IP 02, University of Oregon
"Sahaptian and the evolution of hierarchical systems"

Team
Spike Gildea (Principal Investigator)
Joana Jansen (PhD student / postdoc)

IP aims and objectives
This IP is has concrete descriptive goals and also diachronic typology goals. The concrete descriptive goals are to annotate a large corpus of texts in the Sahaptin (Sahaptian). We will categorize texts by genre and dialect of speaker, and we will tag individual sentences for the information described in the CRP, including semantic factors such as animacy, pragmatic factors such as global frequency and topic persistence, as well as for episode boundary. The analysis of the Sahaptin texts will contribute to the CRP typological databases. The diachronic typological goal is to characterize different types of hierarchical systems not just in terms of their grammatical engagement with the hierarchy, but also in terms of their **source** (etymological origins) and **pathway** (developmental stages, including mechanisms used to introduce new stages). Towards the ultimate goal of generating an overall diachronic typology of hierarchy effects in grammar, this project will reconstruct the sources of hierarchical grammar in Cariban and Sahaptian language families, as well as in isolate Movima. I now briefly summarize each goal independently.

Sahaptin Corpus Development
The Sahaptin and Nez Perce languages comprise the Sahaptian Family, which is classified, along with Klamath, into the Plateau branch of Penutian; Sahaptin is still spoken in Oregon and adjacent states. By the most optimistic estimates, Sahaptin has 125 speakers, but all fully fluent speakers are over 60 years of age. Both Sahaptin and Nez Perce are syntactically nominative-accusative, but both present a series of second-position SAP person-marking clitics (SAP > 3), and Sahaptin (but not Nez Perce) independently uses both case-marking and an inverse prefix to signal a distinction between 3PROX > 3OBV (c f. Rude 1994, Rude and Rigsby 1996, Zúñiga 2006). The Sahaptin system is typologically unusual in that the full hierarchy, 1 > 2 > 3PROX > 3OBV, is distinguished grammatically, but via the interaction of different grammatical devices, which do not together form a grammatically unified system. 1 > 2 is distinguished by a unique 1A2O second-position clitic (direct) versus a second person clitic plus the inverse prefix for 2A1O. SAP > 3 is distinguished by SAP always appearing in second position (not varying for grammatical role), 3 always appearing as a verbal prefix (differentiating A and O), and 3A further bearing the 'INVERSE/ERGATIVE' case, whereas 3O bears an accusative case suffix sensitive to animacy (obligatory for human O, optional for animate O). 3PROX > 3OBJ is distinguished by a direct clause with an unmarked A, differentially-marked O, and a simple 3S/A prefix on the verb; in the inverse clause, the A takes an ERGATIVE/OBVIATIVE case-marker, the O is differentially-marked, and the inverse prefix *pá-* occurs on the verb. These complications are interesting not only for synchronic description and typological classification, but also for reconstruction (to be discussed shortly). For now, we focus on the problem of predicting which third person participant in a 3A3O clause will be PROXIMAL, and which OBVIATIVE.

Two studies have attempted to determine the function of the Sahaptin inverse voice. Rude 1994 uses a text-counting methodology with 22 pages of Jacob's Klickitat texts (from Jacobs 1929; these texts will be included in the project database). Rude claims that the Northwest Sahaptin inverse voice is a patient topicalizing construction in which the agent still has high topicality. Blackburn Morrow (2006), in extending Rude's text counting methodology to the Umatilla dialect, found tentative support for this theory, but with an unsettling amount of residue (also seen in Rude 1994). In her subsequent experimental study, five native speakers produced narratives in response to wordless storybooks designed to elicit direct versus inverse voice. Blackburn Morrow found that, although neither was without exception, animacy and global frequency (defined as the total number of times a referent was mentioned in an entire text) of A and O are also key factors to explain Umatilla Sahaptin speakers' decision to use the inverse. However, examples are found in texts that don't quite fit the explanations offered in prior research, and Blackburn Morrow notes that in her data (necessarily collected from a very small number of speakers given the language's stage of endangerment), not all of the elders' uses of direct versus inverse were as predicted. She indicates that factors such as episode theme, episode boundaries, and the backgrounding of A may also be important factors.

