammar fibration: Let $\mathscr{C}_G$ be the syntax category generated by a production grammar G. Define $\mathrm{Aut}(\mathscr{C}_G)$ as the group of all invertible grammar functors $F : \mathscr{C}_G \to \mathscr{C}_G$ and $\mathrm{Aut}_\pi(\mathscr{C}_G)$ as the subgroup of grammar functors that fix the nonterminal symbols pointwise. The idea is now to consider two syntax categories as structurally equivalent if there is an equivalence of category via a pair of grammar functors between them (cf. [18]). In itself this is not enough because no morphisms witness the structural equivalence of expressions in $\mathscr{C}_G$. But application of the Grothendieck construction ([2, 7]) to $\mathrm{Aut}_\pi(\mathscr{C}_G)$ viewed as a category with a single object and $F \in \mathrm{Aut}_\pi(\mathscr{C}_G)$ as morphisms and $\mathscr{C}_G$ viewed as a constant functor yields a category $\int_{\mathrm{Aut}_\pi} \mathscr{C}_G$ with isomorphisms between expressions in the same $\mathrm{Aut}_\pi(\mathscr{C}_G)$-orbit. There is an exact sequence $\mathscr{C}_G \hookrightarrow \int_{\mathrm{Aut}_\pi} \mathscr{C}_G \to \mathrm{Aut}_\pi(\mathscr{C}_G)$: $\mathscr{C}_G$ can be recuperated as the fiber over the identity functor. When $\mathscr{C}_G$ had not enough morphisms then $\int_{\mathrm{Aut}_\pi} \mathscr{C}_G$ has too much and we still have to quotient by the congruence relation generated by $(\mathrm{id}_C, F) \simeq (\mathrm{id}_C, H)$ on the arrows of $\int_{\mathrm{Aut}_\pi} \mathscr{C}_G$. We consider now two syntax categories as equivalent if there is an equivalence of category between the respective quotient categories $\int_{\mathrm{Aut}_\pi} \mathscr{C}_G / \simeq$.

Conclusion

We have discussed the concept of structural equivalence from Bare Grammar. The introduction of a sort concept clarified the structure of the automorphism group. Then language extensions were discussed: This permitted to introduce a concept of relative symmetry. We sketched a concept of structural equivalence that behaves well with respect to lexical extensions. We emphasized the role of exact sequences and fibrations throughout as this highlights the connection to familiar symmetry concepts from algebra and algebraic topology. Although this "syntactic Galois theory" is embryonic at the present stage

9) See [3, 4, 17] for details and terminology of this approach.

there is hope that more than concise formulations can be gained from it: Fiber bundles are generalized products and come equipped with a notion of locality. They might hence provide the right framework to deal with the modalization of proof-theoretic grammar frameworks considering Lawvere's "modality as local truth": product grammars overshoot ([14]) or undershoot ([11]) but fibring over the identity localizes to a nice language in a Chomsky-Schützenberger-type manner[10)]. On this view a grammar is an extension of a kernel by a transformational hull, as in the Chomsky-Harris tradition, tamed by a symmetry group, that modulates the passage from local to global structures.

References

[1] Baez. *Higher Categories, Higher Gauge Theory* - I. Unni Namboodiri Lectures, 2006. (available at http://math.ucr.edu/home/baez/namboodiri).

[2] Barr/Wells. *Category Theory for Computing Science*. Prentice-Hall New York, 1990.

[3] Benson. *Syntax and Semantics: A Categorical View*. Information and Control **17** pp.145-160, 1970.

[4] Benson. *The Basic Algebraic Structure in Categories of Derivations*. Information and Control **28** pp.1-29, 1975.

[5] Borceux/Janelidze. *Galois Theories*. Cambridge University Press, 2001.

[6] Bosch. *Algebra*. 4th ed. Springer, Heidelberg 2001.

[7] Brown. *Topology and Groupoids*. 3rd ed. BookSurge Deganwy, 2006.

[8] Burmeister. *Lecture Notes on Universal Algebra - Many Sorted Partial Algebras*. Ms. Technische Universität Darmstadt, 2002. (available in electronic form).

[9] Cohn. *Further Algebra and Applications*. Springer Heidelberg, 2003.

[10] Grätzer. *Universal Algebra*. Van Nostrand Princeton, 1968.

[11] Kambites. *Formal Languages and Groups as Memory*. Ms. Universität Kassel, 2006. (arXiv:math.GR/0601061 v1).

[12] Keenan/Stabler. *Bare Grammar*. CLSI Stanford, 2003.

[13] Kobele. *Structure and Similarity*. Ms. UCLA, 2002. (available in electronic form).

[14] Kobele/Kracht. *On Pregroups, Freedom, and (Virtual) Conceptual Necessity*. U.Penn Working Papers in Linguistics **10** 1, 2005. (available in electronic form).

[15] Mac Lane/Birkhoff. *Algebra*. Macmillan New York, 1967.

[16] Nakahara. *Geometry, Topology and Physics*. 2nd ed. Taylor&Francis New York, 2003.

[17] Nelson. *Categorical and Topological Aspects of Formal Languages*. Math. Systems Theory **13** pp.255-273, 1980.

[18] Straßburger. *What is a Logic, and What is a Proof ?* Ms. Universität des Saarlandes, 2005. (available at http://www.ps.uni-sb.de/˜lutz).

[19] Walter. *Verallgemeinerte Pullbackkonstruktionen bei Semi-Thuesystemen und Grammatiken*. EIK **6** 4/5 pp.239-254, 1970.

thomas.holder@freenet.de

[10)] To illustrate: the Chomsky-Schützenberger Theorem about contextfree languages C should yield a bundle of the sort $C \to R \to D$ where R is regular and D is a Dyck-language. Fibered products of grammars are also studied in [19].