This project will build on ongoing projects with the Sahaptin tribe at the University of Oregon, to create an annotated corpus of all previously published Sahaptin texts and a subset of recently recorded Sahaptin text. It is especially crucial for us to work with the older speakers now, both to record additional texts and to analyze older texts, as the speech of the middle-aged generation of

Umatilla Sahaptin speakers differs from that of their Elders with regard to exactly this grammatical feature. Blackburn Morrow 2006 found that the inverse voice construction varied by generation in her experimental data, with the inverse voice construction of the three middle-aged speakers not matching that of the two elders in form, function, or frequency of use. Thus, data collected from middle-aged speakers may not show this grammatical feature, and when the current elders pass on, the remaining consultants may not be sensitive to its use in texts collected from older generations.

Current UO Ph.D. student Joana Jansen will finish her dissertation in August of 2009, a reference grammar of Yakima Sahaptin; she will then begin postdoctoral work sponsored by this project. In collaboration with Yakima elder (and UO Ph.D. student) Virginia Beavert, Jansen has recorded, transcribed, and analyzed a corpus of roughly 3000 clauses of Sahaptin texts in Toolbox. An additional 25 hours of recordings (with various genres, including conversation) have been collected, but not analyzed. In addition, Northwest Sahaptin myth and narrative texts were collected by Melville Jacobs from 1926 to 1930, and were published in three volumes (Jacobs 1929, 1934, 1937). This rich collection is comprised of 81 myths and narratives, with roughly 18,000-19,500 clauses. Currently, the texts are printed in English and Sahaptin, with no interlinear analysis; the orthography, while sufficient to represent the sound contrasts of the language, is not in current use. In this project, Jansen will put Jacobs' texts into the modern orthography, then put them into Toolbox, where they will be interlinearized and annotated for the needs of the CRP. Jansen's Toolbox analysis will be enhanced to match CRP standards. The Jacobs texts will be added to the Referential Hierarchies corpus along with the subset of Jansen's texts already approved for dissemination. Additional texts recorded by Jansen will be entered and analyzed, and will thus be available for frequency counts and illustrative examples, but public dissemination of the entire texts cannot be assured without explicit permission from both individual speakers and the tribal council.

The Evolution of Hierarchical Systems
The second primary aim of this project is to identify sources for inverse alignment (SAP > 3, also called hierarchical alignment), inverse voice (3PROX > 3OBV; cf. Gildea 1994) and integrated inverses (SAP > 3PROX > 3OBV) wherever possible, targeting in particular Cariban, Sahaptian, and Movima.

In contrast to the reasonably robust literature on the evolution of nominative and ergative alignment patterns, there is very little literature addressing the evolution of hierarchical effects in language. Both DeLancey (2001) and Zúñiga (2006) assert the theoretical pre-eminence of deixis (SAP > 3) in both inverse alignment and integrated inverse systems, which suggests that the integrated inverse would evolve by adding a distinction between 3PROX > 3OBV to a prior SAP > 3. Possible sources for inverse alignment include reanalysis of zero third-person marking (this may have happened in Huastec, cf. Zavala 1994, and Reyesano, cf. Guillaume to appear), and reconstructed sources of explicit inverse morphemes include incorporation of cislocative markers into the verbal complex (cf. arguments in DeLancey 2001 that this is the origin of inverse morphemes in Kuki-Chin and Dravidian languages) and reanalysis of third person morphology (cf. Guillaume to appear for Reyesano). While the independent genesis of inverse alignment appears to be well attested, no specific case of an integrated inverse system has been reconstructed to extension of a prior inverse alignment system.

In contrast, Givón (1994) explicitly argues that inverse voice systems (his "pragmatic inverse") can arise out of an active-passive alternation, such that active > direct and passive > inverse, or out of word order alternations, such that canonical order > direct and patient topicalization > inverse. It is well-attested that "overuse" of a passive construction can result in its reanalysis as a new structural type, an active (ergative) clause, and Givón specifically argues that inverse voice must be a transitional stage between passive and active ergative: passive > inverse > active ergative. It remains an open question whether the shift from passive > inverse voice is merely a change in function, or if it also can result in a new structural type of transitive inverse clause. Givón futher asserts that the grammar of inverse voice can then become obligatory in the interaction between inherently more topical SAPs and third persons (Givón's "semantic inverse"), thus creating an integrated inverse. This appears to be instantiated in the Kiowa-Tanoan i9ntegrated inverse, in which the grammar of the obligatory inverse is isomorphic to a passive, cf. Mithun 1999, Zúñiga 2006). In Givón's scheme, inverse alignment would be created as yet another developmental stage, the "semantic inverse" left behind when an integrated inverse loses the inverse voice; this development is not yet attested or reconstructed for a specific language or language family.

The sole proposal for origins of hierarchical ergative case-marking is reanalysis of oblique instrumental phrases as inanimate ergative subjects (Garrett 1990), with the later extension of ergative case-marking up the hierarchy to all third persons. The origins of DOM are transparently extension to P of the marking for Goal/Recipient, but the starting point and later extensions are unknown.

This project will reconstruct sources of the inverse agreement system reconstructed to Proto-Cariban (Gildea 1998) and found in almost all modern languages of the family. Recent descriptive

work on Pekodian Branch languages Bakairi (Meira 2003) and Inkpéng (Pacheco 2001) suggest that phonological loss of third-person marking might have given rise to a formally hierarchical pattern. In collaboration with Scott DeLancey, we will reconstruct the separate elements of the Sahaptian hierarchical grammar: SAP clitics from free personal pronouns in Plateau Penutian, case-marking to directional postpositions (source morphemes known, source construction unknown), and the source of the inverse prefix is currently completely mysterious. In collaboration with Haude (IP 03 from this CRP), we will also reconstruct the integrated inverse in Movima (the source is probably an inverse voice arising out of reanalysis of nominalizing morphology, cf. preliminary discussion in Haude to appear). Within the CRP, the PI will bring in a diachronic perspective to all CRP meetings and workshops.

References

Blackburn Morrow, Jesse. 2006. *Topicality and the inverse voice in Umatilla Sahaptin: An experimental study.* University of Oregon MA Thesis.

DeLancey, Scott. 2001. Lectures on Functional Syntax. Notes for LSA Summer Insitute, UC Santa Barbara. (http://darkwing.uoregon.edu/~delancey/sb/fs.html)

Gildea, Spike. 1994. Semantic and pragmatic inverse — "inverse alignment" and "inverse voice" — in Carib of Surinam. In Givón (ed), 187-230.

Gildea, Spike. 1998. *On Reconstructing Grammar: Comparative Cariban Morphosyntax.* Oxford: Oxford University Press.

Gildea, Spike. In press. Linguistic Studies in the Cariban Family. *Handbook of South America*, ed. by Lyle Campbell and Veronica Grondona. Berlin: Mouton de Gruyter.

Guillaume, Antoine. To appear. Hierarchical agreement and split intransitivity in Reyesano. *International Journal of American Linguistics.*

Givón, T. 1994. Introduction. *Voice and Inversion*, ed. by T. Givón, 3-44. Amsterdam: John Benjamins.

Haude, Katharina. To appear. The intransitive basis of Movima clause structure. *Ergativity in Amazonia*, ed by Francesc Queixalós and Spike Gildea. Amsterdam/Philadelphia: John Benjamins.

Jacobs, Melville. 1929. Northwest Sahaptin texts. University of Washington Publications in Anthropology, Vol. 2, pp. 175-244. Seattle: University of Washington Press.

Jacobs, Melville. 1934. Northwest Sahaptin Texts (English Translations). Columbia University Contributions to Anthropology, Vol. 19, Part 1. New York: Columbia University Press.

Jacobs, Melville. 1937. Northwest Sahaptin Texts (Indian Text). Columbia University Contributions to Anthropology, Vol. 19, Part 2. New York: Columbia University Press.

Meira, Sérgio. 2003b. A marcação de pessoa nos verbos em Bakairi (Karíb). *Ergatividade na Amazônia II*, ed. By Francesc Queixalós. Paris: CNRS~CELIA.

Mithun, Marianne. 1999. *The Languages of Native North America.* Cambridge University Press: Cambridge.

Pacheco, Frantomé. 2001. *Morfossintaxe do verbo Ikpeng.* Ph.D. dissertation, Universidade Estadual de Campinas (UNICAMP).

Rigsby, Bruce, and Noel Rude. 1996. Sketch of Sahaptin, a Sahaptian language. Handbook of North American Indians, Vol. 17, Languages, ed. I Goddard, pp. 666-92. Washington, D.C.: Smithsonian Institution.

Rude, Noel. 1994. Direct, inverse and passive in Northwest Sahaptin. Voice and Inversion, ed. T. Givón, pp.101-119. Amsterdam: John Benjamins.

Tomlin, Russell. 1997. Mapping conceptual representations into linguistic representations: The role of attention in grammar. in Jan Nuyts and Eric Pederson, eds., *Language and conceptualization*, pp. 162-89. Cambridge: Cambridge University Press.

Zavala, Roberto. 1994. Inverse Alignment in Huastec. Función 15/16.27-81

Zúñiga, Fernando. 2006. Deixis and Alignment: Inverse Systems in Indigenous Languages of the Americas. Amsterdam: John Benjamins.